

REFLECTION

FULL WAVEFORM INVERSION

Kenneth Otabor Irabor

Department of Earth Science and Engineering

Imperial College London

Supervisor :: Professor Mike Warner

A thesis submitted in fulfilment of the requirement for the degree of
Doctor of Philosophy and The Diploma of Imperial College London

30th September 2016

Disclaimer

Legal Disclaimer

I declare that this thesis is entirely my original work and that where any material could be misconstrued as the work of others, it had been fully cited and referenced, and/or with appropriate acknowledgements given.

Name of Student: **Kenneth Otabor Irabor**

Signature:

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

This thesis is dedicated to the memory of my late dad Mr. Benedict
Imhanralu Irabor, to my elder brother Osazuwa and to my wife
Tonye.

Acknowledgements

The person I am today is the hand work of the supernatural, through the efforts of beautiful human beings.

My sincere appreciation goes to my supervisor Professor Mike Warner, for giving me the opportunity to do a PhD with him, and providing a balanced environment that allowed for the development of my research skills. I am most grateful for everything Mike did and continues to do for me. I came to Imperial College knowing close to nothing about computer programming. I received two Fortran programming books from Mike the very first day I started my PhD and they launched me into this ever beautiful world of computational geophysics. And for that and so many more I am indeed thankful.

I thank all the members of the Fullwave3D and Fullwave Game Changer teams, Imperial College for being wonderful and making the past five years worth every moment. I am forever grateful for this opportunity to have met you all. I thank Mrs Akela Okoye for providing all the laughter especially during those very gloomy days, bringing down the roof when things worked so well, and continually encouraging me. Lluís, for those many emails, and for always coming down from his desk with a smile to set things straight for me, many Thanks. Ivan Stkl for the extremely useful discussions in and out of group meetings. Adrian gave me loads of programming tips and was always available for yet another brain picking. Gang provided the many laughs at group meetings and constantly helped me challenge myself, I appreciate you in many ways you will never know. Nuno for giving me a renewed belief in the technique developed here.

My immediate family have been of tremendous support especially at the very final stages of my research when I panicked on many occasions. My wife Tonye constantly checked on me to make sure I my sanity was still in check. I have learnt so much about life from you, and I look forward to learning more. For many things I cannot say, thank you. I thank my mum and my siblings for giving me the foundation I had growing up as a kid in Lagos Nigeria. Those days formed the basis of who I am, and I am ever grateful to you all. My elder brother Godswill Osazuwa Irabor was my first teacher and continues to challenge me each day to be better. Thank you for seeing me through school and paying my fees.

I thank the Exxon Mobil for paying for my MSc, and the sponsors of Fullwave3D and Fullwave Game Changer for paying for my PhD. Many thanks to the developers and maintainers of free and/or open source softwares that I used to develop my codes. Thanks to SEG and Chevron for providing the datasets used for my blind experiments. And finally, a special thanks to Institut Franais du Petrole for providing the popular Marmousi model used for my synthetic examples.

Finally, my ultimate help and strength comes from the Lord God who made heaven and earth. I am grateful for Life, Grace and Prayers answered.

Thank You Lord

Abstract

The Full Waveform Inversion (FWI) gradient is composed of a low wavenumber tomographic component and a high wavenumber migration component. A successful application of FWI requires that the low wavenumber parts of the model be recovered before the high wavenumbers. This process becomes difficult in datasets dominated by pre-critical angle reflection energies. Reflection waveform inversion (RWI) has been proposed as an alternative to help bootstrap the FWI method for reflection data.

In this thesis, I have made a novel contribution to RWI using Finite Difference Explicit Wavefield Decomposition (FDEWD). This method improves the wavefield decomposition process by cleanly decomposing the wavefields into four components using finite difference method and Fourier transform. Four component wavefields travelling left, right, up and down are simultaneously derived in this method compared to just opposite directions possible with most other methods. FDEWD also lacks the evanescent energy present in traditional Fourier based separation. The extra layer of separation introduced by FDEWD ensures that the tomographic component of the gradient is formed by energies propagating within and close to the first Fresnel zone, hence yielding a cleaner tomographic update.

The FDEWD method developed here was then used in an RWI context to successfully invert a synthetic dataset and a blind dataset. The scheme involved a migration update step with an exaggerated

step length and a tomographic update step with true step length computation. The results obtained shows that the new method produces superior results compared to the method based on direct separation of the total wavefields. FDEWD also allows for transmission FWI to be performed without the need to mute the data in any way.

We have implemented the scheme here in a 2-D constant density acoustic wave equation. It is, however, possible to extend this method to 3-D, anisotropic and elastic problems.

Abbreviations

FWI	Full Waveform Inversion
AWI	Adaptive Waveform Inversion
FDEWD	Finite Difference Explicit Wavefield Decomposition
RTM	Reverse Time Migration
MBTT	Migration Based Travel Time
...	...

Contents

1	Introduction	1
1.1	Overview	1
1.2	Full Waveform Inversion	3
1.2.1	Seismic Inverse Problem	3
1.2.2	Full Waveform Inversion	5
1.2.3	Scale Separation and Full Waveform Inversion	6
1.3	Reflection FWI	7
1.4	Thesis Overview	8
1.5	Convention	9
2	FWI Theory	11
2.1	General Inverse Theory	11
2.1.1	Iterative Linearised inversion - Local Optimisation	15
2.2	Forward problem in FWI	29
2.2.1	Acoustic Wave Propagation	29
2.2.2	Finite Difference Solution	30
2.2.3	Practicalities of applying finite difference method	34
2.3	Inverse problem in FWI	36
2.3.1	Gradient computation in Conventional FWI	36
2.3.2	Resolution of the FWI gradients	38
2.4	Theory of Reflection FWI	39
2.4.1	Reflection FWI Gradient in a Smooth Model	41
2.4.2	Reflection FWI Gradient in a Rough Model	46
2.5	Conclusion	52

3	Finite Difference Explicit Wavefield Decomposition	53
3.1	Theory of Finite Difference Explicit Wavefield Decomposition . . .	54
3.1.1	Fourier Wavefield Decomposition	54
3.1.2	Finite Difference Explicit Wavefield Decomposition	59
3.2	Gradient Computation in FWI with Finite Difference Explicit Wave- field Decomposition	70
3.2.1	Analysis of the gradients of FDEWD	77
3.3	Applications of FDEWD	81
3.3.1	FDEWD and the PML Absorbing Boundary Condition . .	81
3.3.2	Gradients Mixing, Scaling and the Subspace Method . . .	83
3.3.3	FDEWD and Conventional FWI	85
3.3.4	FDEWD and Reflection FWI	85
3.3.5	FDEWD and RTM	85
3.4	Conclusion	85
4	Numerical Testbed for Seismic Tomography	87
4.1	Introduction	87
4.2	Design	88
4.3	Implementation	92
4.3.1	Procedural Implementation	92
4.3.2	Object-Oriented Implementation	96
4.3.3	Benchmarking	101
4.4	Refactoring and Improvement	101
5	Synthetic Data Examples	102
5.1	Introduction	102
5.2	Problem Definition	103
5.3	Experiment 1 : Reflection FWI with a low background velocity . .	121
5.3.1	Inversion procedures and strategies	121
5.3.2	Results from the low (slow) background starting model . .	123
5.4	Experiment 2 : Reflection FWI with a high background velocity .	129
5.4.1	Inversion procedures and strategies	129
5.4.2	Results from the high (fast) background starting model . .	130

6	Blind Data Example	153
6.1	Problem Set-up	153
6.1.1	Data	155
6.1.2	Start Model	155
6.1.3	Source Wavelet	157
6.1.4	Geometry	157
6.2	QC and Pre-processing	161
6.3	Inversion Strategies	162
6.4	Results	164
6.4.1	Frequency continuation FWI	164
6.4.2	Transmission FWI without muting	165
6.4.3	Reflection FWI via finite difference wavefield decomposition	167
6.5	Discussion	171
7	Concluison	174
7.1	Conclusion	174
7.2	Recommendation and Future Work	176
A	Wavefields	178
A.1	Evolution of wavefields through the Marmousi model	178
B	Intermediate Wavenumbers	181
	References	201

List of Figures

1.1	Application of FWI to 3D anisotropic dataset (Warner et al., 2013). Full waveform inversion was applied to a 3-D anisotropic data, and produced a high resolution model of the overburden	2
1.2	The forward problem involves computing data response from a known model. The inverse problem on the other hand involves estimating model parameters, given the data response	4
2.1	A linear map from model space into data space	13
2.2	A nonlinear map from model space into data space	15
2.3	Approximation of the sine function by series expansion about 0. In each of the approximations, n is the highest order of the derivatives used. Using higher orders in the expansion shows the sine functions being fitted further away from 0	16
2.4	A single variable function with multiple minima	19
2.5	Local approximation of the function using first order information. A linear graph has no minimum value hence higher orders series approximations are used	19
2.6	Local approximation of an objective function using second order derivative information around iterate x_k	21
2.7	Local approximation of an objective function using second order derivative information around iterate x_{k+1}	21
2.8	Local approximation of an objective function using second order derivative information around iterate x_{k+2}	22

LIST OF FIGURES

2.9	Local approximation of an objective function using second order derivative information around iterate x_{k+3}	22
2.10	Local approximation of an objective function using second order derivative information around iterate x_{k+4}	23
2.11	Local approximation of the function using first order information. A linear graph has no minimum value hence higher orders series approximations are used	24
2.12	Local approximation of an objective function using second order derivative information around iterate x_{k+4}	25
2.13	Gradient of an objective function $f(x_1, x_2)$. Each arrow points in the direction of maximum ascent from where point at which it is drawn	25
2.14	Gradient descent method solution for the minimization of the Rosenbrock function	27
2.15	A discretized Marmousi model. The wavefields computed in the model are only available at the grid points seen in this model . . .	31
2.16	A standard 4th order stencil for computing the Laplacian of the wave equation. The node represented by the red sphere represents the grid point at which where the Laplacian is currently evaluated. All other nodes (green spheres), together with red node, contribute to the evaluation of the Laplacian.	32
2.17	A compact 2D finite difference stencil. This stencil yields a more accurate solution to the wave equation as it incorporates additional neighbouring grid points	33
2.18	The wavenumber spectrum presents in the gradient has dependency on the local angular aperture (θ), the local velocity (c) and the local angular frequency (ω). Large angular apertures give rise to low wavenumber updates	39
2.19	Conventional FWI gradients are shown here. The gradient points in the direction in which the model is updated	40
2.20	Forward wavefields propagating through a smooth model in forward time. The smoothness of the model and the lack of scatterers in it makes the wavefields travel only downwards	42

LIST OF FIGURES

2.21	Backward wavefields propagating through a smooth model in backward time. Absence of scatterers also allows the energies propagate only downwards	43
2.22	Backward wavefields propagating through a smooth model in forward time. The direction of propagation has been reversed by the reversal in the temporal direction	43
2.23	Migration update in a smooth model formed by cross-correlating the source wavefield in forward time and the receiver wavefield in forward time. The migration attempts to build scatterers in the model by forming the update around the subsurface location that acts as a virtual source of energy	44
2.24	Migration update in a smooth model formed by cross-correlating the source wavefield in forward time and the receiver wavefield in forward time. The migration attempts to build scatterers in the model by forming the update around the subsurface location that acts as a virtual source of energy	45
2.25	A cartoon showing a rough model (with scatterers having wavenumbers as to match the dominant wavelength of the source and thus generate reflections)	46
2.26	Source wavefields through a rough velocity model. The solid line represents direct wavefields from the source (propagating downwards in forward time), while the dashed line represents the scattered wavefield (propagating upwards in forward time)	47
2.27	Receiver wavefields in a rough velocity model in backward time. The solid line represents direct wavefields from the receiver (propagating downwards in backward time), while the dashed line represents the scattered wavefield (going upwards in backward time)	48
2.28	Receiver wavefields in a rough velocity model in forward time. The time reversal means that the direct receiver wavefields propagate upwards in forward time, while the scattered receiver wavefields propagate downwards in forward time	48
2.29	Reflection FWI gradient formed by cross-correlating the source wavfield in forward time with the adjoint wavefields in forward time	49

LIST OF FIGURES

2.30	Migration component of FWI gradient formed by source and adjoint wavefields travelling in the opposite direction in forward time	49
2.31	Tomography update in RWI formed by source and adjoint wavefields propagating in same direction in forward time.	50
2.32	The total reflection FWI update. notice that the migration component has a much stronger amplitude compared to the tomography updates, with much less amplitude	50
2.33	The total reflection FWI update. notice that the migration component has a much stronger amplitude compared to the tomography updates, with much less amplitude	51
3.1	Wavefield decomposition via Fourier transform. The wavefields are decomposed into left and right going waves. However, energies from "standing" waves appear as noise (horizontal lines) and are not desirable for applications in FWI and RTM	55
3.2	Directional wavefield splitting using Fourier transform. In this scheme the horizontal and vertical splitting is performed on the same wavefield. The splitting shows the present of evanescent energies	57
3.3	Decomposition of the wavefield into horizontal and vertical components. The sum of these components exactly produces the original wavefield	60
3.4	Source illumination of Finite Difference Decomposed Wavefields .	62
3.5	Directional wavefield splitting using Finite Difference Explicit wavefield Decomposition and Fourier transform.	64
3.6	Percentage difference between a snapshot of the sum of the four components of the wavefield produced by FDEWD, and a snapshot of the same wavefield computed with a conventional finite difference stencil. Although the phase of the original wavefield appears a bit distorted, however, inversion experiments, as discussed in later chapters, do not show any significant issue associated with this distortion.	65

LIST OF FIGURES

3.7	Evolution of a wavefield in time. The component wavefields also evolve as does the full wavefield. The leftmost column is the evolution of total wavefield, the middle column is the evolution of the horizontal wavefields, while the rightmost column is the evolution of the vertical wavefields.	67
3.8	An evolution of wavefields in time through a section of the Marmousi model	68
3.9	An evolution of wavefields in time through a section of the Marmousi model	69
3.10	Total gradient	72
3.11	Tomographic gradient contributions from cross-correlating horizontal wavefields propagating along same direction in time	73
3.12	Tomographic gradient contributions from cross-correlating vertical wavefields propagating along same direction in time	74
3.13	Migration gradient contributions from cross-correlating horizontal wavefields propagating along different directions in time	75
3.14	Migration gradient contributions from cross-correlating vertical wavefields propagating along different directions in time	76
3.15	Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 4Hz	78
3.16	Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 6Hz	79
3.17	Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 8Hz	80
3.18	The computational and the PML domain, showing supports for decay constant, α and damping distance, d	84
4.1	PDe Solver Component.	90
4.2	Tomography Component.	90
4.3	Optimiser Component.	91

LIST OF FIGURES

4.4	Utils Component.	91
4.5	Seismic Utils Component.	92
4.6	Relationships amongst all components.	93
5.1	The true velocity model with the locations of the well logs. The well logs will be used to analyse the accuracy of the inverted model	103
5.2	A sample shot data generated from the true model. The presence of different energy modes makes this data very friendly for seismic inversion via the multiscale FWI	104
5.3	A sample shot data generated from a smooth Marmousi model. Though lacking reflections and back-scattered energies, the transmission energies present in the data still make the model a suitable starting model for multiscale FWI.	105
5.4	A sample shot data from the true Marmousi model	107
5.5	A sample shot data from the true Marmousi model	108
5.6	Source wavelet and its amplitude spectrum	109
5.7	Low Background Start Model (Model-B), and its residuals	111
5.8	High Background Start Model (Model-C), and its residuals	112
5.9	Comparisons of well logs from the true model, true background model (Model-A) and high background start model (Model-C). Red logs represent the true Marmousi model. Blue logs are the very good start model. Black logs represent the high background start model.	114
5.10	Comparisons of well logs from the true model, true background model (Model-A) and low background start model (Model-B). Red logs represent the true Marmousi model. Blue logs are the very good start model. Black logs represent the low background start model.	115
5.11	Sample predicted data from the smooth, high background start model	117
5.12	Data residuals for the previous predicted data	118
5.13	Another sample predicted data from the smooth, high background start model	119

LIST OF FIGURES

5.14	Data residuals for the previous predicted data	120
5.15	Comparison between shot data	122
5.16	Initial high and low wavenumber updates. The low wavenumber update contains the wavenumbers needed to build the background model. The high wavenumber update is cycle-skipped and does not contribute to building the background model directly	125
5.17	Well logs from the Intermediate result, compared with logs from the start model (Model-B). Black represents the start model, Red represents the true model, while Blue represents the updated model using the tomographic components of the gradient.	126
5.18	Results obtained by doing tomographic updates of the model starting from 5Hz following the scheme outlined in this section	127
5.19	Final results obtained using finite difference explicit wavefield decomposition and conventional FWI. Colour scale same as in Figure 5.18	127
5.20	Well logs from the final inversion result, compared with logs from the start model (Model-B). Black represents the start model, Red represents the true model, while Blue represents the final updated model.	128
5.21	Initial high and low wavenumber updates. The high wavenumber updates are initially injected into the model. These then generate scattered energies used in building the low wavenumber updates during subsequent iterations.	131
5.22	Well logs from the initial result (3 to 6Hz), compared with logs from the start model (Model-C). Black represents the start model, Red represents the true model, while Blue represents the updated model. The surface location of each log is shown on the log.	132
5.23	Well logs from the initial result (3 to 6Hz), compared with logs from the very good start model (Model-A). Black represents the the very good start model, Red represents the true model, while Blue represents the updated model. The surface location of each log is shown on the log.	133

LIST OF FIGURES

5.24	Well logs from the final inversion result, compared with logs from the start model (Model-C). Black represents the start model, Red represents the true model, while Blue represents the final updated model.	135
5.25	Final results obtained using finite difference explicit wavefield decomposition and conventional FWI	136
5.26	Inverted models and its residuals	137
5.27	Sample shot data from the right side of the model, computed using the inverted model	138
5.28	Comparison of near traces from the shot gather in Figure 5.27. The red traces represent those from the start model, while the black are from the true model.	139
5.29	Comparison of near traces from the shot gather in Figure 5.27. The red traces represent those from the recovered model, while the black are from the true model.	140
5.30	Comparison of mid traces from the shot gather in Figure 5.27. The red traces represent those from the start model, while the black are from the true model.	141
5.31	Comparison of mid traces from the shot gather in Figure 5.27. The red traces represent those from the recovered model, while the black are from the true model.	142
5.32	Comparison of far traces from the shot gather in Figure 5.27. The red traces represent those from the start model, while the black are from the true model.	143
5.33	Comparison of far traces from the shot gather in Figure 5.27. The red traces represent those from the recovered model, while the black are from the true model.	144
5.34	Sample shot data from the center of the model, computed using the inverted model	145
5.35	Sample shot data from the left side of the model, computed using the inverted model	146

LIST OF FIGURES

5.36	Comparison of near traces from the shot gather in Figure 5.35. The red traces represent those from the start model, while the black are from the true model.	147
5.37	Comparison of near traces from the shot gather in Figure 5.35. The red traces represent those from the recovered model, while the black are from the true model.	148
5.38	Comparison of mid traces from the shot gather in Figure 5.35. The red traces represent those from the start model, while the black are from the true model.	149
5.39	Comparison of mid traces from the shot gather in Figure 5.35. The red traces represent those from the recovered model, while the black are from the true model.	150
5.40	Comparison of far traces from the shot gather in Figure 5.35. The red traces represent those from the start model, while the black are from the true model.	151
5.41	Comparison of far traces from the shot gather in Figure 5.35. The red traces represent those from the recovered model, while the black are from the true model.	152
6.1	Well log and source wavelet provided by Chevron	154
6.2	Sample Shot Data (with full bandwidth)	156
6.3	Initial model provided. Notice the well defined nature of the sea bed	157
6.4	Sample Shot Data for a low pass filtered source propagated through the start model. There is a rich supply of transmission energies in this data, which can help in building the overburden via transmission FWI	158
6.5	Well data comparison between the initial model and the true model	159
6.6	Original source wavelet and a low pass version (Ormsby : 2.5 - 4.5 - 6.6 - 9.8)	160
6.7	Near Trace Plot	162
6.8	Observed Data filtered to acceptable propagation frequency using a 25 m mesh	163
6.9	Inverted results from conventional FWI	164

LIST OF FIGURES

6.10	Gradients from transmission mode FWI via the use of the horizontal component of the wavefields in FDEWD. Here, horizontal smoothing is applied to the gradient.	165
6.11	Enhanced gradient from transmission mode FWI via the use of the horizontal component of the wavefields in FDEWD. Here, the horizontal smoothing is varied and applied to the gradient. A vertical adaptive weighting factor was applied to enhance shallow parts of the gradient	166
6.12	Initial gradient model from reflection mode FWI via the use of the vertical component of the wavefields in FDEWD. The start model for this mode was the model derived from the transmission mode FWI	167
6.13	Well data comparison between the initial model, the recovered model from transmission FWI and the true model	168
6.14	Well data comparison between the initial model, the recovered model from recovered FWI and the true model	169
6.15	Model obtained when FWI is applied on the model obtained from transmission FWI	170
6.16	Final results of finite difference explicit wavefield decomposition and subsequent FWI	170
6.17	Final Inverted Model	172
6.18	Published results on the dataset using AWI and subsequent FWI (Warner et al., 2015)	172
6.19	Published results on the dataset using scattering angle filters and subsequent FWI (Alkhalifah, 2016)	173
6.20	Final results of finite difference explicit wavefield decomposition and subsequent FWI	173
A.1	Evolution of a wavefield in time through a section of the Marmousi model. Shot position is on the left side of the model	179
A.2	Evolution of a wavefield in time through a section of the Marmousi model. Shot position is on the right side of the model	180

LIST OF FIGURES

B.1	Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 3Hz	182
B.2	Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 4Hz	183
B.3	Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 5Hz	184
B.4	Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 6Hz	185
B.5	Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 7Hz	186
B.6	Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 8Hz	187

Chapter 1

Introduction

1.1 Overview

Full Waveform Inversion (FWI) has become a sought after technique in building high-resolution, high-fidelity quantitative models (velocity, density, anisotropy, attenuation etc) of the subsurface. In general, full waveform inversion involves a search for a model that explains the data from a seismic experiment, wiggle for wiggle, and subject to the applicable physics of wave propagation. This search can be done either in a Monte Carlo-type approach, where we calculate zillions of models and try to choose from amongst them the models that best explain the data, along with their corresponding uncertainties; or in a multiply iterated linearised inversion process, where we start from an initial model, close enough to the true model, and employ various adjoint-type techniques to update the model.

In the past decade, FWI has been successfully applied to various 3D datasets ([Warner et al., 2008](#)), anisotropic datasets ([Warner et al., 2013](#)), multi-parameter datasets ([da Silva et al., 2016](#); [Yang et al., 2014](#)), land datasets ([Al-Yaqoobi, 2013](#)), a few elastic datasets ([Brossier et al., 2009](#)), and in global seismology ([Fichtner et al., 2009, 2013](#)). The adoption of FWI as the preferred inversion technique can be attributed to the improvements in seismic data acquisition methods; and to an increase in computation power versus reduction in cost of modern

computers. Primarily, the models from FWI are used as input for depth migration. However, in some cases, the high resolution of the models generated also allows for a direct geologic interpretation of such FWI models. There have also been reported cases of the application of FWI technique to non-seismic datasets, showing similar success stories as seen in the seismic cases.

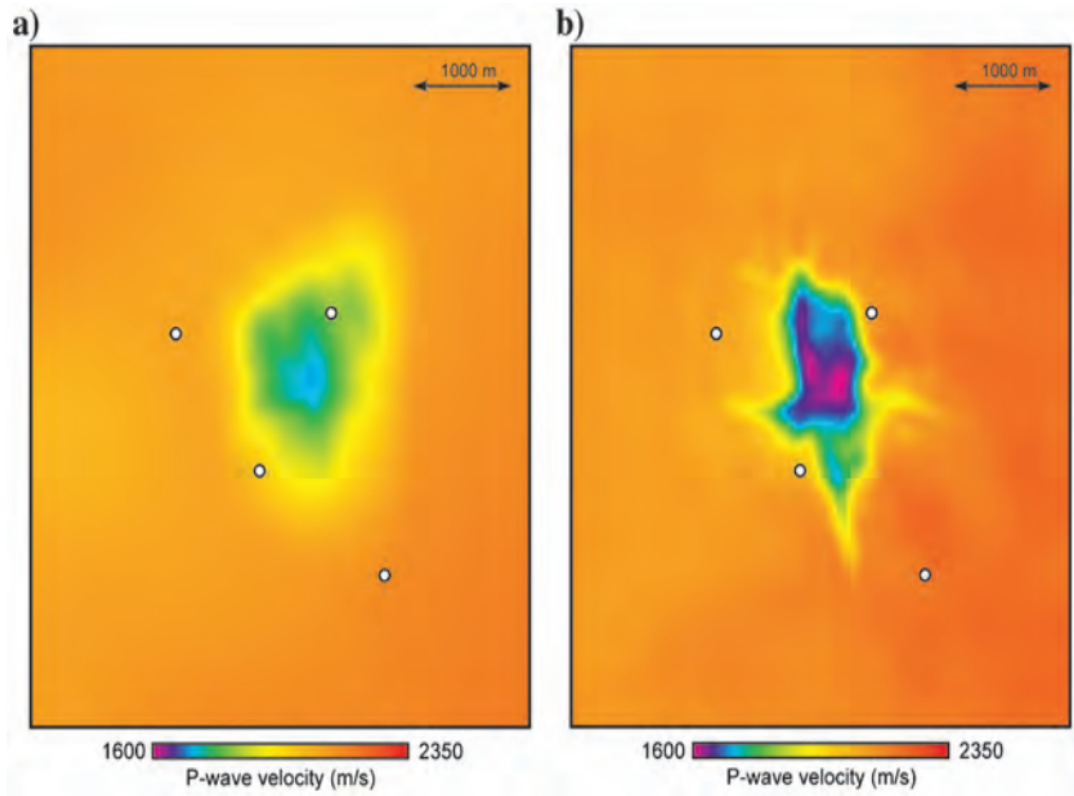


Figure 1.1: Application of FWI to 3D anisotropic dataset ([Warner et al., 2013](#)). Full waveform inversion was applied to a 3-D anisotropic data, and produced a high resolution model of the overburden

FWI has been successfully applied to datasets with good starting models and rich in turning waves, low frequencies and long offsets. However, the depth of satisfactory resolution of the inverted model has been restricted by the the depth of penetration of the turning waves. In other to invert the deep targets of the model, there is a necessity to move from the transmission dominated FWI regime to a reflection dominated FWI. Building starting velocities necessary for a successful

application of reflection FWI nonetheless has been a difficult problem for the geophysics community, and has been a subject of recent research papers. In this dissertation, I explore some of the challenges with reflection FWI (in the acoustic approximation) and propose novel techniques to overcome them, with particular reference to building starting models sufficient enough to run a successful FWI experiment. Importantly, the techniques proposed here can be easily extended to account for elasticity and anisotropy. They can also enhance and improve the results of transmission FWI.

1.2 Full Waveform Inversion

1.2.1 Seismic Inverse Problem

The scientific method usually begins with making an observation. This typically involves making some measurement. In exploration geophysics, we are concerned with physical processes within the earth that results in observable data that can be used to understand the structure, composition and dynamics of the earth. An understanding of the parameterization of a medium and the physics of a process that occurs within the medium (also known as a model) makes it possible to mimic such process and generate data similar to that observed in the medium. This is referred to as a forward problem and is a well-established problem in physics. The solution of a forward problem is referred to as a synthetic data. The necessary ingredients for the solution of a forward problem therefore consist of an assumed model (and its various parameterizations), and an appropriate governing physics. This can be extended to a seismic problem, where the model is parameterized by a distribution of elastic constants, and the appropriate physics in this case is that of elastic wave propagation. If the model were parameterized by a distribution of acoustic velocities (and densities), an appropriate physics will the acoustic wave equation. The solution of such a forward problem is a synthetic seismic data.

In a geophysical study, we only have access to the observable data, and are therefore concerned with generating a set of parameters that can describe an

earth model, given the right physics. This is known as an inverse problem.

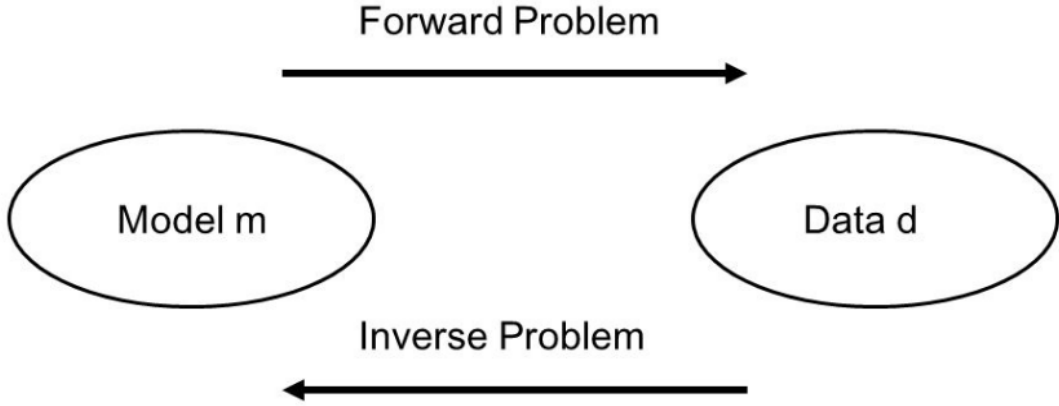


Figure 1.2: The forward problem involves computing data response from a known model. The inverse problem on the other hand involves estimating model parameters, given the data response

The seismic inverse problem, therefore, consists of estimating the physical properties of the subsurface from data recorded during a seismic experiment. These data can be surface data, cross-hole data, vertical seismic profiling data etc. The parameters estimated as a solution to a seismic inverse problem could be velocities, densities, acoustic impedances, reflectivities, anisotropic parameters, etc. The solution of the seismic inverse problem, typically involves making prediction from an assumed set of model parameters. Hence, the forward problem and the inverse problem form this close loop (Figure 1.2).

In reflection seismology, the seismic inverse problem originally involved estimating reflectivity values for subsurface interfaces between layers of varying acoustic impedances. The Born approximation, which provides the fundamental framework for the solution of this type of seismic inverse problem, requires that the model be decomposed into a smoothly varying background component (macro velocity) and a highly varying perturbation component (reflectivity) (Cohen and Bleistein, 1977; Fokkema and Berg, 1989). The Born approximation also requires that the macro model be known before the model perturbations can be determined. Therefore, traditional methods of seismic data inversion involved building the smooth, low-varying background model with travel time tomographic

methods (Billette and Lambaré, 1998; Bishop et al., 1985; Chapman and Pratt, 1992; Woodward et al., 2008) and Migration Velocity Analysis (MVA) (Symes and Carazzone, 1991). The high-varying perturbations (reflectivity/images) were built using migration.

This two-step approach works well in areas where the subsurface model is not too complex, and we can readily build the macro model using any of the methods above. However, in complex models, building reliable velocity models for accurate depth migration is very prohibitive.

1.2.2 Full Waveform Inversion

Full Waveform Inversion (FWI) came with the promise of mitigating some of the limitations of the traditional seismic inverse problem by applying the 'full' physics of wave propagation, and by combining the original two-step approach into a single inversion step (Tarantola, 1984). FWI, since its inception in the early eighties (Gauthier et al., 1986; Tarantola, 1984) and its revival in the late nineties (Pratt et al., 1996), has been established to be a method for building high-resolution, high-fidelity velocity models of the subsurface, suitable for accurate seismic imaging (depth migration) and a direct interpretation. An excellent review of the history, methodology and applications of FWI is found in Virieux and Operto (2009). As explained in the pioneering works of Tarantola (1984) and Lailly (1983), FWI searches for a model that best predicts the observed data of a seismic acquisition experiment, subject to the applicable Physics of wave propagation. Mathematically, FWI is posed as an optimization problem (Nocedal and Wright, 2006), where we seek to minimise a functional (equation 1.1) that measures the difference between some observed (real) data and the synthetic data, generated from some estimate of the model, with respect to the model parameters. In the classical formulation, this difference is the L-2 norm in some defined Hilbert space.

$$F(\mathbf{m}) = \| \mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{pred}} \|_2 \quad (1.1)$$

In its local optimisation framework, the search for a model that best predicts the data commences from an initial model close enough to the true model. This initial model is then iteratively updated till a criterion is met when the model obtained is deemed acceptable. This process follows the Born approximation (Hudson and Heritage, 1981; Tarantola, 1984; Weglein et al., 2003), where we linearised the inverse problem, and assume small model perturbation at every iteration. However, because we solve the full non-linear forward problem at each iteration, we end up solving the full non-linear inverse problem after multiple iterations.

Though originally posed as a local optimization problem, FWI has been formulated and applied in a global optimisation sense (Boschetti et al., 1996; Datta and Sen, 2016; Jin and Madariaga, 1993; Sen and Stoffa, 1991). The prohibitive computational cost of performing FWI in a global optimisation context makes the local optimisation context the preferred and common mode of application.

Conventional FWI works quite well with datasets having low frequencies, high offset-to-depth ratio, rich in turning waves and a fairly accurate starting model. There is a wide range of published works on the successful applications of FWI to field datasets (Sirgue et al., 2009; Warner et al., 2013) and various synthetic exercises (Bunks et al., 1995; Pratt et al., 1998). In some cases (Warner et al., 2013) the models generated are highly resolved to allow for a direct interpretation. The successes enjoyed by FWI in these settings have led to a renewed interest in FWI research. Though capable of direct interpretation, FWI results are still used mainly as input for depth imaging (Reverse Time Migration (RTM)) because of the limitations of the inverted frequencies.

1.2.3 Scale Separation and Full Waveform Inversion

The theoretical set-up for FWI shows that it is capable of inverting the full velocity model (macro model and perturbations). However, It became clear to early investigators that conventional full waveform inversion on reflection dataset,

which was predominantly the main type of dataset inverted initially, cannot recover the macro model (Gauthier et al., 1986) simultaneously with the model perturbations. This is partly due to the poor sensitivity of reflection data to intermediate and low wavenumber components of the model (Claerbout, 1985), and partly due to velocity-depth ambiguity that is inherent in FWI. More importantly, it was realised that progressively introducing higher frequencies into the inversion process works better than attempting to recover all the frequencies at the same time (Gauthier et al., 1986) and (Bunks et al., 1995). This approach, now referred to as multiscale FWI, starts by inverting first the low wavenumber (macro) velocity components, and gradually adds higher wavenumbers into the inversion as higher frequencies are introduced into the inversion.

Multiscale FWI in the time domain involves applying a low cut filter to both the source wavelet and the observed data. In the frequency domain, however, it is applied by systematically selecting discrete frequencies. Multiscale FWI opened new opportunities for the application of FWI.

Most of the published works on the application of FWI is still limited to building the overburden and performing inversion with frequencies below 12Hz (Warner et al., 2013). This is because Multiscale FWI is driven by the availability of low frequencies and diving waves in the dataset. However, many exploration targets lie beneath the depths covered by diving waves, and the observed datasets also lack low frequencies. In cases where the low frequencies are present, they are often contaminated by noise. Hence, there has remained the decades old puzzle of building models good enough for a successful application of FWI to reflection dominated seismic datasets, including datasets where the target is below the target depths of available diving waves. This need led to the introduction of a new variant of wave inversion referred to as Reflection Waveform Inversion (RWI)

1.3 Reflection FWI

Reflection datasets are short offset datasets, that lack in low frequencies due to source and receiver ghosts in marine environment (Lindsey, 1960), various noises,

including ground rolls on land. Reflection datasets also have their targets below the coverage of available diving waves. Hence, multiscale FWI as outlined in the previous section cannot be successfully applied directly to reflection datasets. The FWI gradient has been noted to be composed of a migration-like component and a tomography-like component (Mora, 1987, 1989). In reflection dataset inversion, the migration component is a couple of order higher in magnitude than the tomographic component. Hence, FWI on reflection datasets build the reflector but fails to build the background velocity.

Reflection-based waveform inversion (RWI) includes all the various methods used to suppress the migration component of FWI, and improve the strength of the tomographic component. It is usually performed by decomposing the model into a high and a low wavenumber component, and executing the inversion in an alternating fashion similar to migration based traveltime tomography (MBTT) (Chavent et al., 1994). In the first step, we build some reflectivity (or high wavenumber component) from initial smooth model, while in the second step we perform some form of demigration to update the background model (Chi et al., 2015; Ma and Hale, 2013; Mora, 1989; Tang et al., 2013; Wang, Chen, Zhang, Shen et al., 2013; Wang and Rao, 2009; Xu, Chen, Lambaré, Zhang, Wang et al., 2012; Yao et al., 2014).

However, reflection-based Waveform Inversion still suffers from some challenges. In most implementations of RWI, the model is decomposed into a macro velocity image and a reflectivity image. Mapping the reflectivity into the background velocity via some demigration operation requires that the migration be true amplitude image

1.4 Thesis Overview

This thesis consist of six related chapters.

Chapter two reviews the theory of FWI and reflection FWI. The chapter begins by presenting the formulation of inverse problem in geophysics. FWI is then presented as a local optimisation problem, with discussion on how its gradient is

computed. In the concluding part of this chapter, reflection full waveform inversion is reviewed, and it sets the platform for introducing the method developed in chapter three.

Chapter three introduces and discusses Finite Difference Explicit Wavefield Decomposition (FDEWD) and how it relates to FWI. The fundamentals of wavefield decomposition in FWI via the Fourier transform is reviewed. The decomposition method presented in this thesis (FDEWD) is then formally derived. The various gradients produced by the method are also discussed. The chapter concludes with a discussion of the various possible applications of this decomposition scheme.

Chapter four discusses the software developed and used in this thesis. It gives a quickly summary of the design and implementation of a testbed for doing 2D FWI on a shared memory computer.

In chapter five, the application of finite difference explicit wavefield decomposition to a synthetic dataset starting from two models is presented. An inversion scheme that incorporates the MBTT and multiscale FWI is also presented and applied in inverting the dataset.

Chapter six discusses the application of FDEWD to a blind dataset. The results are compared with published results.

Chapter seven concludes this thesis with a summary of the work and recommendations for further development.

1.5 Convention

This thesis discusses the terms **tomographic component** and **migration component** of the FWI gradient. No attempt is redefine the general meanings and usages of these terms.

Also, the word **migration** has been used in two contexts. In the first context, it has been used to denote the conventional meaning of the term migration: A

technique used to create an image of earth structures from recorded seismic data. In the other context, it denotes the high wavenumber component of full waveform inversion gradient. Context has been employed to separate the meaning.

Chapter 2

FWI Theory

2.1 General Inverse Theory

The quest for knowledge about physical laws, the internal structure of the earth, and the universe in general begins with the collection of data. Having collected the data, we then proceed to understanding, reducing and processing these data in order to extract the require information. This is the solution to the inverse problem. Inverse solutions are key fundamental problems to the understanding of the physical world we live in. The observed data are governed by well established physical laws, and by the nature of the earth and the universe ([Aki and Richards, 2002](#); [Zhdanov, 2002](#)). The process of predicting observed data using established laws of physics ([Ikelle and Amundsen, 2005](#)), for given sources within given media is referred to as the forward problem solution ([Bal, 2012](#); [Kirsch, 2011](#)). The method of reconstructing the sources of some sources of some physical phenomena, as well as the parameters of the media under study, from the observed data is referred as the inverse problem solution ([Zhdanov, 2015](#)). Inverse problem in geophysics makes use of physical fields generated by artificial or natural sources and propagated through the earth as its observed data. The solution of an inverse problem typically assumes that the forward problem is a well-posed problem of mathematical physics ([Isakov, 2006](#)).

The study of the earth's phenomena (and indeed those of any physical system, process or model) begins by parametrising a model of the earth ([Sorenson, 1980](#); [Tarantola, 2005](#)) using a minimal set of parameters that captures a complete or relevant description of such phenomena. Parametrisation is generally non-unique, as the same model can be parametrised differently using a different set of parameters. For example, if we were interested in studying seismic phenomena within the earth, a model of the earth can be parametrised using only its P-wave acoustic velocity, when we make the assumption that an acoustic earth model with a constant density explains the phenomena under observation. It can also be parametrised using the so-called Lamé's parameters ([Aki and Richards, 2002](#)), when an isotropic non-attenuating elastic earth model is under study.

In seismology, the values of the model parameters are not directly observable. However, the physical fields that propagate through the model and interact with the model can be observed. While it is much easier to record the seismic response of the earth, a direct measurement of say the velocity and density distributions of the earth cannot be carried out. There are also the additional complications introduced by the noise contamination of the observed data and the limited observation points.

The reliability of the predicted data, in terms of how well they match observed data, helps constrain the assumed physics. When data can be repeatedly predicted that optimally match their observed counterparts, such that every other differences in their match can be explained, then such physics is accepted as being sufficient to the phenomena under study. Seismology as a subject, has matured such that it has well established laws that govern its various phenomena. The problem of estimating the model parameters from measured data is known as an inverse problem .

Linear Inverse Problem

The linear forward problem is expressed as:

$$\mathbf{d} = \mathbf{G}\mathbf{m}, \quad \mathbf{m} \in M, \mathbf{d} \in D \quad (2.1)$$

where $\mathbf{d}$ is the observed data from some Banach space representing the space of geophysical data D and $\mathbf{m}$ is some model parameterization from M , another Banach space representing the space of permissible model parameters. $\mathbf{G}$ is a linear operator that maps from the model parameter space to the space of geophysical data.

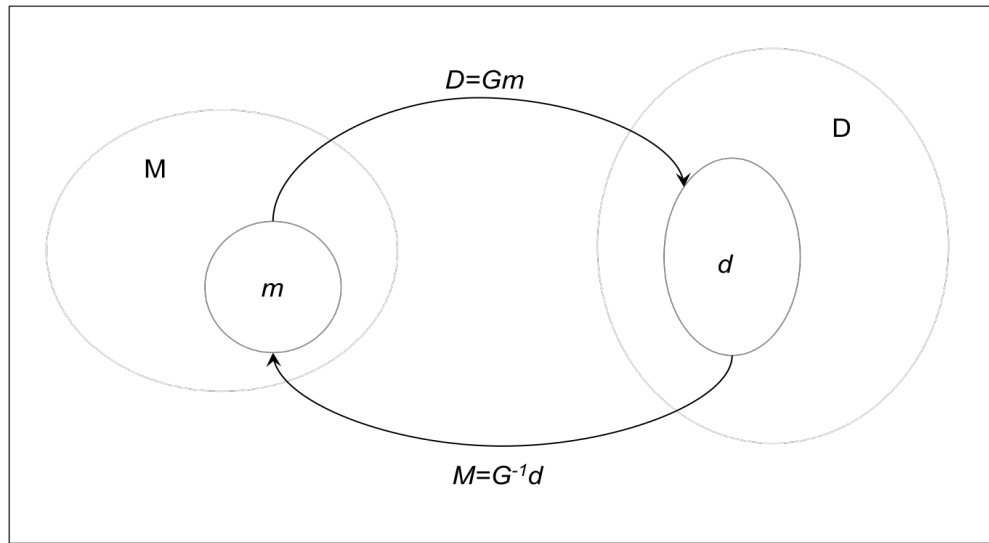


Figure 2.1: A linear map from model space into data space

Let

$$\mathbf{F} = \mathbf{G}^{-1}$$

where $\mathbf{G}^{-1}$ is the inverse of the forward operator $\mathbf{G}$.

In general, the solution of the linear inverse problem can then be expressed expressed as:

$$\mathbf{m} = \mathbf{F}\mathbf{d} \quad (2.2)$$

When the forward operator $\mathbf{G}$ is a non-square matrix, then the inverse operator $\mathbf{F}$ is not defined. We can solve the problem via the least squares method. Starting from equation 2.1, we can proceed as:

$$\mathbf{G}^T \mathbf{G} \mathbf{m} = \mathbf{G}^T \mathbf{d} \quad (2.3)$$

Equation 2.3 is called the normal system for linear forward problem. The matrix $\mathbf{G}^T \mathbf{G}$ is a square matrix, hence, the solution of the normal system is

$$\mathbf{m} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{d} \quad (2.4)$$

Nonlinear Inverse Problem

The general nonlinear forward problem can be expressed mathematically as:

$$\mathbf{d} = \mathbf{G}(\mathbf{m}), \quad \mathbf{m} \in M, \mathbf{d} \in D \quad (2.5)$$

where $\mathbf{d}$, D , $\mathbf{m}$ and M are as defined above. $\mathbf{G}$ is a nonlinear operator (or set of physical laws) describing some explicit relationship between the model parameter estimate and the observed data.

The solution to the inverse problem is

$$\mathbf{m} = \mathbf{F}(\mathbf{d}) \quad (2.6)$$

In many practical science and engineering problems, the inverse operator has no analytic expression. We proceed to solving these problem by some indirect

methods of solution. The most common type involve various gradient methods (Zhdanov, 2002, 2015)

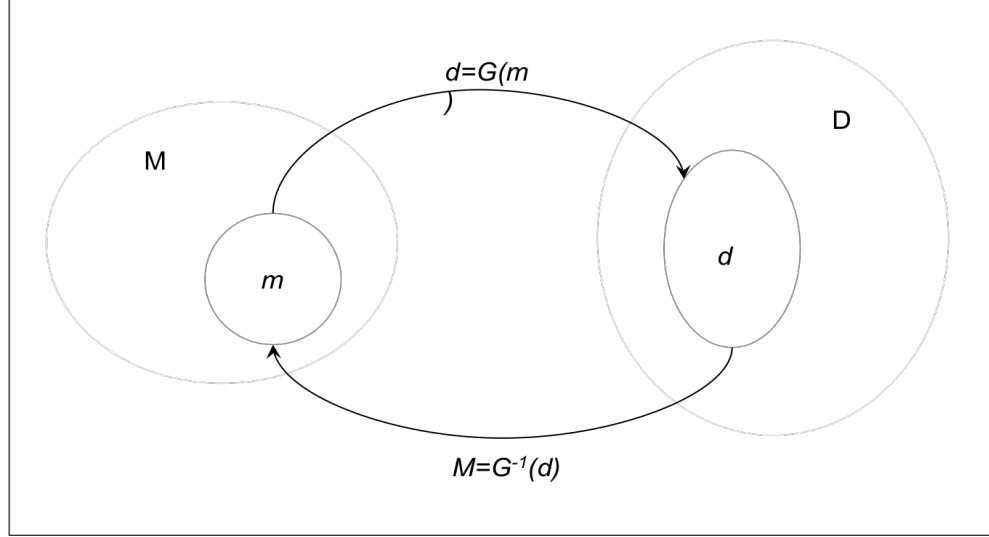


Figure 2.2: A nonlinear map from model space into data space

2.1.1 Iterative Linearised inversion - Local Optimisation

Local approximation of a function by series expansion

A smooth function is mathematically defined as one that is infinitely differentiable in its domain (Hijab and Hijab, 2007). Given an unknown smooth function, it is possible to approximate the function locally (centred about a predetermined position) as a series expansion (equation 2.7).

$$f(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + a_3(x - x_0)^3 + a_4(x - x_0)^4 + \dots \quad (2.7)$$

One way of achieving this is by using the derivatives of the function as the coefficients of the series expansion (Callahan et al., 1995). The most common of

this class of series expansions via the derivatives of a function is the Taylor series (Kreyszig, 2007) as given in equation 2.8.

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \frac{f'''(x_0)}{3!}(x-x_0)^3 + \frac{f^{(4)}(x_0)}{4!}(x-x_0)^4 + \dots \quad (2.8)$$

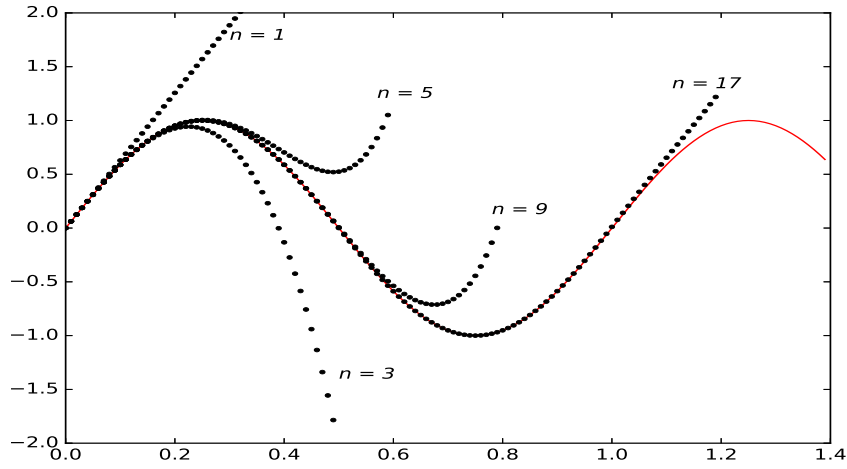


Figure 2.3: Approximation of the sine function by series expansion about 0. In each of the approximations, n is the highest order of the derivatives used. Using higher orders in the expansion shows the sine functions being fitted further away from 0

As higher order derivatives of the function are used in its approximation, the series expansion about the localised position matches the function at further distances away from the localised position (Courant and Hilbert, 1965; Heinbockel, 2006). Figure 2.3 shows the Taylor's series expansion for the sine function centred about 0.

Local optimization overview

Given a function of several variables, with various minima (and/or maxima). Optimization seeks to minimize or maximize such a function, subject to constraints

on its variables (Gill et al., 1981; Nocedal and Wright, 2006; Snyman, 2005). In case of finding the minimum of a function, the optimization problem can be formulated mathematically as given in equation 2.9.

$$\min_{x \in R^n} f(x) \quad \text{subject to} \quad \begin{cases} c_i(x) = 0, & i \in E \\ c_i(x) \leq 0, & i \in I \end{cases} \quad (2.9)$$

where x is the vector representing the variables in some space R^n , f is the objective function (or functional) that we want to maximize or minimize and c is the set of constraints satisfied by the unknowns. I and E are sets of indices, while f and c_i are scalar-valued functions of the variables x . Optimization problems that involves definition of a functional and an explicit set of constraints are known as constrained optimization (Bertsekas, 2014; Floudas and Pardalos, 2001; Schäffler, 2012).

There is another class of optimization problems where there are natural constraints on the variables. Hence, the constraints are not represented explicitly in some equation, as it assumes that they have no effect on the optimal solution (Bartholomew-Biggs, 2006; Lange, 2004; Nocedal and Wright, 2006). This class of problems, known as unconstrained problems, can also arise by reformulating a constrained optimization problem, and replacing the constraints by penalization terms in the objective function (Nocedal and Wright, 2006).

Given a minimization problem, the objective function may have several minima. These can be categorised into local minima and global minimum. A point, $x^* \in R^n$ is a local minimum if there exists a $\delta > 0$ such that

$$f(x^*) \leq f(x) \quad \text{for all } x \text{ in } \{x \in R^n : |x - x^*| \leq \delta\}$$

i.e. the local minimum is the smallest value of the function in some feasible neighbourhood. On the other hand, the global minimum is defined as the smallest value of the function over all feasible points, i.e. if x^* is a global minimum of the function $f(x)$, then

$$f(x^*) \leq f(x) \quad \text{for all } x \in R^n$$

The solution of an optimization problem is derived using a global optimization techniques ([Liberti and Maculan, 2006](#); [Pintér, 2006](#); [Schäffler, 2012](#)) or a local optimization methods ([Lange, 2004](#); [Michiels et al., 2007](#); [Nocedal and Wright, 2006](#)). Local optimizations methods appeal more when we seek a local minimum of the functional, the gradient of the objective function is available and/or when there are limitations in computational resources. Local optimization techniques generally starts from a point assumed to be within the neighbourhood of the local minimum.

Generally, local optimization algorithms work in an iterative manner by updating the current iterate (x_k) using a direction, (p_k) and a step-length, α , such that:

$$x_{k+1} = x_k + \alpha p_k \tag{2.10}$$

and

$$f(x_{k+1}) \leq f(x_k) \tag{2.11}$$

Local optimization techniques fall into two broad groups: the line search method and the trust region method ([Nocedal and Wright, 2006](#)), depending on which of the search direction and the step-length is computed first. The line search method has a close link with the concept of series approximation. The search direction is got by approximating the objective function around the current iterate using series expansion, and seeking the minimum of the series expansion.

Given a single variable objective function as depicted in [Figure 2.4](#), with multiple minima. A local optimization can be performed to locate a local minimum of the functional, by using only the first order derivatives (gradients) to approximate the function via Taylor series expansion ([equation 2.12](#)).

$$f(x) = f(x_k) + f'(x_k)(x - x_k) \tag{2.12}$$

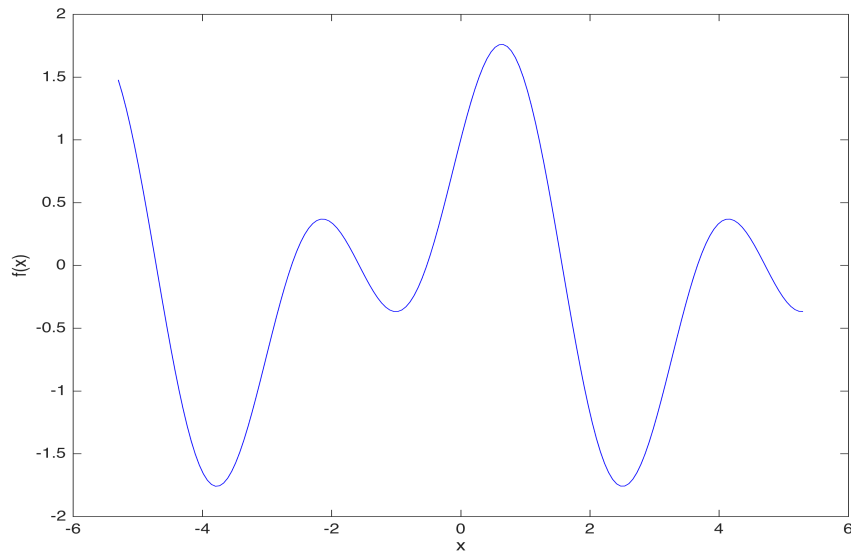


Figure 2.4: A single variable function with multiple minima

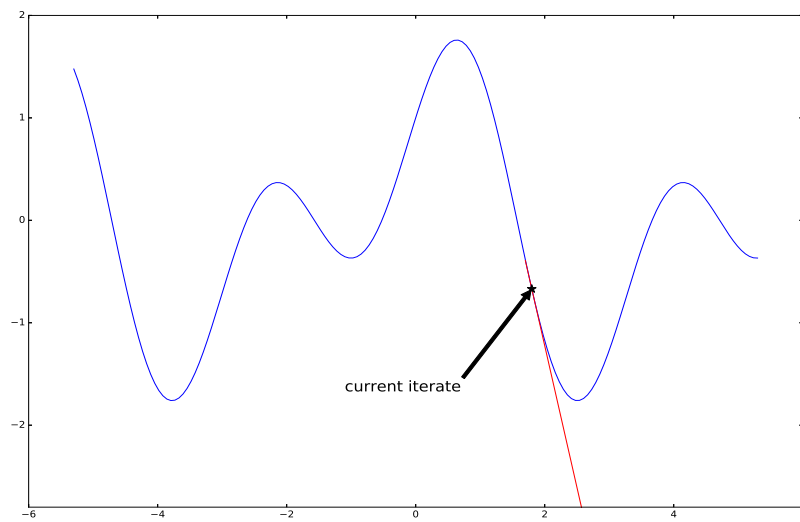


Figure 2.5: Local approximation of the function using first order information. A linear graph has no minimum value hence higher orders series approximations are used

2.1 General Inverse Theory

Equation 2.12, however, has got no minimum value. A better approximation will be to incorporate the second derivative in the series approximation. The new function becomes equation 2.13.

$$f(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{f''(x_k)}{2!}(x - x_k)^2 \quad (2.13)$$

This function can be minimised by equating its first order derivative to zero. A fundamental requirement on the second derivative of the function is that it should be positive (Fletcher, 2013). In a maximization problem, it is however expected to be negative. These requirements are connected to the concepts of convexity (and concavity) of the objective function (Drábek and Milota, 2013).

Taking the first order derivative of the functional and setting it to zero (equation 2.14) will result in equation 2.15.

$$\frac{df(x)}{dx} = \frac{d(f(x_k) + f'(x_k)(x - x_k) + \frac{f''(x_k)}{2!}(x - x_k)^2)}{dx} = 0 \quad (2.14)$$

$$x_{k+1} = x_k - f'(x_k) * (f''(x_k))^{-1} \quad (2.15)$$

Comparing equation 2.10 with equation 2.15 shows that the search direction is:

$$p_k = -f'(x_k) * (f''(x_k))^{-1} \quad (2.16)$$

while the step-length α is 1 (Boyd and Vandenberghe, 2004; Fletcher, 2013). On Figure 2.6, the minimum of the series expansion does not coincide with the local minimum of the functional. Nonetheless, iteratively approximating and minimizing the quadratic series expansion brings the minimum of the approximate function close to the local minimum of the functional (Figures 2.7 to 2.10)

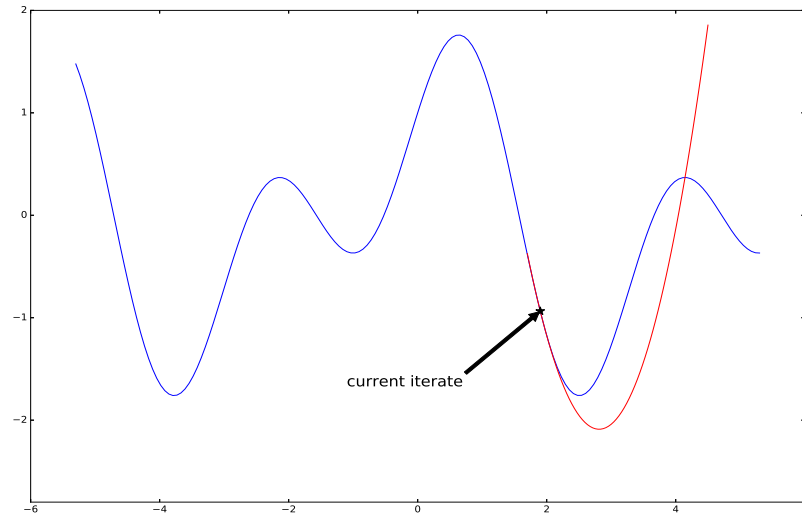


Figure 2.6: Local approximation of an objective function using second order derivative information around iterate x_k

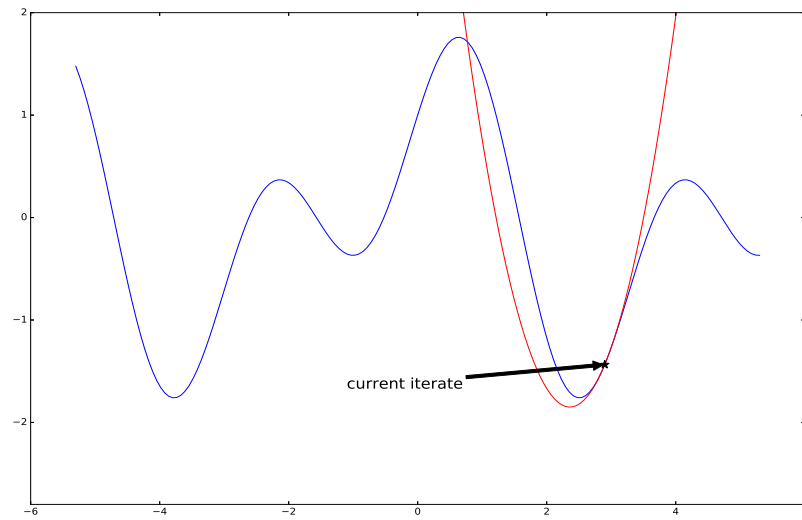


Figure 2.7: Local approximation of an objective function using second order derivative information around iterate x_{k+1}

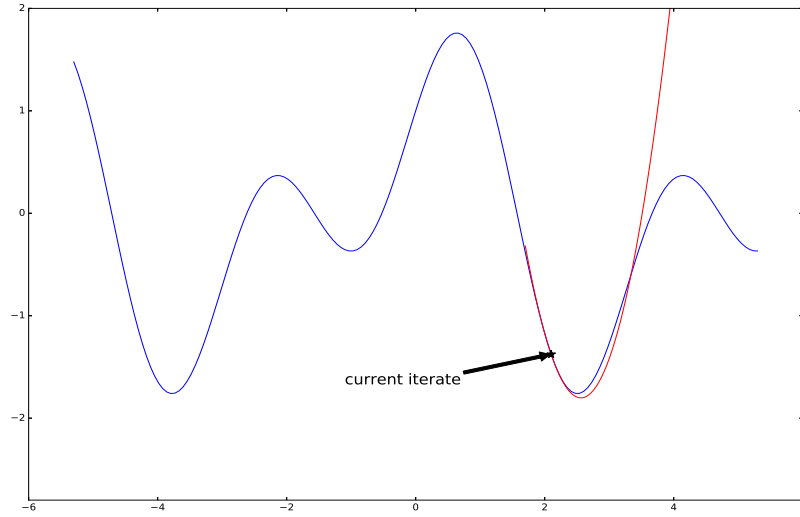


Figure 2.8: Local approximation of an objective function using second order derivative information around iterate x_{k+2}

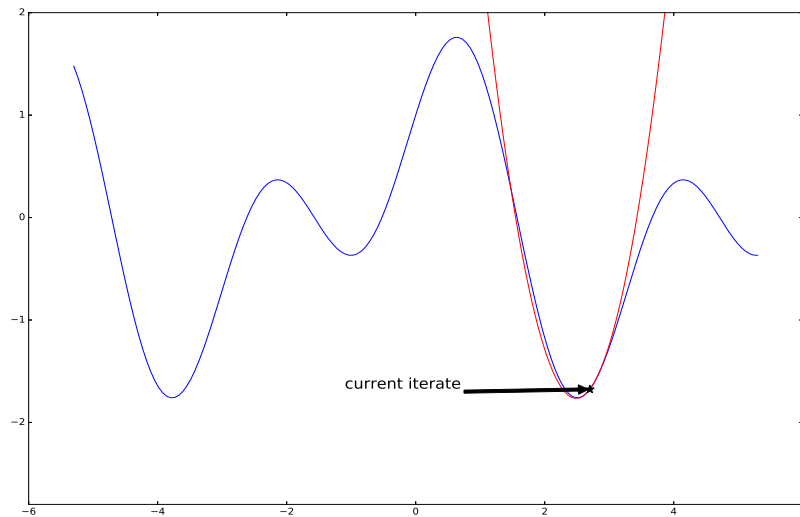


Figure 2.9: Local approximation of an objective function using second order derivative information around iterate x_{k+3}

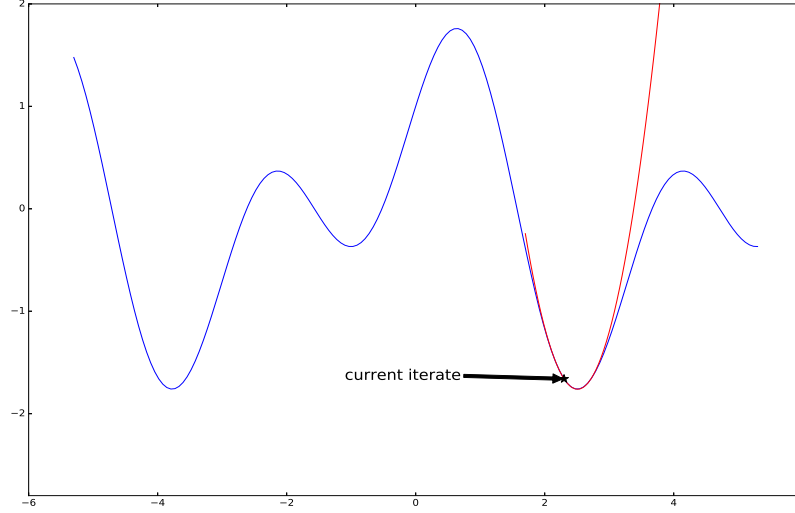


Figure 2.10: Local approximation of an objective function using second order derivative information around iterate x_{k+4}

When the second order derivative of the functional cannot be computed, the first order derivative can still be used as a search direction, with the step-length in this case not assumed to be one (Drábek and Milota, 2013; Fletcher, 2013; Nocedal and Wright, 2006).

Figure 2.12 shows the functional with the first derivative take at two different regions. If the first derivatives of the functional were to be taken as the search directions at these locations, then clearly the search directions will be in the opposite directions as the first derivatives.

Equation 2.15 that gives the parameter (or variable) update then becomes:

$$x_{k+1} = x_k - \alpha f'(x_k) \quad (2.17)$$

From this (equation 2.17), the search direction will be $p_k = -f'(x_k)$, while the step-length, α will have to be computed.

If α is optimally computed, the minimisation process will be carried out using

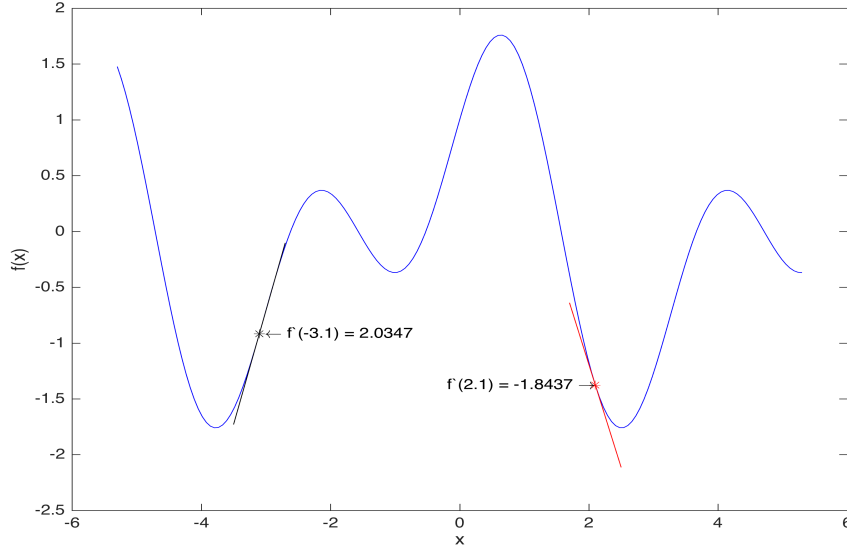


Figure 2.11: Local approximation of the function using first order information. A linear graph has no minimum value hence higher orders series approximations are used

only first derivative information (Figure 2.12). This technique is known as the steepest descent method.

Practically, problems in physics and engineering are multivariable in nature, involving thousands of unknown (and millions in some cases). The ideas developed for the single variable optimization problem can be easily extended to the multivariable problem. In multivariable optimization problems, directional derivatives take the place of the first order derivative of a single variable function (Gelfand et al., 2000; Heinbockel, 2006). The more generic term, gradient, as used in single variable problems, no longer represent the slope of the tangent to the curve. The gradient of a multivariable objective function (Figure 2.13) is a vector that represents the maximum directional derivative of the function (Jost and Li-Jost, 1998; Kreyszig, 2005; Riley et al., 2006).

Mathematically, if $\mathbf{x}$ is a vector of variables x_i for $i \in 1, 2, \dots, n$ and given that

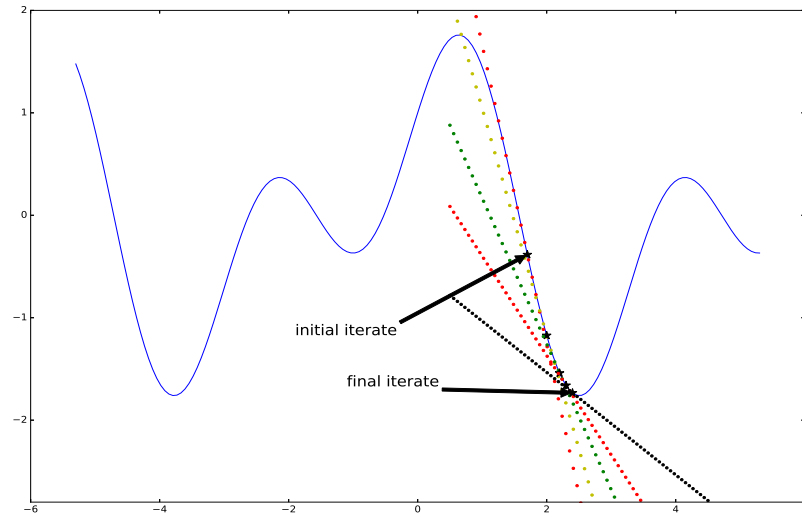


Figure 2.12: Local approximation of an objective function using second order derivative information around iterate x_{k+4}

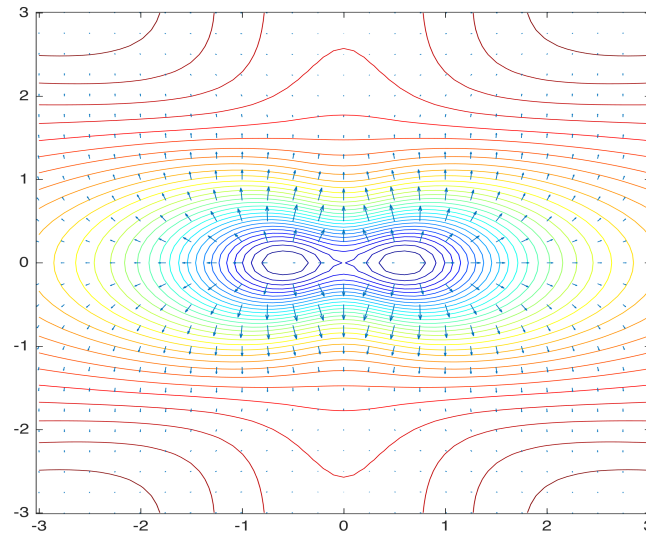


Figure 2.13: Gradient of an objective function $f(x_1, x_2)$. Each arrow points in the direction of maximum ascent from where point at which it is drawn

$f(\mathbf{x})$ is a multivariable objective function, the gradient is given by equation 2.18.

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} \quad (2.18)$$

Similarly, the second derivative for the single variable objective function becomes a matrix for a multivariable objective function. This matrix (equation 2.19) is known as the Hessian matrix($\mathbf{H}$), and it is required that it is positive definite for minimization problems (Nocedal and Wright, 2006).

$$\mathbf{H}(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix} \quad (2.19)$$

The equation for updating the solution then becomes:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha \mathbf{H}^{-1} \nabla f \quad (2.20)$$

and the search direction is thus given by:

$$\mathbf{p}_k = -\mathbf{H}^{-1} \nabla f \quad (2.21)$$

As the number of variables (N) gets considerable large, computing the inverse of the Hessian or its action on the gradient becomes prohibitive. Various approximations are made to either the inverse of the Hessian, or its action on the gradient.

The gradient descent method is the most popular local optimization technique as it does not require the computation of the inverse Hessian matrix or its action on

the gradient. In the gradient descent method, the inverse Hessian matrix is an identity matrix, thus the search direction is taken as the negative of the gradient, i.e. $\mathbf{p}_k = -\nabla f$. Figure 2.14 shows an application of the gradient descent method to compute the minimum value for the Rosenbrock function (Rosenbrock, 1960).

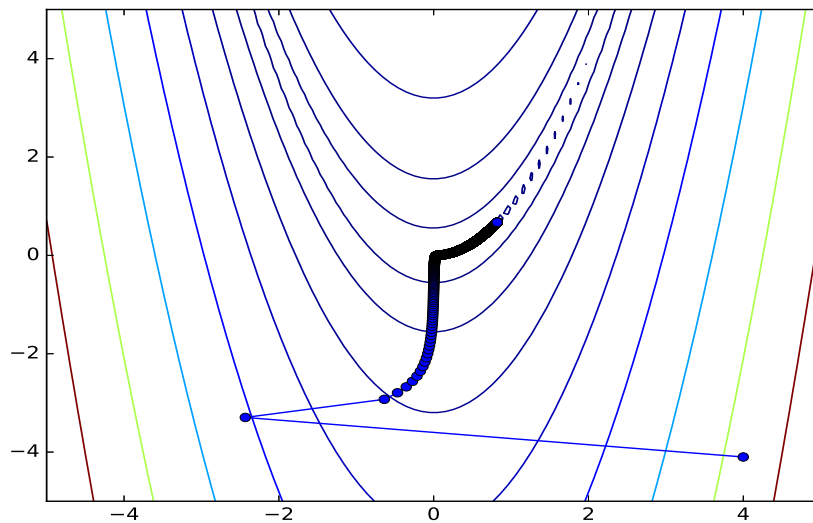


Figure 2.14: Gradient descent method solution for the minimization of the Rosenbrock function

Local optimization and FWI

A solution of an optimization problem, as discussed in the previous subsection, seeks to optimize (maximise or minimise) the value of a functional. In the conventional FWI set-up, the problem is setup as a local optimization problem, where the functional (equation 2.22) we seek to minimise is the L_2 norm of the difference between the observed data and the predicted data (this is also known as the *data residual*)

$$f(\mathbf{m}) = \| \mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{pred}} \|_2 \quad (2.22)$$

f denotes the data misfit expressed as a function of the vector model parameters $\mathbf{m}$. $\mathbf{d}_{\text{obs}}$ is a vector representing the observed field data and $\mathbf{d}_{\text{pred}}$ is a vector for the predicted data. Note that the L_2 norm is computed as a sum of the data residuals over all time steps, for all source and all receivers. The only constraint we place on the model ($\mathbf{m}$) is the natural requirement that it satisfies the wave equation for every iteration step.

Let us represent the wave equation by

$$\mathbf{A}(\mathbf{m}) = s(t)$$

where $\mathbf{A}$ is the wave equation operator, $\mathbf{m}$ is the set of model parameters and s is some source of energy expressed as a function of time t . The complete optimization problem is now expressed as in equation 2.23 (van Leeuwen and Herrmann, 2013)

$$\begin{aligned} \min \quad & f(\mathbf{m}) \\ \text{s.t.} \quad & \mathbf{A}(\mathbf{m}) = s(t) \end{aligned} \quad (2.23)$$

FWI has also been set in some situations as a maximization problem (Luo and Sava, 2011; Van Leeuwen and Mulder, 2010), where the functional is expressed as a cross-correlation (or deconvolution) between the synthetic and the observed

data. In this case, however, the L-2 norm difference-based functional can be used. Subsequent discussions will therefore be along this theory.

2.2 Forward problem in FWI

The problem of computing synthetic data from a model is referred to as a forward problem. In seismic data inversion, the solution to the forward problem is provided by the wave equation. There are various variants of the wave equation. The operator, $\mathbf{A}$ in equation 2.24 is usually a non-linear differential operator. Its form changes, depending on the variant of the wave equation considered.

$$\mathbf{A}(\mathbf{m}) = \mathbf{s}(t) \quad (2.24)$$

The solution to the constant density acoustic wave equation is described here. However, its formalism can be adapted to any of the variants of the wave equation.

2.2.1 Acoustic Wave Propagation

The scalar wave equation (Bleistein et al., 2013; Chapman, 2004; Stolt and Weglein, 2012) is an approximation of the seismic wave equation (Aki and Richards, 2002; Shearer, 2009), and widely applied to many acoustic seismic dataset. It is formulated as equation 2.25 (Gisolf and Vershuur, 2010).

$$\frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} - \nabla^2 p = s(t) \quad (2.25)$$

Where p is some scalar potential, related to the acoustic pressure which is propagated through the medium, and is a function of space and time; $\mathbf{c}$ is the medium velocity, which is a function of space and is assumed to be constant for the duration of the wave propagation; and $s(t)$ is the same forcing function that triggers the wave propagation.

The scalar acoustic wave equation formalism, though restrictive as it ignores the effects of elasticity, attenuation and anomalous density variations (Warner et al., 2012), is the most widely used variant of the wave equation . It has the unique advantages of being computationally less expensive and is better posed for the optimization problem as only the dominant parameter (P-wave velocity) is involved in the inversion (Virieux and Operto, 2009).

In realistic earth model, the velocity distribution c , has no analytic representation, hence the solution to equation 2.25 relies on numerical methods (Gockenbach, 2010; Morton and Mayers, 1998; Smith, 1985). The common numerical methods used in seismic forward modeling include finite-element methods, finite-difference methods and pseudospectral methods. The finite difference method is considered the classical and most common in seismic forward modelling, and is the formalism used in this thesis.

2.2.2 Finite Difference Solution

The finite difference method represents a continuous function by its values at discrete points, and approximates derivatives using differences (Gisolf and Vershuur, 2010; Jovanović and Süli, 2013; Smith, 1985). Thus,

$$p(x, t) = p(x_i, t_n)$$

and

$$c(x) = c(x_i)$$

where $p(x_i, t_n)$ and $c(x_i)$ are the discrete forms of the continuous variables $p(x, t)$ and $c(x)$. Similarly, an analytical spatial derivative is approximated as

$$\frac{\partial^n p}{\partial x^n} = \frac{1}{h}(a_0 p(x_i, t_n) + a_1 p(x_{i\pm 1}, t_n) + a_2 p(x_{i\pm 2}, t_n) + \dots)$$

where h is separation between adjacent discrete points, and a_i is a scaling factor for the discrete point x_i . With these discretizations, the analytical scalar wave equation then reduces to an algebraic equation

In the conventional finite difference method for the scalar wave equation, the model through which the seismic waves propagate, is discretised using rectangular grids. Each discrete point is then called a node or a grid point. Wavefields are thus, discretized and only available at these grid points, where the model values have been defined (Figure 2.15). The wavefields, being time dependent, are also discretised in time.

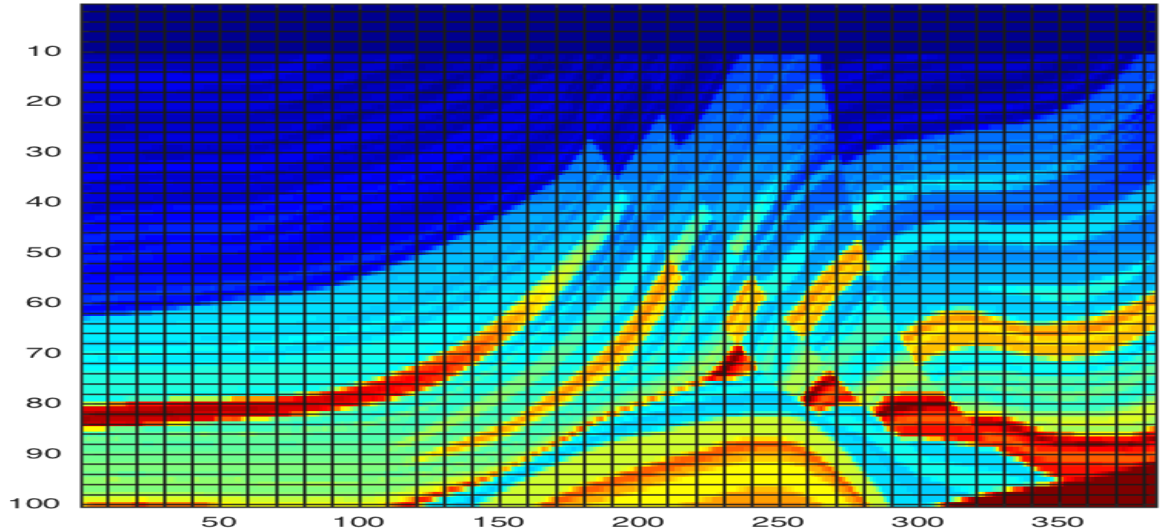


Figure 2.15: A discretized Marmousi model. The wavefields computed in the model are only available at the grid points seen in this model

The solution proceeds by evaluating the spatial derivatives (Laplacian) of the wavefields ($\nabla^2 p$) at a grid point, x_i using neighbouring values of the wavefields (Figure 2.16). A general expression for the spatial derivative of the wave equation in one dimension is given by equation 2.26.

$$\left. \frac{\partial^2 p^t}{\partial x^2} \right|_{i,j} \approx \frac{1}{dx^2} \left[c_0 p_{i,j}^t + \sum_{n=1}^N c_n (p_{i-n,j}^t + p_{i+n,j}^t) \right] \quad (2.26)$$

where c_n are the finite-difference coefficients, and have the property that their sum is zero(0). The upper limit of the summation term in Equation 2.26 determines the order of the finite difference stencil (geometric arrangement of the neighbouring grid points used in the evaluation of a spatial derivative). When the upper limit (N) is 2, the finite difference stencil is referred to a 4th-order.

The accuracy of the finite difference solution can be improved using compact finite difference scheme (Figure 2.17) or a higher order stencil. The higher the order of the stencil, the more accurate the results. However, additional computational costs are also introduced. Higher order stencils have been studied by many authors (Etgen and OBrien, 2007; Liu and Sen, 2009; Moczo et al., 2007).

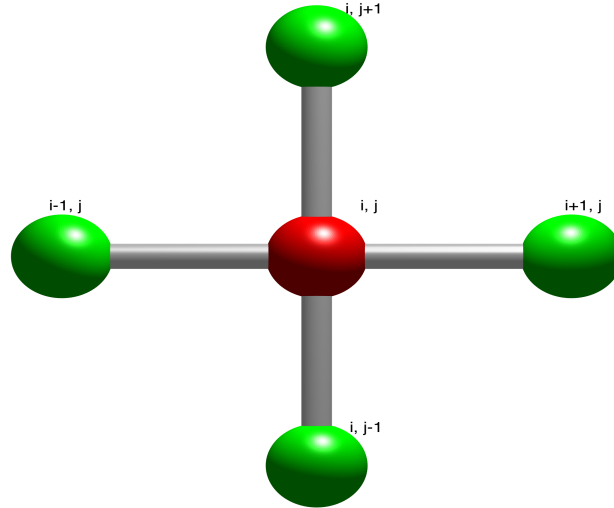


Figure 2.16: A standard 4th order stencil for computing the Laplacian of the wave equation. The node represented by the **red** sphere represents the grid point at which where the Laplacian is currently evaluated. All other nodes (**green** spheres), together with red node, contribute to the evaluation of the Laplacian.

Similarly, a 2nd-order finite difference solution for the temporal derivatives is given by equation 2.27. Higher order solutions for the temporal derivatives exist (Chen, 2007, 2011; Dablain, 1986; Long et al., 2013). In general, higher-order

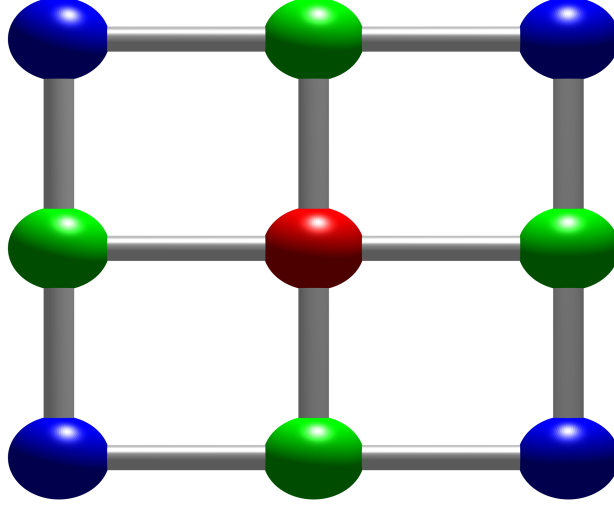


Figure 2.17: A compact 2D finite difference stencil. This stencil yields a more accurate solution to the wave equation as it incorporates additional neighbouring grid points

finite difference evaluation of the temporal derivatives demand large computer resources and are generally unstable (Etgen and OBrien, 2007).

$$\left. \frac{\partial^2 p}{\partial t^2} \right|_t \approx \frac{1}{dt^2} [-2p_{i,j}^t + (p_{i,j}^{t-1} + p_{i,j}^{t+1})] \quad (2.27)$$

Injecting equations 2.26 and 2.27 into equation 2.25 yields equation 2.28.

$$\begin{aligned} & \frac{1}{cdt^2} [-2p_{i,j}^t + (p_{i,j}^{t-1} + p_{i,j}^{t+1})] \\ & - \frac{1}{dx^2} \left[c_0 p_{i,j}^t + \sum_{n=1}^N c_n (p_{i-n,j}^t + p_{i+n,j}^t + p_{i,j-n}^t + p_{i,j+n}^t) \right] \\ & = s(t) \end{aligned} \quad (2.28)$$

Re-arranging the terms of equation 2.28 results in the final expression for the

solution (equation 2.29).

$$p_{i,j}^{t+1} = 2p_{i,j}^t - p_{i,j}^{t-1} + \frac{cdt^2}{dx} \left[c_0 p_{i,j}^t + \sum_{n=1}^N c_n (p_{i-n,j}^t + p_{i+n,j}^t + p_{i,j-n}^t + p_{i,j+n}^t) \right] + s(t) \quad (2.29)$$

2.2.3 Practicalities of applying finite difference method

The theory of the finite difference method presents the solution to the wave equation in a straight forward approach. However, like any numerical method there are certain considerations to be taken into account when implementing the method. Some of these considerations include stability, grid dispersion, source implementation, absorbing boundary implementation, free surface implementation, etc

Stability

The level of discretisation, measured in terms of the grid size and the time step affects the stability and accuracy of the finite difference solution. The stability of the solution is controlled by the maximum propagation velocity in the model, the size of the grid and the time step. In regular grids, with equal grid spacing in all directions, it is chosen such that the wavefields do not propagate more than a fraction of the grid for a particular time step. An expression for ensuring the stability of the finite difference solution is given by equation 2.30 and 2.31 (Lines et al., n.d.).

$$\frac{dt * v_{max}}{dx} \leq r \quad (2.30)$$

and

$$r = \sqrt{\frac{a_1}{a_2}} \quad (2.31)$$

Equations 2.30 and 2.31 controls the maximum velocity that can be propagated for a particular implementation of a finite difference method, for a specific grid size and time step.

Grid Dispersion

Grid dispersion is a numerical error in finite difference solution when different frequencies in the source spectrum propagate at difference velocities. It is usually attributed to a situation where the model is sparsely sampled compared to the lowest wavelength present in the source. Every finite difference implementation has a minimum sampling requirement for the smallest wavelength present in the source wavelet. This is sometimes referred to as the minimum number of grid points per smallest wavelength allowed by the finite difference implementation. This number varies according to the order of the finite difference. It is about 5 for the fourth-order in space and second-order in time finite difference implementation. For higher order finite difference implementations, the number decreases to about 2.5 in some cases.

The minimum wavelength present is given by

$$\lambda_{min} = \frac{v_{min}}{f_{max}}$$

where v_{min} is the minimum velocity present in the velocity model, and f_{max} is the maximum frequency in the source spectrum. Hence, the number of grid points per wavelength determines the grid size of the finite difference given by equation 2.32.

$$dx \leq \frac{\lambda_{min}}{N} \tag{2.32}$$

where N is the minimum number of grid points by wavelength allowed by the particular finite difference implementation.

2.3 Inverse problem in FWI

2.3.1 Gradient computation in Conventional FWI

We can re-write equation 2.22 as equation 2.33, a compact form.

$$f(\mathbf{m}) = \frac{1}{2} \delta \mathbf{d}^\dagger \delta \mathbf{d} \quad (2.33)$$

where $\delta \mathbf{d}$ is the data residual, and is a function of the model parameters. Equation 2.33 expresses the objective function as a function of the model parameters. When solved as an iterative local optimization problem, the model update is expressed as

$$\mathbf{m} = \mathbf{m}_0 + \alpha \delta \mathbf{m} \quad (2.34)$$

where $\delta \mathbf{m}$ is the search direction, and α is the step-length. The search direction is given by:

$$\delta \mathbf{m} = -\mathbf{H}^{-1} \nabla_{\mathbf{m}} f$$

where $\nabla_{\mathbf{m}}$ and $-\mathbf{H}^{-1}$ are evaluated at $\mathbf{m}_0$.

The gradient of the objective function is:

$$\begin{aligned} \nabla_{\mathbf{m}} f &= \frac{\partial}{\partial \mathbf{m}} \left(\frac{1}{2} \delta \mathbf{d}^\dagger \delta \mathbf{d} \right) \\ &= \left(\frac{\partial \delta \mathbf{d}}{\partial \mathbf{m}} \right)^\dagger \delta \mathbf{d} \\ &= \left(\frac{\partial \mathbf{d}_{\text{pred}}}{\partial \mathbf{m}} \right)^\dagger \delta \mathbf{d} \\ &= \mathbf{J}^\dagger \delta \mathbf{d} \end{aligned} \quad (2.35)$$

$\mathbf{J}$ is the sensitivity or Fréchet derivative matrix.

In typical seismic inversion problems, with large model parameter sets, the Fréchet matrix is expensive to compute. The functional, however, depends on the model parameters through state variables ($\mathbf{p}$), where

$$\mathbf{p} = \mathbf{G}(\mathbf{m})$$

This allows the gradient of the functional to be computed using adjoint state techniques (Plessix, 2006).

An alternative derivation of the gradient of the objective function, with a similar result as that of the adjoint state technique, begins by expressing the observed data as

$$\mathbf{d} = \mathbf{D}\mathbf{p}$$

where $\mathbf{D}$ is some picking matrix that extracts the modelled wavefields ($\mathbf{p}$) at receiver locations. The modelled wavefields have a linear relationship with the source wavelets

$$\mathbf{A}\mathbf{p} = \mathbf{s} \tag{2.36}$$

where

$$A = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2$$

We then take the derivative of equation 2.36 with respect to the model parameter $\mathbf{m}$:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{m}}(\mathbf{A}\mathbf{p}) &= \frac{\partial \mathbf{s}}{\partial \mathbf{m}} \\ \frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{p} + \mathbf{A} \frac{\partial \mathbf{p}}{\partial \mathbf{m}} &= \mathbf{0} \end{aligned} \tag{2.37}$$

Equation 2.37 is equal to $\mathbf{0}$ as the source function is independent of the model parameters. Re-arranging equation 2.37 gives equation 2.38

$$\frac{\partial \mathbf{p}}{\partial \mathbf{m}} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \mathbf{p} \quad (2.38)$$

Observe that $\frac{\partial \mathbf{d}}{\partial \mathbf{m}} = \mathbf{D} \frac{\partial \mathbf{p}}{\partial \mathbf{m}}$.

Equation 2.38 can then be substituted into equation 2.35 to derive an expression for the gradient computation (equation 2.39).

$$\begin{aligned} \nabla_{\mathbf{m}} F &= \left(\mathbf{D} \frac{\partial \mathbf{p}}{\partial \mathbf{m}} \right)^\dagger \delta \mathbf{d} \\ &= \left(-\mathbf{D} \mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \mathbf{p} \right)^\dagger \delta \mathbf{d} \\ &= -\mathbf{p}^T \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^T (\mathbf{A}^{-1})^T \delta \mathbf{d} \end{aligned} \quad (2.39)$$

Equation (2.39) can be re-writtten as

$$\nabla_{\mathbf{m}} F = -\mathbf{p}^T \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^T \delta \mathbf{p} \quad (2.40)$$

where $\delta \mathbf{p} = (\mathbf{A}^{-1})^T \delta \mathbf{d}$. $\mathbf{p}$ is interpreted as the wavefield generated by propagating a source function, $\mathbf{s}$ forward in time via the modelling operator $\mathbf{A}$. $\delta \mathbf{p}$ is interpreted as the wavefield generated by propagating the data residual, $\delta \mathbf{d}$ backward in time via the adjoint operator $\mathbf{A}^T$.

2.3.2 Resolution of the FWI gradients

The resolution of the FWI operator determines the wavenumber spectrum present in the model update vector. The local wavenumber is controlled by the local

angular aperture, θ , the local velocity, c , and the local angular frequency, ω , (Figure 2.18) via equation 2.41 (Alkhalifah, 2016).

$$\mathbf{k}_m = \mathbf{k}_s + \mathbf{k}_r = 2\left(\frac{2\pi}{\lambda}\right) \cos\left(\frac{\theta}{2}\right) \mathbf{n} \quad (2.41)$$

where

$$\lambda = \frac{c}{\omega} \quad (2.42)$$

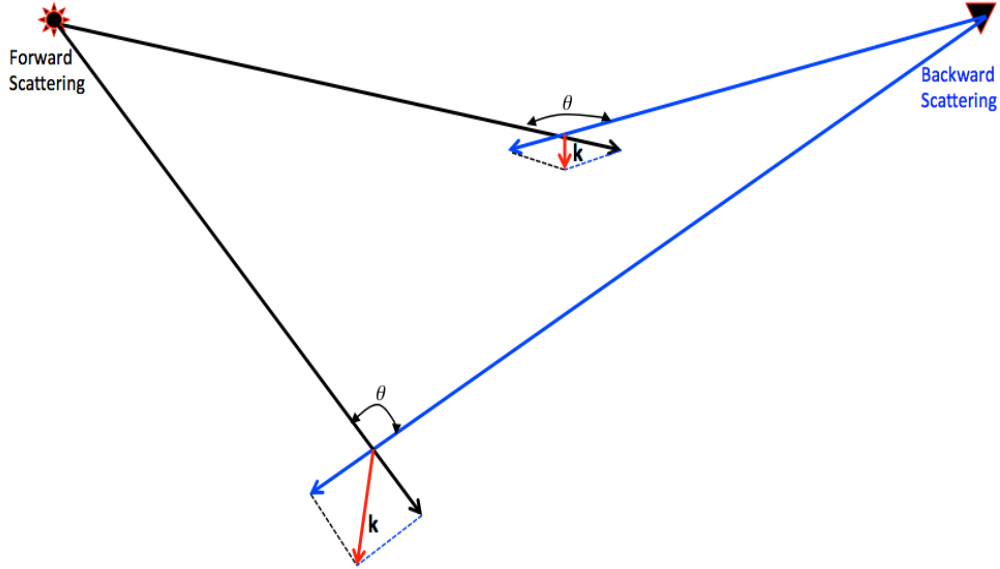


Figure 2.18: The wavenumber spectrum presents in the gradient has dependency on the local angular aperture (θ), the local velocity (c) and the local angular frequency (ω). Large angular apertures give rise to low wavenumber updates

2.4 Theory of Reflection FWI

The model of the subsurface can be decomposed into a series of wavenumbers, ranging from zero to a maximum (known as the Nyquist wavenumber component), determined by the sampling interval of the model. Wavenumber components close

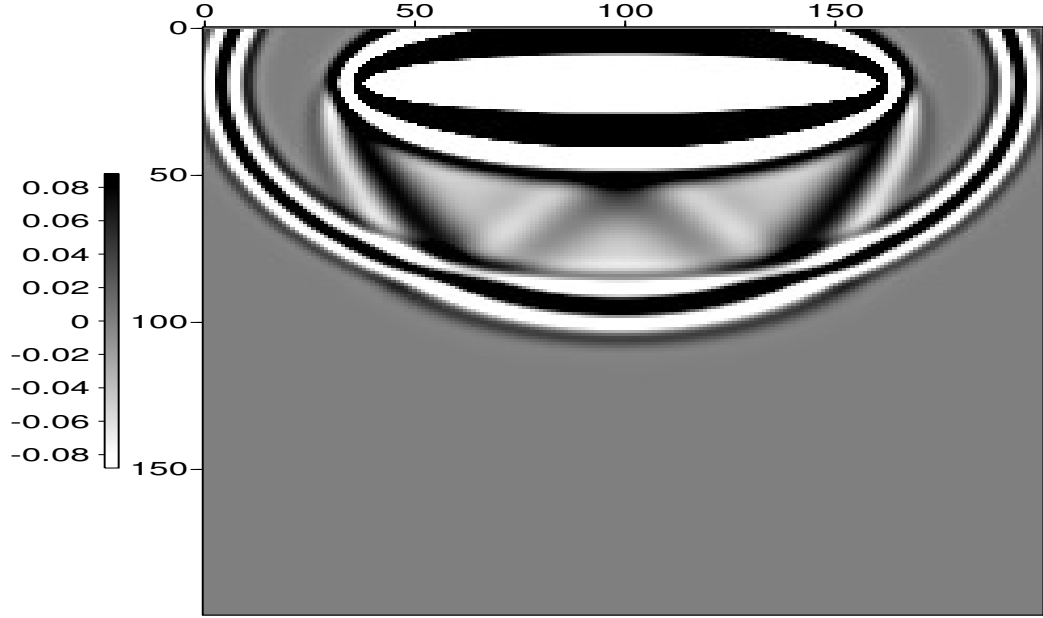


Figure 2.19: Conventional FWI gradients are shown here. The gradient points in the direction in which the model is updated

to zero represent the slowly varying components of the model, and are referred to as low wavenumber (long wavelength) components. Higher wavenumbers represent highly oscillatory (short wavelength) components of the model. The ability of FWI to recover the varying wavenumbers present in the model is dependent on the migration operator. Therefore, this recovered wavenumber is the time recorded wavelengths mapped to depth by the local velocity and the aperture (Xu, Wang, Chen, Zhang and Lambare, 2012), and is given by equation 2.43 (Alkhalifah, 2016)

$$\mathbf{k} = 2 \frac{\omega}{v} \cos \frac{\theta}{2} \mathbf{n} \quad (2.43)$$

where $\mathbf{k}$ is the recovered wavenumber, ω the angular frequency, v the local velocity and θ , the local angular aperture. Since the desired preference in FWI is to recover the local wavenumber components of the model, prior to recovering the high wavenumber components, it is therefore necessary to have a richer supply of low frequencies and a wide aperture. This is in order to recover the low wavenumber

components of the model, as given by equation 2.43.

Transmission and refraction datasets have large local angular aperture that allows for the recovery of large wavenumber components of the model (Gauthier et al., 1986). However, in reflection dominated datasets with short offsets and lacking in low frequencies due to the presence of source and receiver ghost (Lindsey, 1960), only high wavenumbers are recovered as a results of the limited range of the local angular aperture values. A work to the problem is to have scatterers present in the model. These will act as secondary sources of energies and help in generating low wavenumber components of the model by fulfilling the condition of having large local angular aperture.

Subsequent subsections discuss the general theory of reflection FWI. The cartoons shown in Figures 2.20 through 2.32 help illustrate the formation and nature of reflection FWI updates.

2.4.1 Reflection FWI Gradient in a Smooth Model

In a smooth model with no scatterers, the forward wavefields propagate only downwards (ignoring the horizontal directions) in the forward time (Figure 2.20). The backward wavefields also propagate downwards in the forward time (Figure 2.21). This is equivalent to saying the backward wavefields propagate upwards if time is reversed (Figure 2.22). If these energies are cross-correlated, the result is a migration-like update, as a result of the small aperture angle between the source wavefields and the receiver wavefields (Figure 2.33). This simplified experiment shows that the first iteration update of FWI in a smooth velocity model is a migration-like update.

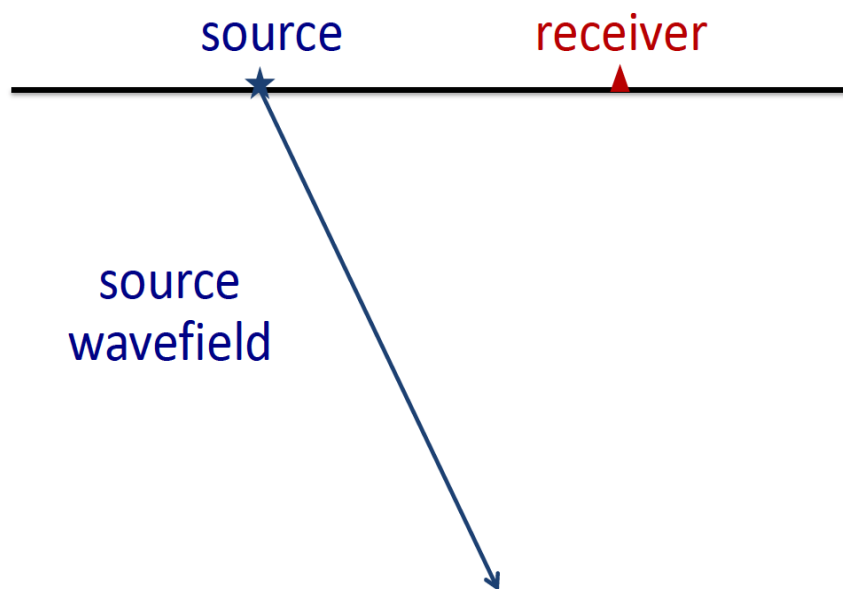


Figure 2.20: Forward wavefields propagating through a smooth model in forward time. The smoothness of the model and the lack of scatterers in it makes the wavefields travel only downwards

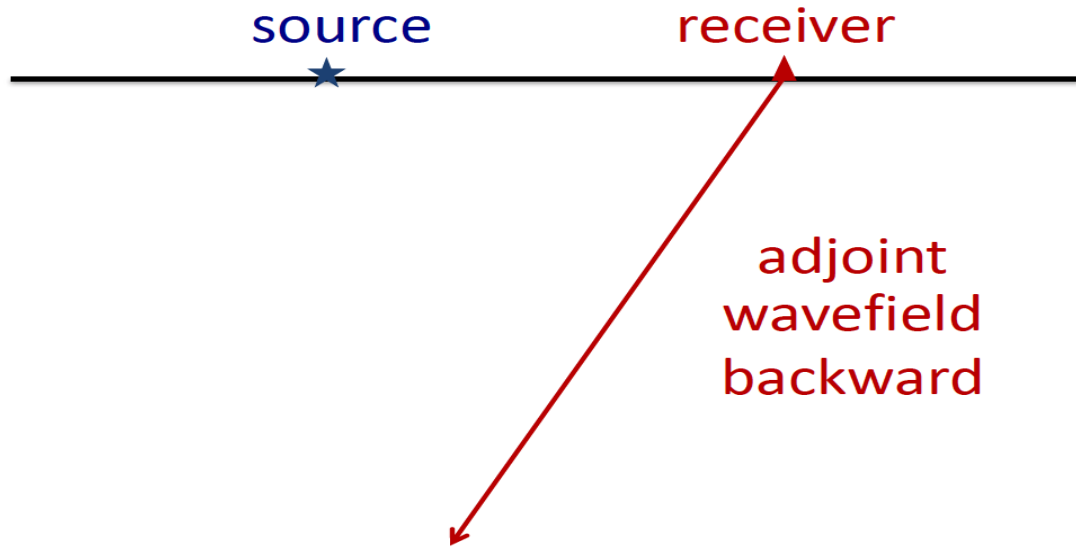


Figure 2.21: Backward wavefields propagating through a smooth model in backward time. Absence of scatterers also allows the energies propagate only downwards

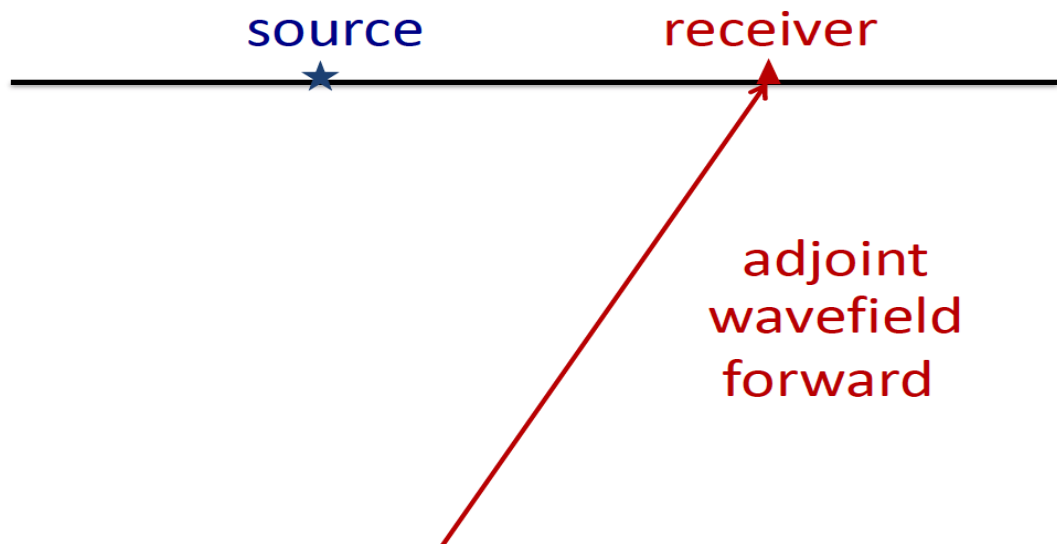


Figure 2.22: Backward wavefields propagating through a smooth model in forward time. The direction of propagation has been reversed by the reversal in the temporal direction

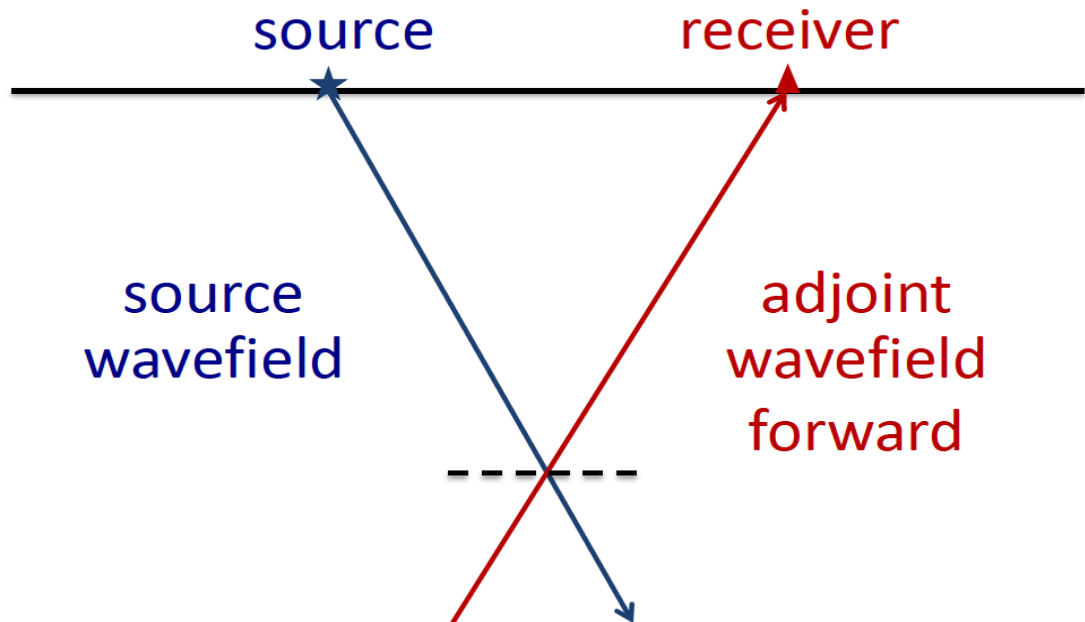


Figure 2.23: Migration update in a smooth model formed by cross-correlating the source wavefield in forward time and the receiver wavefield in forward time. The migration attempts to build scatterers in the model by forming the update around the subsurface location that acts as a virtual source of energy

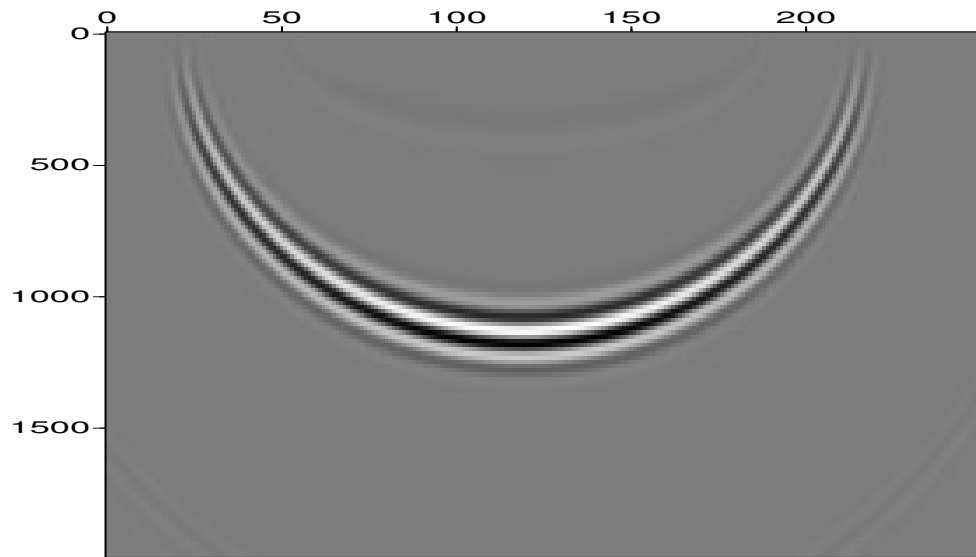


Figure 2.24: Migration update in a smooth model formed by cross-correlating the source wavefield in forward time and the receiver wavefield in forward time. The migration attempts to build scatterers in the model by forming the update around the subsurface location that acts as a virtual source of energy

2.4.2 Reflection FWI Gradient in a Rough Model

In a rough model with scatterers about the same scale as the dominant wavelength in the source (Figure 2.25), the source (forward) wavefields propagate in both the downward and the upward directions (Figure 2.26). If the amplitude of the downward component of the source wavefield is 1, then amplitude of the upward component is then R , where R is the reflection coefficient of the scatterer. The receiver wavefields (in backward time) are shown in Figure 2.27. If the receiver wavefields were to propagate in forward time, the path is shown in Figure 2.28 . The downward component of the receiver wavefields has an amplitude of R , while the upward component has an amplitude of R^2

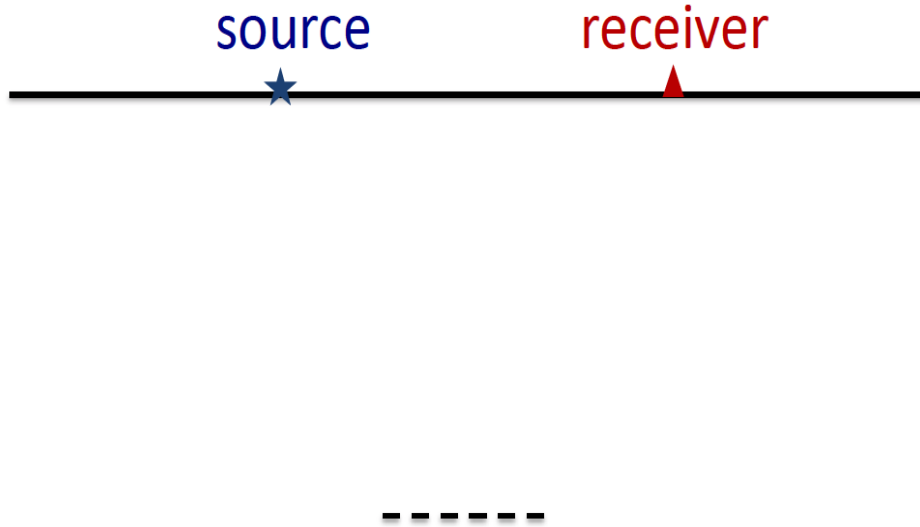


Figure 2.25: A cartoon showing a rough model (with scatterers having wavenumbers as to match the dominant wavelength of the source and thus generate reflections)

The source wavefields are then cross-correlated with the receiver (with the same temporal orientation). The update has two components: a migration component

(Figure 2.30) formed in areas where the source and the receiver wavefields propagate in opposite directions (for the same temporal orientation), hence, a small aperture angle; and a tomography component (Figure 2.31) formed in areas where the source and the receiver wavefields propagate in the same direction (for the same temporal orientation) with a large aperture angle. The sum of both updates are shown in Figure 2.32. The tomographic updates have an amplitude of R^2 , while the migration update has an amplitude of R . Hence, the tomographic updates are weaker than the migration updates by a factor of R .

This realisation that 'weak' tomographic updates are present in FWI of reflection dominated dataset imposes a scheme, where the tomographic and migration updates are alternately built. This alternate "migration-demigration" process forms the foundation on which reflection FWI is built. The fundamental differences among the various reflection FWI flavours lie in the differences in how they compute the migration component of the update. The tomographic update is usually built using conventional FWI techniques.

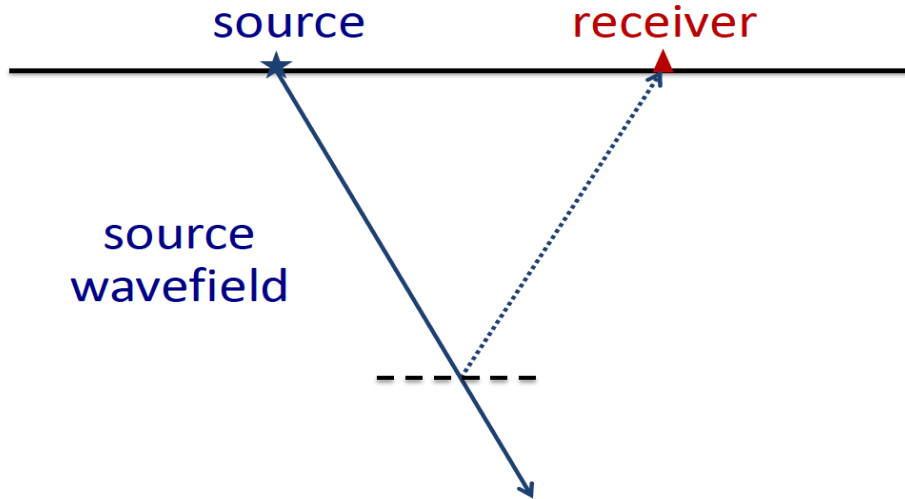


Figure 2.26: Source wavefields through a rough velocity model. The solid line represents direct wavefields from the source (propagating downwards in forward time), while the dashed line represents the scattered wavefield (propagating upwards in forward time)

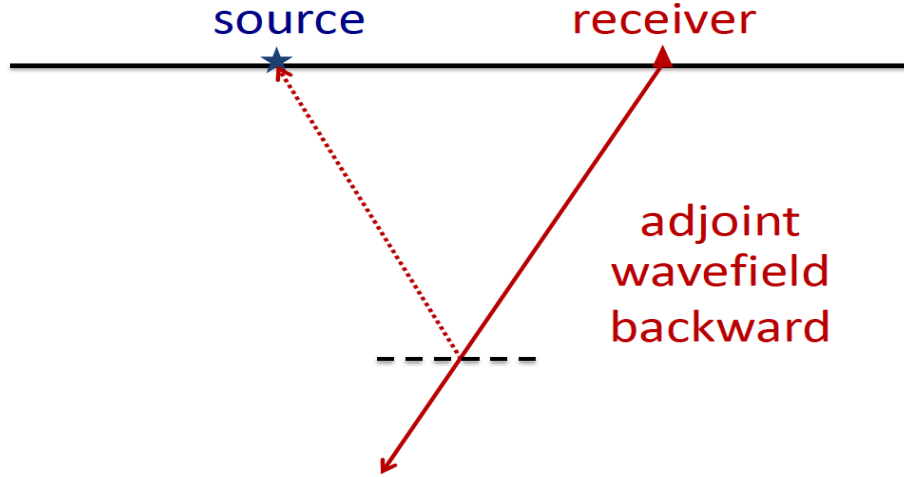


Figure 2.27: Receiver wavefields in a rough velocity model in backward time. The solid line represents direct wavefields from the receiver (propagating downwards in backward time), while the dashed line represents the scattered wavefield (going upwards in backward time)

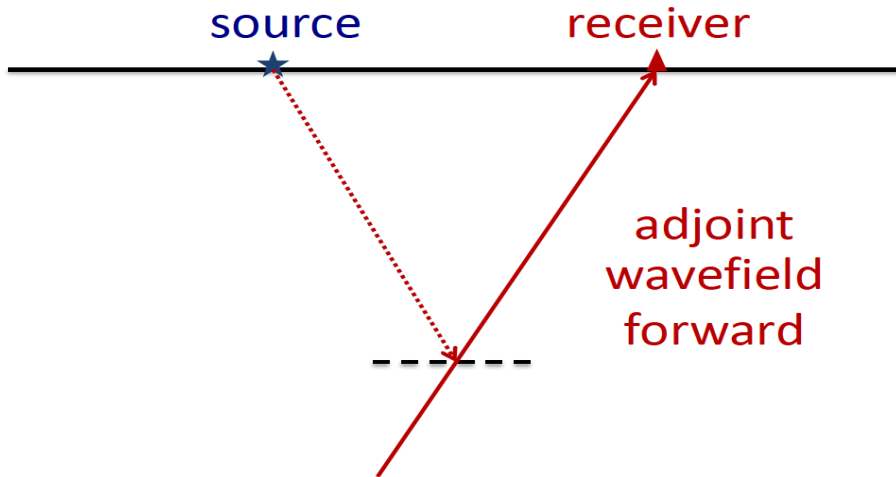


Figure 2.28: Receiver wavefields in a rough velocity model in forward time. The time reversal means that the direct receiver wavefields propagate upwards in forward time, while the scattered receiver wavefields propagate downwards in forward time

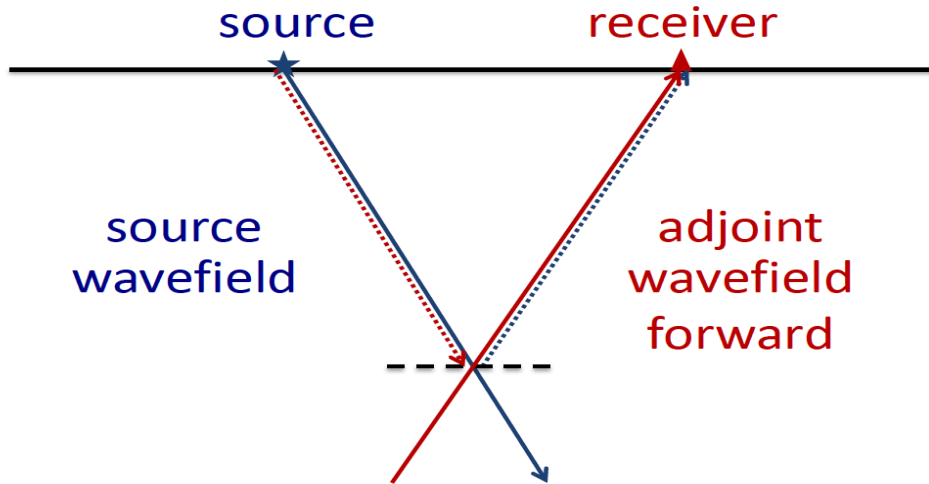


Figure 2.29: Reflection FWI gradient formed by cross-correlating the source wavefield in forward time with the adjoint wavefields in forward time

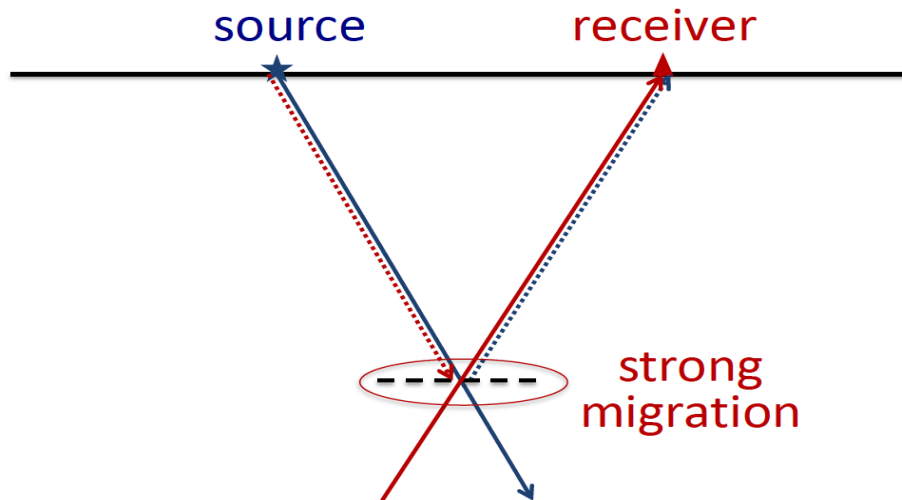


Figure 2.30: Migration component of FWI gradient formed by source and adjoint wavefields travelling in the opposite direction in forward time

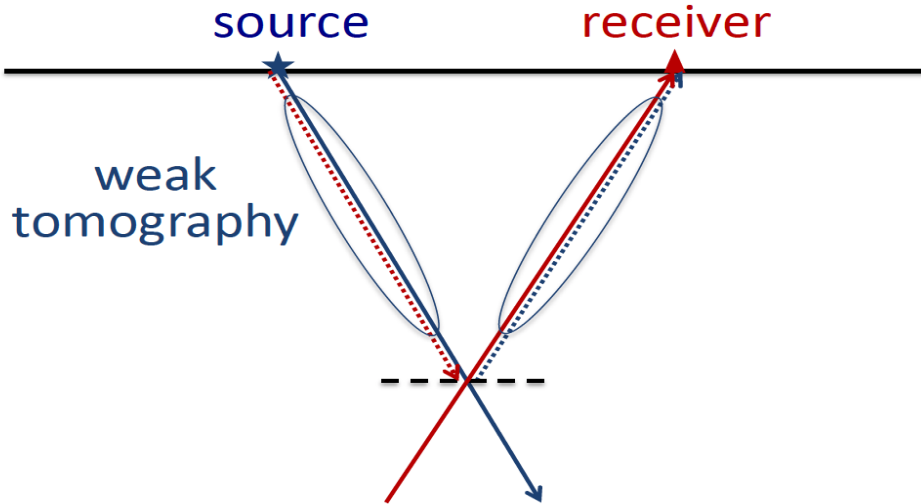


Figure 2.31: Tomography update in RWI formed by source and adjoint wavefields propagating in same direction in forward time.

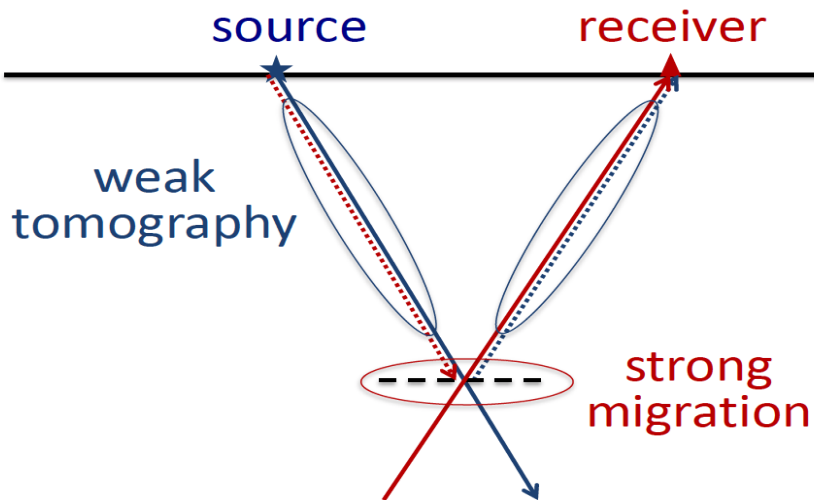


Figure 2.32: The total reflection FWI update. notice that the migration component has a much stronger amplitude compared to the tomography updates, with much less amplitude

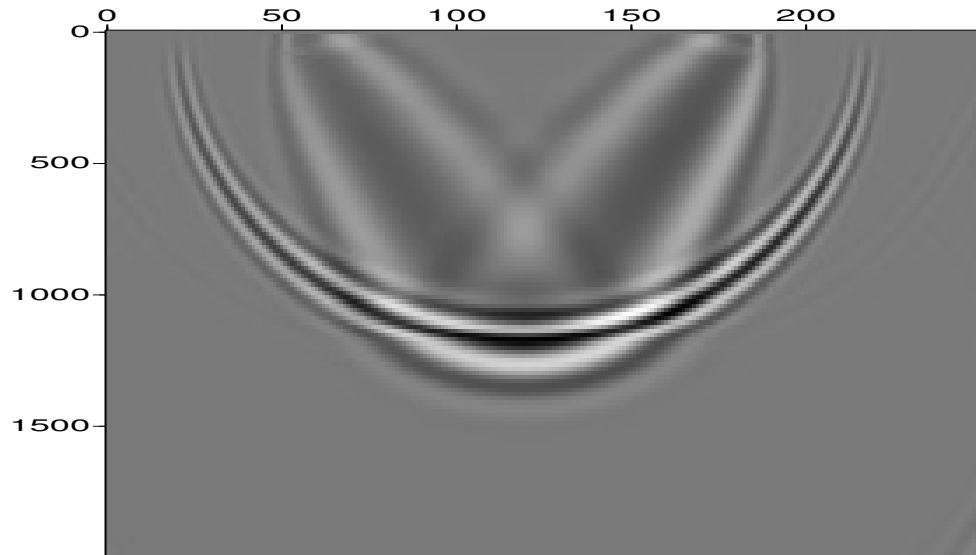


Figure 2.33: The total reflection FWI update. notice that the migration component has a much stronger amplitude compared to the tomography updates, with much less amplitude

2.5 Conclusion

FWI is a multiply-iterated linearised inversion, where we start off from an initial model and iteratively improve the model via some gradient decent method. This local optimization technique involve the computation of the forward wavefields in every FWI iteration. The resolution of FWI is dictated by the migration operator which depends, amongst other things, on the local angular aperture. In transmission and refraction dominated dataset, the large angular aperture entails that low wavenumber components of the model are recovered in conventional FWI. For reflection dominated datasets, the limited local angular aperture makes the recovery of low wavenumbers prohibitive for conventional FWI. This is because the large amplitudes of the high wavenumbers (migration updates), compared to the weak amplitudes of the low wavenumbers (tomographic updates), get the process trapped in a local minimal, where we continually improve on the values of the high wavenumbers.

Reflection-based waveform inversion (RWI) includes all the various methods used to suppress the migration component of conventional FWI for reflection datasets, thus improving the strength of the tomographic component. It is usually performed by executing the inversion in an alternating fashion similar to migration based traveltimes tomography (MBTT) ([Chavent and Clément, 1993](#); [Chavent et al., 1994](#)). In the first step, we build some reflectivity (or high wavenumber component) from an initial smooth model, while in the second step we perform some form of demigration to update the background model ([Chi et al., 2015](#); [Ma and Hale, 2013](#); [Mora, 1989](#); [Tang et al., 2013](#); [Wang, Chen, Zhang, Shen et al., 2013](#); [Wang and Rao, 2009](#); [Xu, Chen, Lambaré, Zhang, Wang et al., 2012](#); [Yao et al., 2014](#)). The various RWI flavours employs different methods of building the tomographic and migration components, but all seem to depend on the alternating inversion scheme.

Chapter 3

Finite Difference Explicit Wavefield Decomposition

The conventional implementation of FWI, as noted in chapter one, fails when the target is deep and the data is dominated (predominantly) by sub-critical angle reflection energies. Many authors have shown that a "migration/demigration" process, which alternates between building a tomographic component and a migration component of the gradient, can help improve the performance of FWI in this scenario. The implementation of the migration/demigration technique, however, differ with authors. Some authors ([Xu, Chen, Lambaré, Zhang, Wang et al., 2012](#)) build the migration component in the reflectivity (migration image) domain, and the tomographic component in the velocity domain. This method requires that the migration operator be true amplitude and can be a source of errors when this condition is not met. Some other authors build the tomographic component with travel time techniques ([Ma and Hale, 2013](#)), while the migration component is build with FWI. Others ([Wang, Chauris, Donno and Calandra, 2013](#)) split the forward and backward wavefields into up-going and down-going components, and the combine energy travelling in the same direction in time to build the tomographic component, while energies travelling in opposite directions in time build the migration component. This approach has the advantage of solving a single equation and having the tomographic and migration component to

3.1 Theory of Finite Difference Explicit Wavefield Decomposition

be of comparable magnitudes.

In this thesis, a similar approach as the last group of authors is followed, by explicitly splitting the forward and backward wavefields, but in this case, into four components. An up-going wavefield, a down-going wavefield, a left-going wavefield and a right-going wavefield. Similar directions in time are then combined to build a tomographic component, while opposite directions in time are combined to build a migration component. While previous methods, in most cases, end up with four gradients, in this approach we end up with eight gradients. These gradients can be systematically combined to improve the performance of FWI.

3.1 Theory of Finite Difference Explicit Wavefield Decomposition

3.1.1 Fourier Wavefield Decomposition

Fourier decomposition involves taking a Fourier transform of the wavefields and separation of the energies according to propagation directions (Figure 3.1).

In a 2D seismic experiment, if $P(x, z, t)$ represent the total wavefields for every point in the model (x, z) and for all times t . It is possible to define a 2D Fourier transform on the depth-time (z - t) slices of the wavefield cube as given by equation 3.1.

$$P(x, k_z, \omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(x, z, t) \exp[-j2\pi(\omega t + k_z z)] dz dt \quad (3.1)$$

$P(x, k_z, \omega)$ represents the 2D Fourier transform, which is a 2D complex number space (Bracewell and Bracewell, 1986). The basis functions $\exp[-j2\pi(\omega t + k_z z)]$ represents sinusoidal waveforms of different scales (or wavelengths) and orientations (Yilmaz, 2001). Separation of the wavefields according to orientations,

3.1 Theory of Finite Difference Explicit Wavefield Decomposition

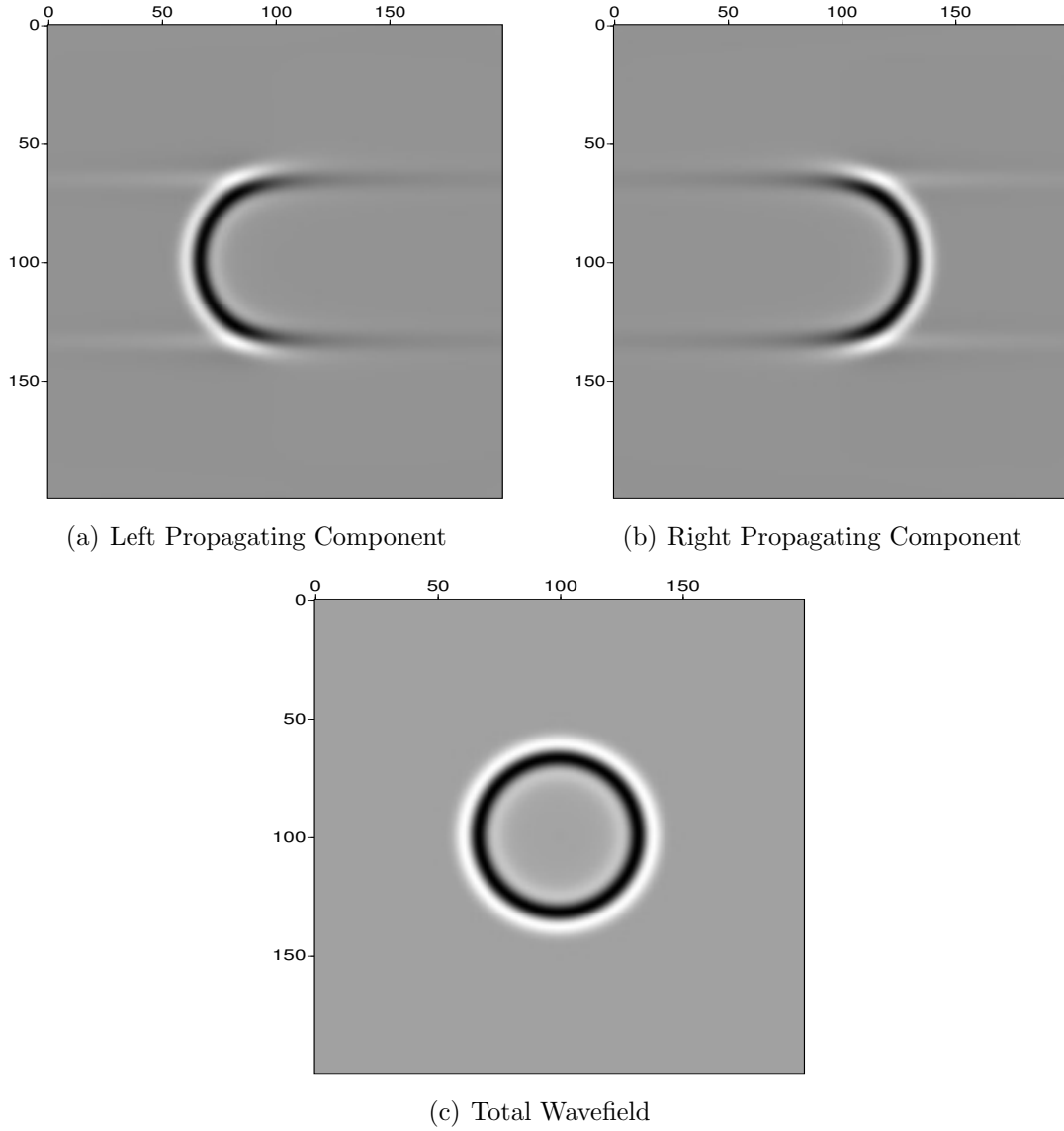


Figure 3.1: Wavefield decomposition via Fourier transform. The wavefields are decomposed into left and right going waves. However, energies from "standing" waves appear as noise (horizontal lines) and are not desirable for applications in FWI and RTM

3.1 Theory of Finite Difference Explicit Wavefield Decomposition

particularly a separation into upward and downward propagating components in the (ω, k_z) domain is given as equations 3.2 and 3.3 (Hu and McMechan, 1987).

$$P_D(x, k_z, \omega) = \begin{cases} P(x, k_z, \omega) & \text{if } \omega k_z \geq 0 \\ 0 & \text{if } \omega k_z < 0 \end{cases} \quad (3.2)$$

$$P_U(x, k_z, \omega) = \begin{cases} P(x, k_z, \omega) & \text{if } \omega k_z \leq 0 \\ 0 & \text{if } \omega k_z > 0 \end{cases} \quad (3.3)$$

$P_D(x, k_z, \omega)$ represent the downwards component of the wavefields in the (ω, k_z) domain, while $P_U(x, k_z, \omega)$ represent the upwards component of the wavefields in the (ω, k_z) domain.

We can now make an inverse Fourier transform back to the space-time domain such as in equations 3.4 and 3.5.

$$P_D(x, z, t) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_D(x, k_z, \omega) \exp [+j2\pi(\omega t + k_z z)] dk_z d\omega \quad (3.4)$$

$$P_U(x, z, t) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_U(x, k_z, \omega) \exp [+j2\pi(\omega t + k_z z)] dk_z d\omega \quad (3.5)$$

$P_U(x, z, t)$ and $P_D(x, z, t)$ respectively represent the upward and the downward components of the original wavefields in the space-time domain. The large errors introduced by energies propagating at right angles to the direction of transformation cause the sum of the two component wavefields to be unequal to the original wavefield (equation 3.6).

$$P_U(x, z, t) + P_D(x, z, t) \neq P(x, z, t) \quad (3.6)$$

Similarly, we can separate the wavefields into a left-going and a right-going component (equations 3.7, 3.8 and 3.9) and then perform a similar inverse transform

3.1 Theory of Finite Difference Explicit Wavefield Decomposition

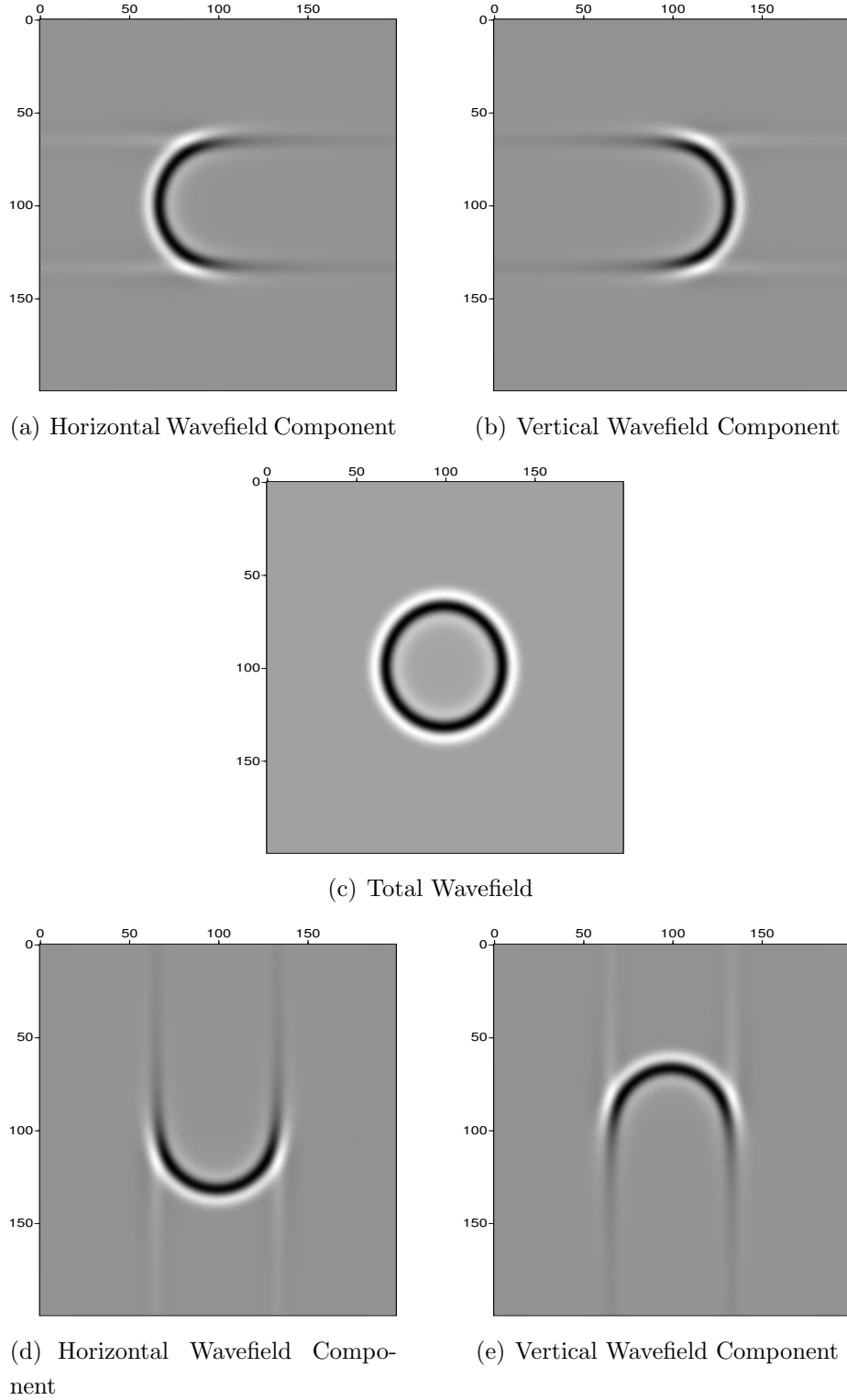


Figure 3.2: Directional wavefield splitting using Fourier transform. In this scheme the horizontal and vertical splitting is performed on the same wavefield. The splitting shows the present of evanescent energies

3.1 Theory of Finite Difference Explicit Wavefield Decomposition

as was done for the up-down separation to derive equations 3.10 and 3.11, with a similar property fulfilled by equation 3.12.

$$P(k_x, z, \omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(x, z, t) \exp[-j2\pi(\omega t + k_x x)] dx dt \quad (3.7)$$

$$P_L(k_x, z, \omega) = \begin{cases} P(k_x, z, \omega) & \text{if } \omega k_x \geq 0 \\ 0 & \text{if } \omega k_x < 0 \end{cases} \quad (3.8)$$

$$P_R(k_x, z, \omega) = \begin{cases} P(k_x, z, \omega) & \text{if } \omega k_x \leq 0 \\ 0 & \text{if } \omega k_x > 0 \end{cases} \quad (3.9)$$

$$P_R(x, z, t) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_R(k_x, z, \omega) \exp[+j2\pi(\omega t + k_x x)] dk_x d\omega \quad (3.10)$$

$$P_L(x, z, t) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_L(k_x, z, \omega) \exp[+j2\pi(\omega t + k_x x)] dk_x d\omega \quad (3.11)$$

$$P_L(x, z, t) + P_R(x, z, t) \neq P(x, z, t) \quad (3.12)$$

Worst still, the sum of these four components is about twice the original wavefield, plus evanescent energies. This is because two snapshots of the wavefield is taken to perform the two Fourier transforms (horizontal and vertical separations).

$$P_U(x, z, t) + P_D(x, z, t) + P_L(x, z, t) + P_R(x, z, t) \neq P(x, z, t) \quad (3.13)$$

The method can separate the energies into the respective propagation directions. However, it is restricted to separating the wavefields only in the orthogonal directions, as the Laplacian is defined in orthogonal directions when working in the Cartesian coordinates. Hence, the method suffers from errors when energies propagate at right angle to the direction of separation. Figure 3.2 shows the results of performing this sort of wavefield decomposition. The presence of the energies

3.1 Theory of Finite Difference Explicit Wavefield Decomposition

leaking in from the orthogonal directions is quite obvious.

Although this method of wavefield decomposition has been employed in RTM (Liu et al., 2011) and FWI (Wang, Chauris, Donno and Calandra, 2013; Wang et al., 2015), the phenomenon where energies leak in from orthogonal directions is not desirable in full waveform inversion. A method that provides a cleaner separation of the wavefields by propagation directions is therefore desirable. The finite difference explicit wavefield decomposition method gives a cleaner wavefield separation by minimising the presence of energies propagating at right angles in the Fourier transform.

3.1.2 Finite Difference Explicit Wavefield Decomposition

The challenge of performing Fourier decomposition on the total wavefield arises from the presence of energies, of equal or near equal amplitudes, propagating at an orthogonal direction to the direction of separation. A scheme that suppresses these orthogonal energies, as in Figure 3.3, minimises the presence of the noises present in wavefield decomposition via Fourier transform as the Fourier decomposition is performed on each of the component wavefields rather than on the total wavefield (as in the conventional method). The components must fulfil the requirement that their sum is exactly the original wavefield.

The scalar (constant density acoustic) wave equation as given in equation 2.25 can be rearranged as equation 3.14.

$$\frac{1}{c^2} \frac{\partial^2 P}{\partial t^2} - \left(\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial z^2} \right) = s(t) \quad (3.14)$$

Let us then perform a variable substitution for p such as:

$$P = P_x + P_z \quad (3.15)$$

P_x, P_z should just be seen for now to be any two components of a wavefield whose wavefield, at every time step, is the original wavefield.

3.1 Theory of Finite Difference Explicit Wavefield Decomposition

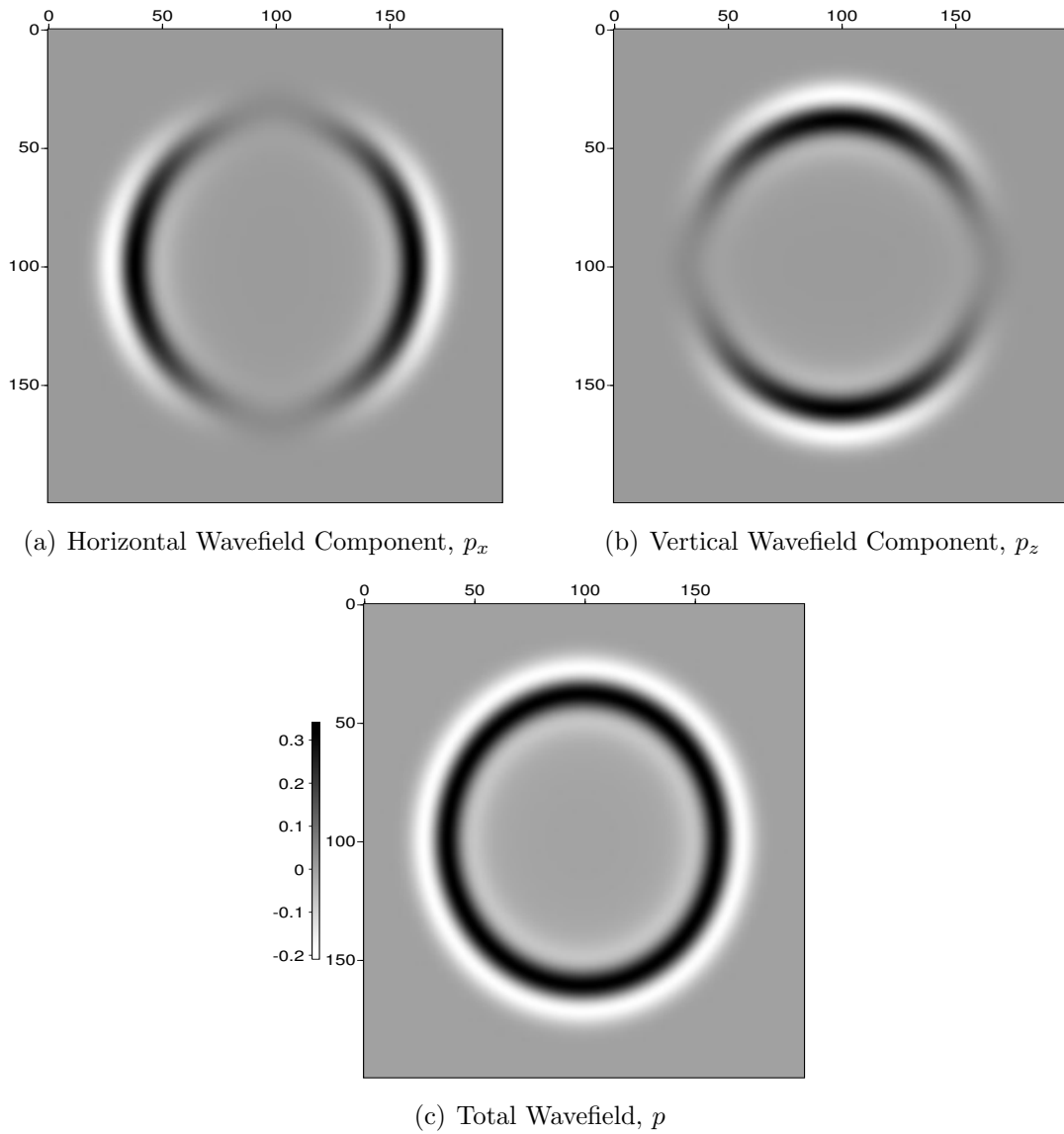


Figure 3.3: *Decomposition of the wavefield into horizontal and vertical components. The sum of these components exactly produces the original wavefield*

3.1 Theory of Finite Difference Explicit Wavefield Decomposition

Substituting into equation 3.14 results in 3.16.

$$\frac{1}{c^2} \frac{\partial^2}{\partial t^2} (P_x + P_z) - \left(\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial z^2} \right) = s(t) \quad (3.16)$$

Equation 3.16 can then be decomposed into a coupled wave equation system comprising equations 3.17 and 3.18.

$$\frac{1}{c^2} \frac{\partial^2 P_x}{\partial t^2} - \frac{\partial^2 P}{\partial x^2} = 0.5 * s(t) \quad (3.17)$$

$$\frac{1}{c^2} \frac{\partial^2 P_z}{\partial t^2} - \frac{\partial^2 P}{\partial z^2} = 0.5 * s(t) \quad (3.18)$$

The coupling of both equations comes from the fact that they share the same wavefield when computing the Laplacian at every time step. It can also be observed that a sum of equations 3.17 and 3.18 yields the original wave equation.

Equations 3.17 and 3.18 can be solved for two wavefields (P_x, P_z) , whose sum always yields the original wavefield P . P_x can be interpreted as the component of the wavefield that travels predominantly in the horizontal direction, while P_z is interpreted as the component of the wavefield that travels predominantly in the vertical direction.

Each of the component wavefields can further be decomposed into wavefields propagation in opposite directions as the wavefields evolve in time, by taking a Fourier transform of the 3D cube holding each component wavefield. This Fourier decomposition operation results in the original component wavefields being expressed as equation 3.19.

$$P_z(x, k_z, \omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_z(x, z, t) \exp[-j2\pi(\omega t + k_z z)] dz dt \quad (3.19)$$

3.1 Theory of Finite Difference Explicit Wavefield Decomposition

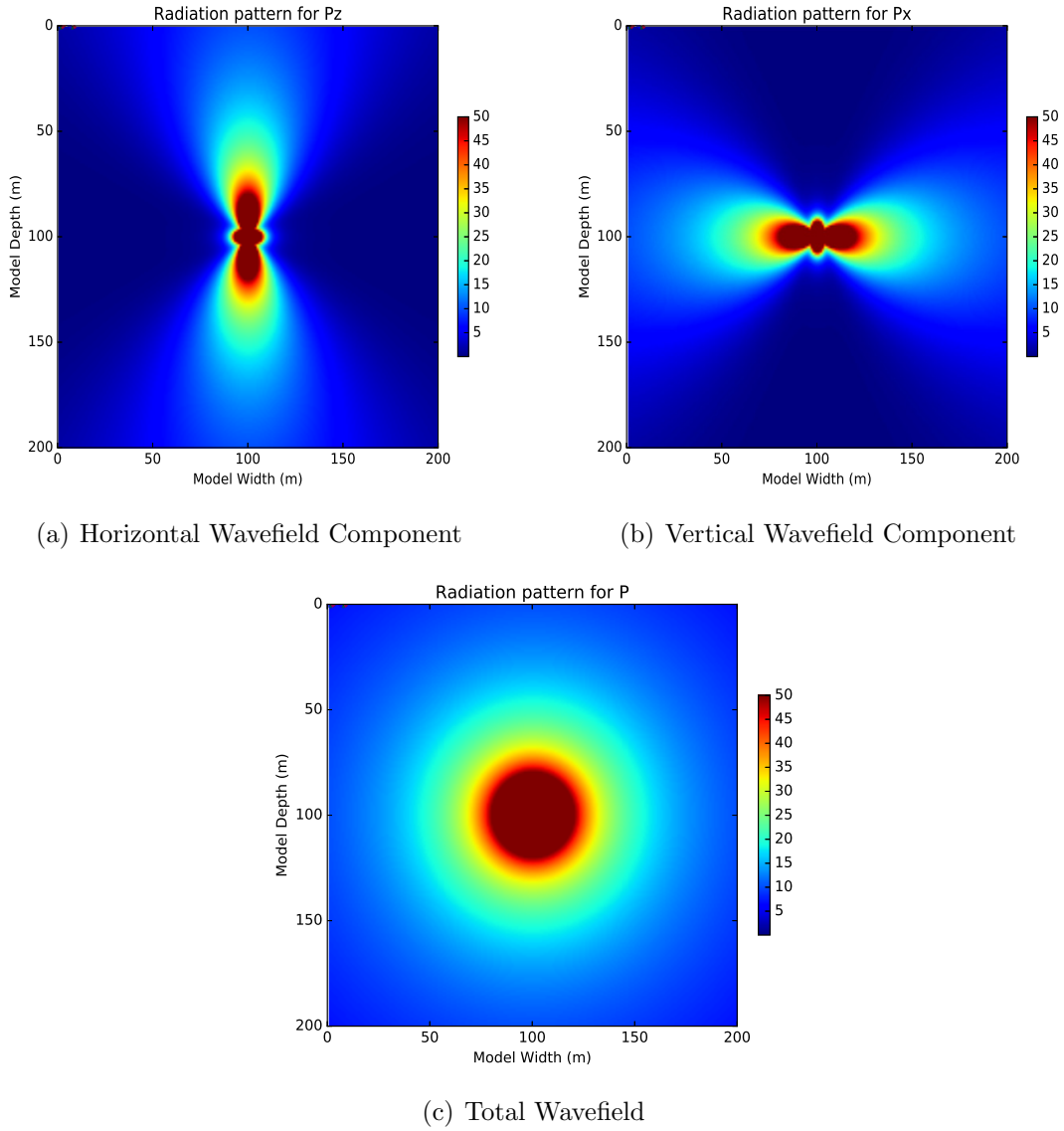


Figure 3.4: Source illumination of Finite Difference Decomposed Wavefields

3.1 Theory of Finite Difference Explicit Wavefield Decomposition

$P_z(x, k_z, \omega)$ represents the 2D Fourier transform of the vertical component of the wavefields. Separation into upward and downward propagating components in the (ω, k_z) domain proceeds as earlier:

$$P_{ZD}(x, k_z, \omega) = \begin{cases} P_Z(x, k_z, \omega) & \text{if } \omega k_z \geq 0 \\ 0 & \text{if } \omega k_z < 0 \end{cases} \quad (3.20)$$

and

$$P_{ZU}(x, k_z, \omega) = \begin{cases} P_Z(x, k_z, \omega) & \text{if } \omega k_z \leq 0 \\ 0 & \text{if } \omega k_z > 0 \end{cases} \quad (3.21)$$

$P_{ZD}(x, k_z, \omega)$ represent the downwards component of the wavefields in the (ω, k_z) domain, while $P_{ZU}(x, k_z, \omega)$ represent the upwards component of the wavefields in the (ω, k_z) domain. This is different from the separation achieved in the original approach as it minimises errors due to energies propagating at right angles to the separation directions.

Similarly, we can separate the horizontal component of the wavefields into a left-going and a right-going component given by equations 3.22, 3.23 and 3.24.

$$P_X(k_x, z, \omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_X(x, z, t) \exp[-j2\pi(\omega t + k_x x)] dx dt \quad (3.22)$$

$$P_{XL}(k_x, z, \omega) = \begin{cases} P_X(k_x, z, \omega) & \text{if } \omega k_x \geq 0 \\ 0 & \text{if } \omega k_x < 0 \end{cases} \quad (3.23)$$

$$P_{XR}(k_x, z, \omega) = \begin{cases} P_X(k_x, z, \omega) & \text{if } \omega k_x \leq 0 \\ 0 & \text{if } \omega k_x > 0 \end{cases} \quad (3.24)$$

This separation scheme has the property that the left propagating and the right propagating components sum up to the total horizontal wavefields (equation 3.25), while the downward propagating and the upward propagating components sum up to the total vertical wavefields (equation 3.26).

$$P_X = P_{XL} + P_{XR} \quad (3.25)$$

3.1 Theory of Finite Difference Explicit Wavefield Decomposition

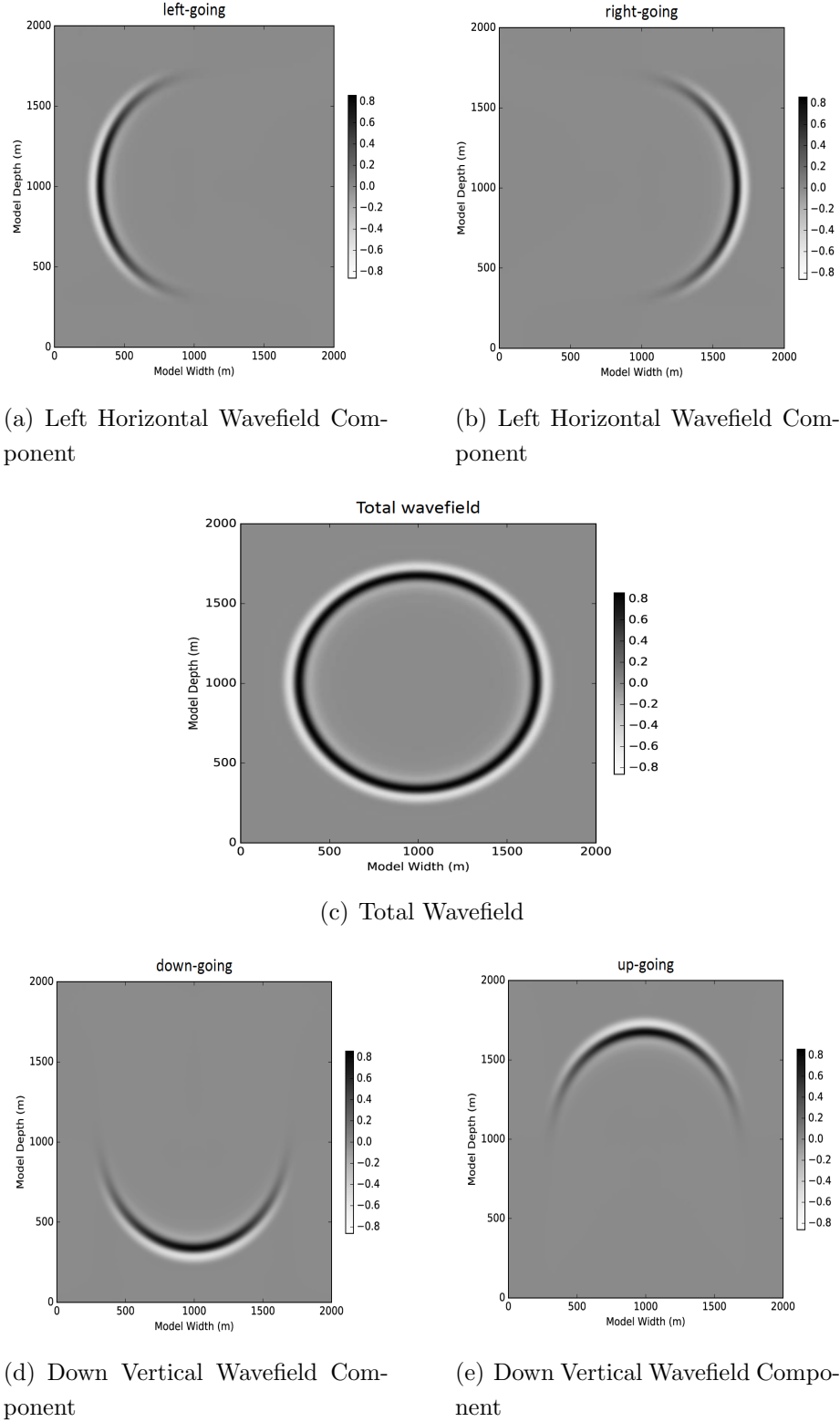


Figure 3.5: Directional wavefield splitting using Finite Difference Explicit wavefield Decomposition and Fourier transform.

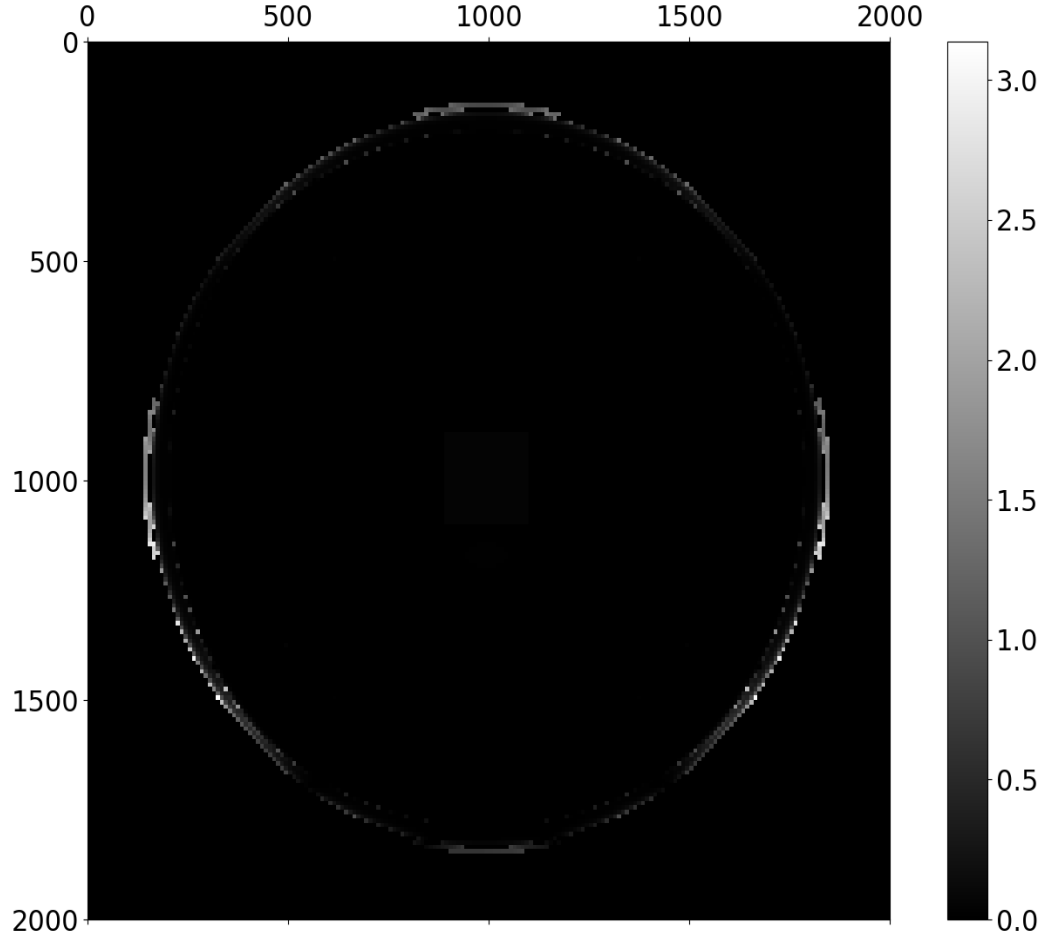


Figure 3.6: *Percentage difference between a snapshot of the sum of the four components of the wavefield produced by FDEWD, and a snapshot of the same wavefield computed with a conventional finite difference stencil. Although the phase of the original wavefield appears a bit distorted, however, inversion experiments, as discussed in later chapters, do not show any significant issue associated with this distortion.*

3.1 Theory of Finite Difference Explicit Wavefield Decomposition

$$P_Z = P_{ZU} + P_{ZD} \quad (3.26)$$

Hence, the total wavefield can be reconstructed via an algebraic sum of the component wavefields, i.e.

$$P = P_{XL} + P_{XR} + P_{XU} + P_{XD} \quad (3.27)$$

Snapshots of this decomposition scheme for wavefields propagating in a smooth velocity model are shown in Figure 3.7. It is quickly observed that at each snapshot the vertical and horizontal components sum to the total wavefield. The behaviour of this decomposition scheme for a model with scatterers is analysed by using a section of the Marmousi model in Figures 3.8 and 3.9. When the dips in the model are gentle and near horizontal, the vertical components of the wavefields contribute more to the total wavefields (Figure 3.8). However, as the dips in the model become steeper and approach vertical (Figure 3.9), the horizontal components contribute more to the total wavefields.

3.1 Theory of Finite Difference Explicit Wavefield Decomposition

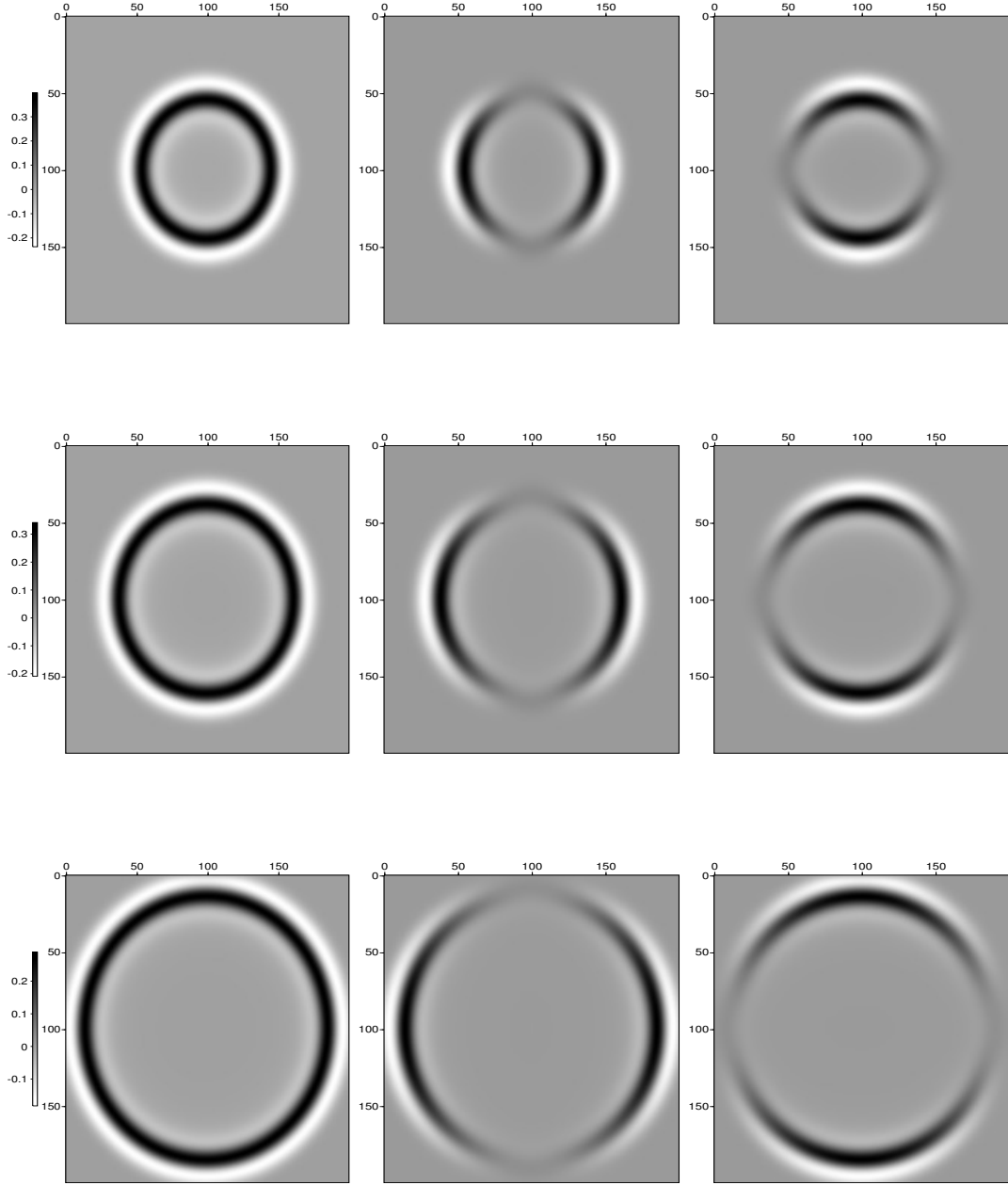
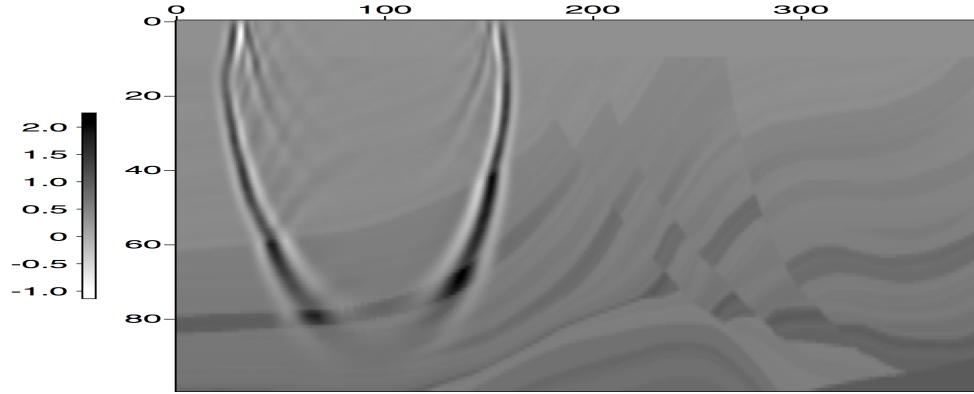
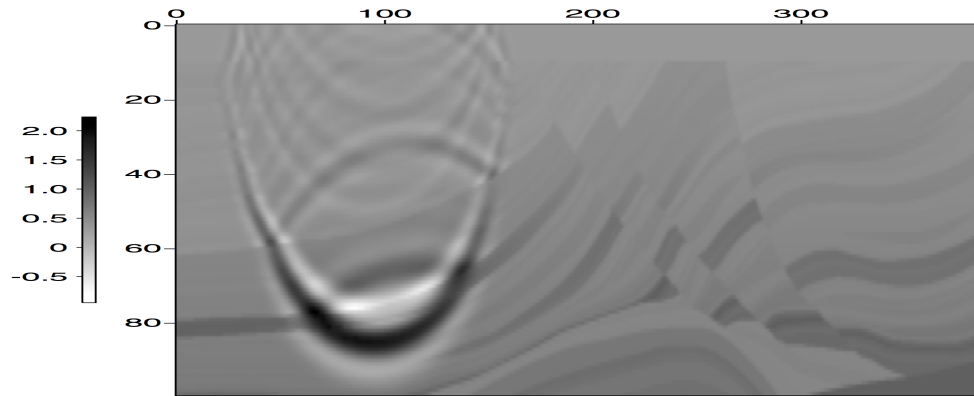


Figure 3.7: Evolution of a wavefield in time. The component wavefields also evolve as does the full wavefield. The leftmost column is the evolution of total wavefield, the middle column is the evolution of the horizontal wavefields, while the rightmost column is the evolution of the vertical wavefields.

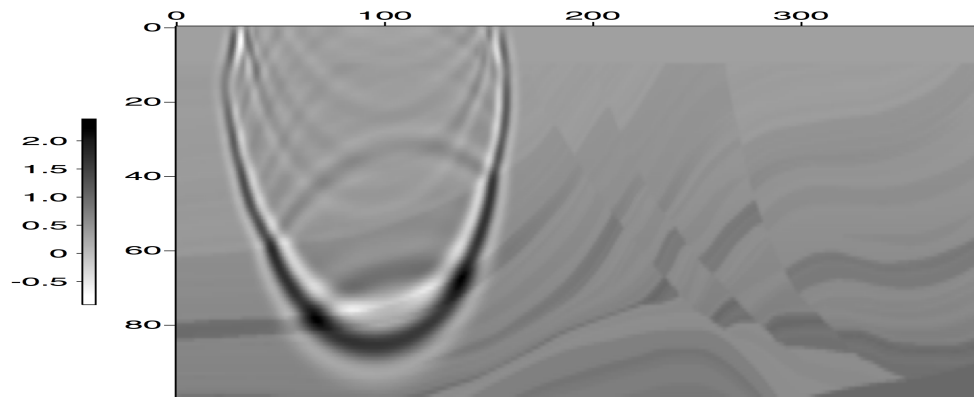
3.1 Theory of Finite Difference Explicit Wavefield Decomposition



(a) Horizontal Wavefields



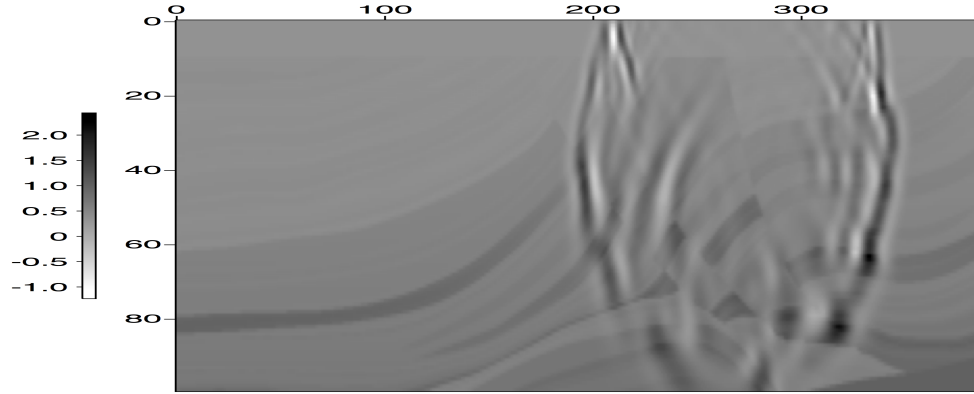
(b) Vertical Wavefields



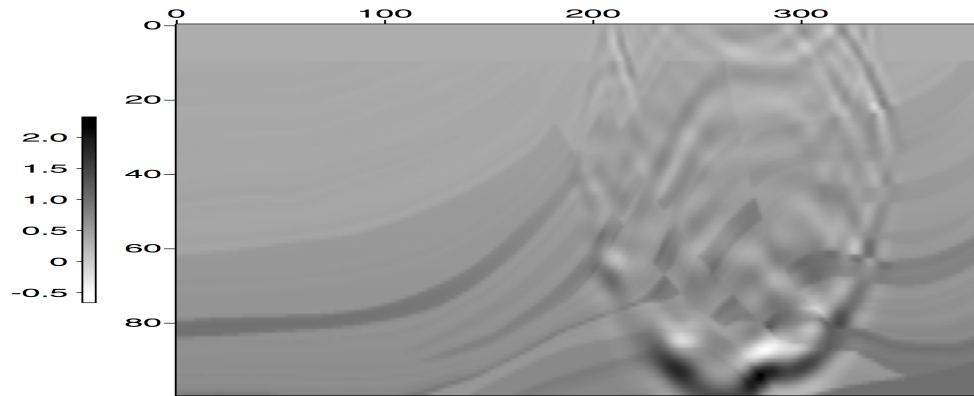
(c) Total Wavefields

Figure 3.8: An evolution of wavefields in time through a section of the Marmousi model

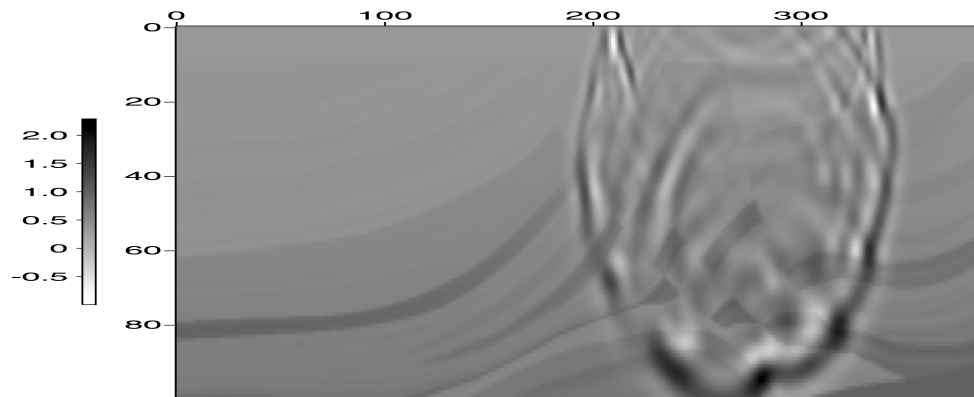
3.1 Theory of Finite Difference Explicit Wavefield Decomposition



(a) Horizontal Wavefields



(b) Vertical Wavefields



(c) Total Wavefields

Figure 3.9: An evolution of wavefields in time through a section of the Marmousi model

3.2 Gradient Computation in FWI with Finite Difference Explicit Wavefield Decomposition

In this section, matrix formalism is applied. Thus, a bold face upper case character represents a matrix, while a bold face lower case character denotes a vector. The letters used take their conventional meaning.

The conventional FWI gradient was derived in equation 2.39. If the decomposition given by equation 3.27 is substituted into equation 2.39, a new expression for the FWI gradient is arrived at (equation 3.28).

$$\nabla_{\mathbf{m}} F = -(\mathbf{p}_{\mathbf{x}l} + \mathbf{p}_{\mathbf{x}r} + \mathbf{p}_{\mathbf{z}u} + \mathbf{p}_{\mathbf{z}d})^T \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^T (\delta \mathbf{p}_{\mathbf{x}l} + \delta \mathbf{p}_{\mathbf{x}r} + \delta \mathbf{p}_{\mathbf{z}u} + \delta \mathbf{p}_{\mathbf{z}d}) \quad (3.28)$$

This particular decomposition scheme results in sixteen components gradients of the original gradient. The sum of these gradients always results in the original gradient. However, eight of these components (Equation 3.29 to 3.36) will be formed by wavefields propagating along similar orthogonal wavepaths, either in same direction in forward time or in opposite directions in forward time.

$$\nabla_{\mathbf{m}} F(ll) = -\mathbf{p}_{\mathbf{x}l}^T \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^T \delta \mathbf{p}_{\mathbf{x}l} \quad (3.29)$$

$$\nabla_{\mathbf{m}} F(rr) = -\mathbf{p}_{\mathbf{x}r}^T \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^T \delta \mathbf{p}_{\mathbf{x}r} \quad (3.30)$$

$$\nabla_{\mathbf{m}} F(uu) = -\mathbf{p}_{\mathbf{z}u}^T \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^T \delta \mathbf{p}_{\mathbf{z}u} \quad (3.31)$$

$$\nabla_{\mathbf{m}} F(dd) = -\mathbf{p}_{\mathbf{z}d}^T \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^T \delta \mathbf{p}_{\mathbf{z}d} \quad (3.32)$$

$$\nabla_{\mathbf{m}} F(lr) = -\mathbf{p}_{\mathbf{x}l}^T \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^T \delta \mathbf{p}_{\mathbf{x}r} \quad (3.33)$$

3.2 Gradient Computation in FWI with Finite Difference Explicit Wavefield Decomposition

$$\nabla_{\mathbf{m}} F(rl) = -\mathbf{p}_{\mathbf{x}r}^T \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^T \delta \mathbf{p}_{\mathbf{x}l} \quad (3.34)$$

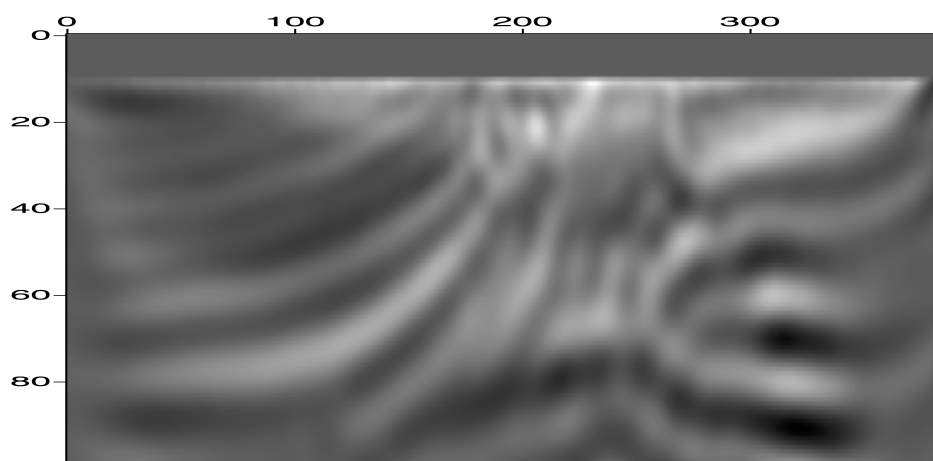
$$\nabla_{\mathbf{m}} F(ud) = -\mathbf{p}_{\mathbf{z}u}^T \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^T \delta \mathbf{p}_{\mathbf{z}d} \quad (3.35)$$

$$\nabla_{\mathbf{m}} F(du) = -\mathbf{p}_{\mathbf{z}d}^T \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^T \delta \mathbf{p}_{\mathbf{z}u} \quad (3.36)$$

Equations 3.29 and 3.30 are interpreted tomographic gradient contribution due to wavefields propagating in the horizontal direction. Similarly, equations 3.31 and 3.32 are interpreted tomographic gradient contribution due to wavefields propagating in the vertical direction. The migration gradient contributions due to horizontally propagating wavefields are given in equations 3.33 and 3.34, and finally the migration gradient contributions due to vertically propagating wavefields are given in equations 3.35 and 3.36.

A presentation of these gradients is shown in Figures 3.10 through 3.14. Figure 3.10 is the total gradient while Figures 3.11 to 3.14 are the eight components of the gradient. The decomposed gradients yield cleaner gradients as they contain minimal crosstalks from orthogonal directions. They only extend the application of FWI to high frequency regime by allowing the building of the background velocities in the absent of low frequencies.

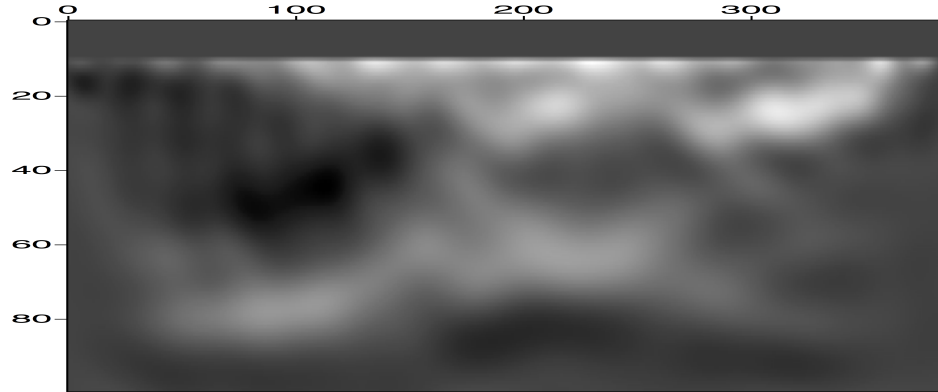
3.2 Gradient Computation in FWI with Finite Difference Explicit Wavefield Decomposition



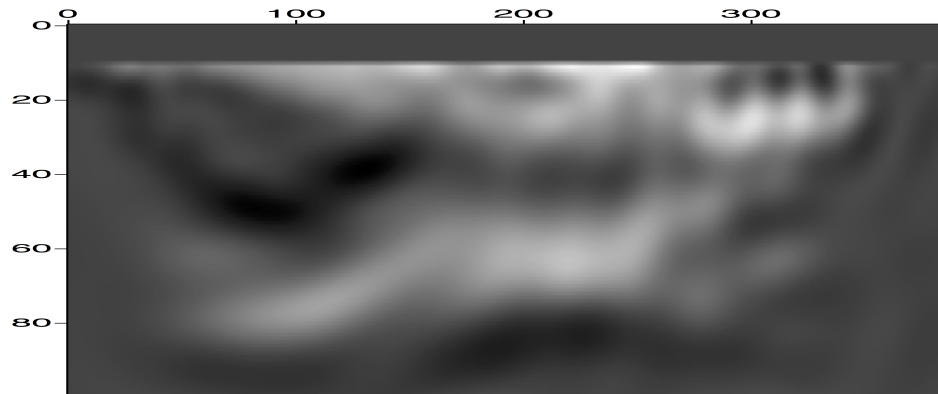
(a)

Figure 3.10: Total gradient

3.2 Gradient Computation in FWI with Finite Difference Explicit Wavefield Decomposition



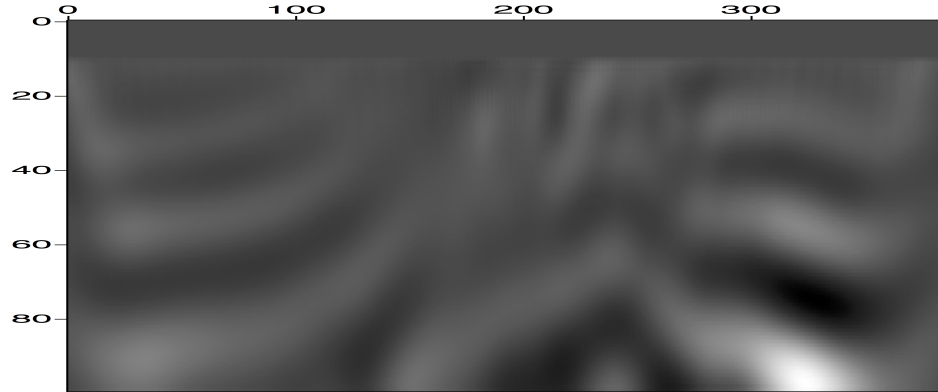
(a) Right going energies gradient



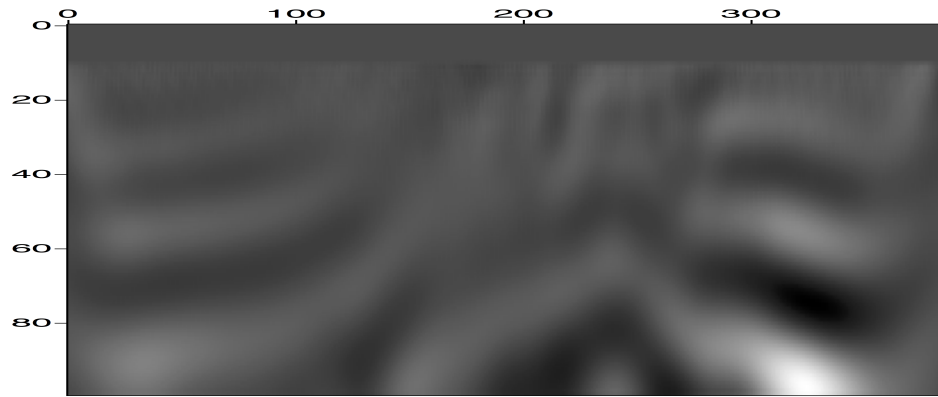
(b) Left going energies gradient

Figure 3.11: *Tomographic gradient contributions from cross-correlating horizontal wavefields propagating along same direction in time*

3.2 Gradient Computation in FWI with Finite Difference Explicit Wavefield Decomposition



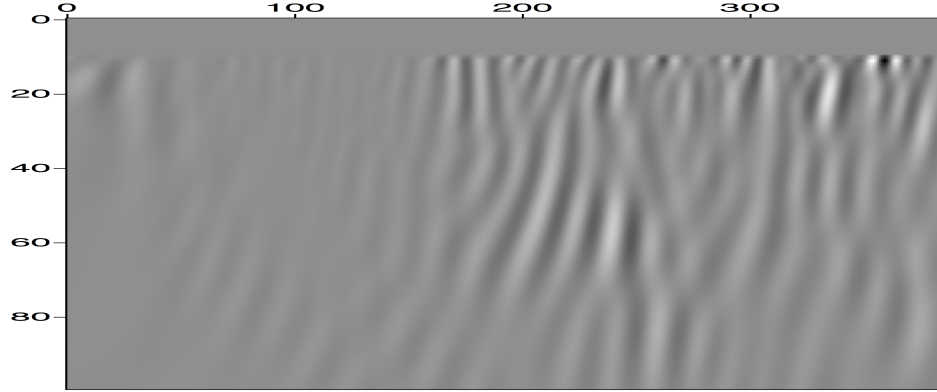
(a) Up going energies gradient



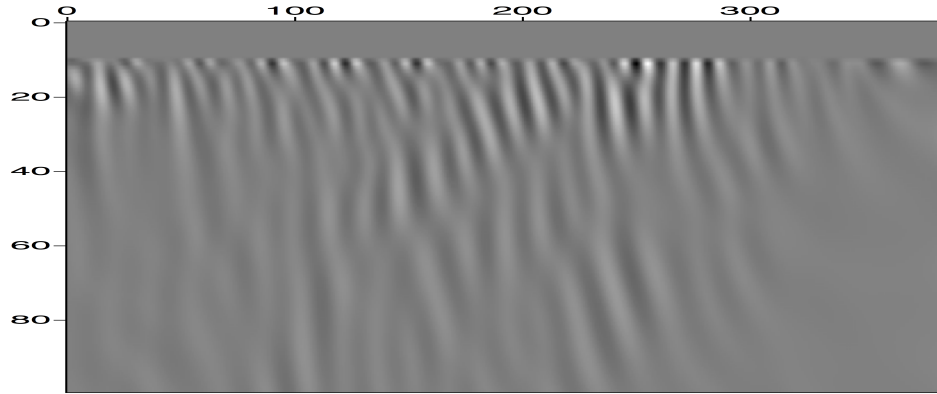
(b) Down going energies gradient

Figure 3.12: Tomographic gradient contributions from cross-correlating vertical wavefields propagating along same direction in time

3.2 Gradient Computation in FWI with Finite Difference Explicit Wavefield Decomposition



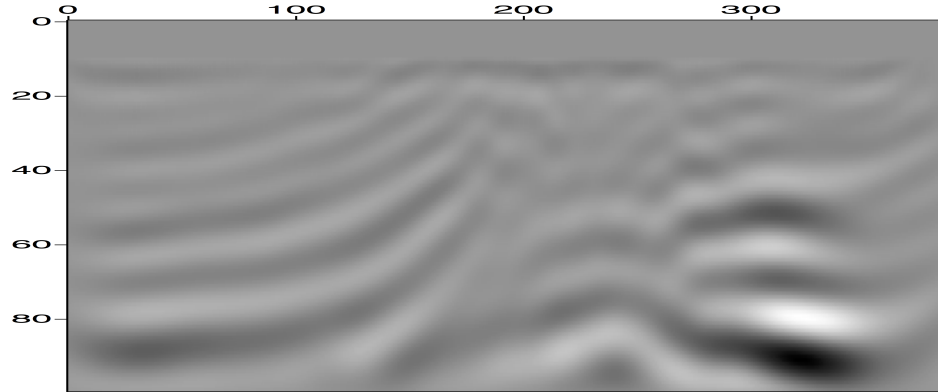
(a) Right-Left going energies gradient



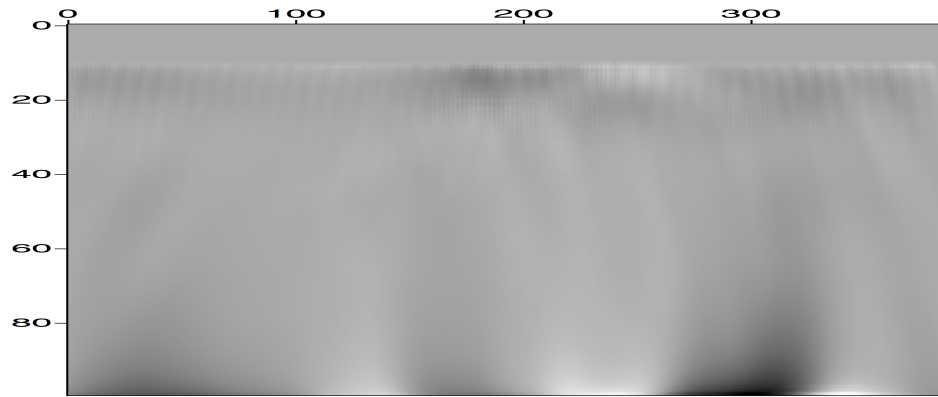
(b) Left-Right going energies gradient

Figure 3.13: Migration gradient contributions from cross-correlating horizontal wavefields propagating along different directions in time

3.2 Gradient Computation in FWI with Finite Difference Explicit Wavefield Decomposition



(a) Up-Down going energies gradient



(b) Down-Up going energies gradient

Figure 3.14: Migration gradient contributions from cross-correlating vertical wavefields propagating along different directions in time

3.2.1 Analysis of the gradients of FDEWD

The gradients generated by the FDEWD scheme has some unique properties that makes it a good candidate scheme for implementing reflection waveform inversion.

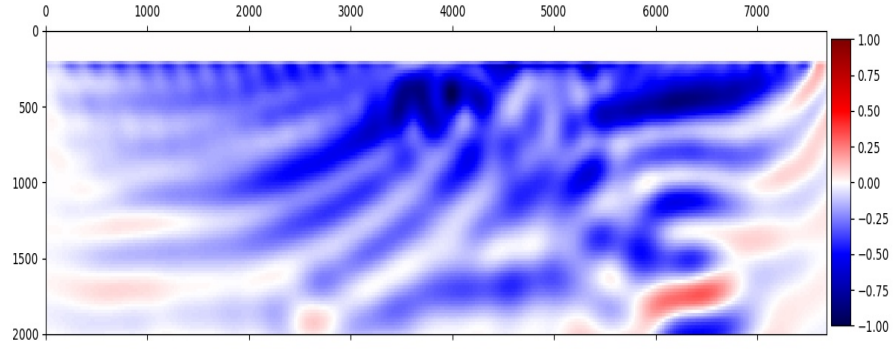
Intermediate wavenumbers

The lack of intermediate wavenumbers in conventional seismic data (Claerbout, 1985; Jannane et al., 1989) hinders the application of FWI in inverting for earth models, especially in scenarios where a generating a reasonable starting model is prohibitive. This wavenumber gap is caused by lack of low frequencies in the source spectrum (low frequencies here refer to frequencies below 1.5Hz). It is no surprise that in situations where low frequency seismic energies have been used (Baeten et al., 2013; Plessix et al., 2010), the inverted results have been quite significantly improved. Low frequencies seismic acquisition, however, are quite expensive.

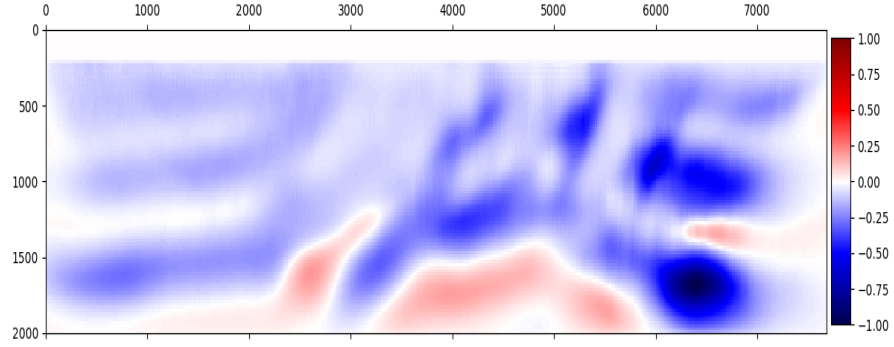
In conventional reflection seismic data lacking the necessary low frequencies, the envelope method has been explored as an alternative to generating the intermediate low wavenumbers. Similar to the envelope method, FDEWD also has the capability to generate intermediate wavenumbers even when low frequencies are absent in the data.

In Figures 3.15 to 3.17 comparisons are shown between the gradient produced by conventional FWI and that produced by FDEWD at frequencies 4Hz, 6Hz and 8Hz. The wavenumbers present in the tomographic gradients of FDEWD are much lower than those present in the corresponding conventional gradient. Also, it can be noticed that the FDEWD tomographic updates at 8Hz have similar vertical wavenumbers in comparison to the gradients of the conventional FWI method. Appendix 3 contains additional examples comparing the wavenumbers from conventional FWI and FDEWD gradients

3.2 Gradient Computation in FWI with Finite Difference Explicit Wavefield Decomposition



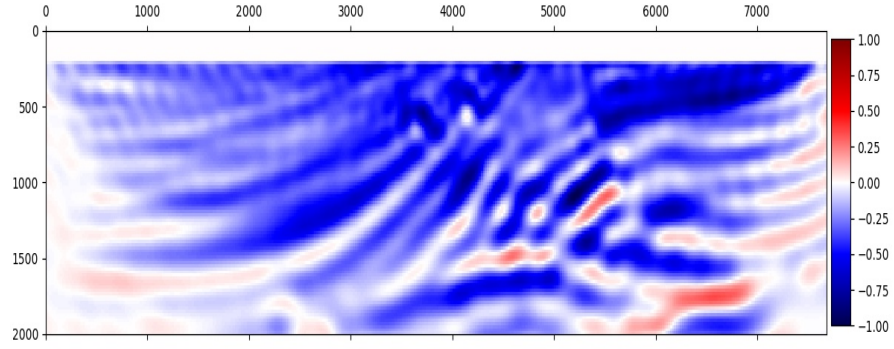
(a) Conventional FWI gradient



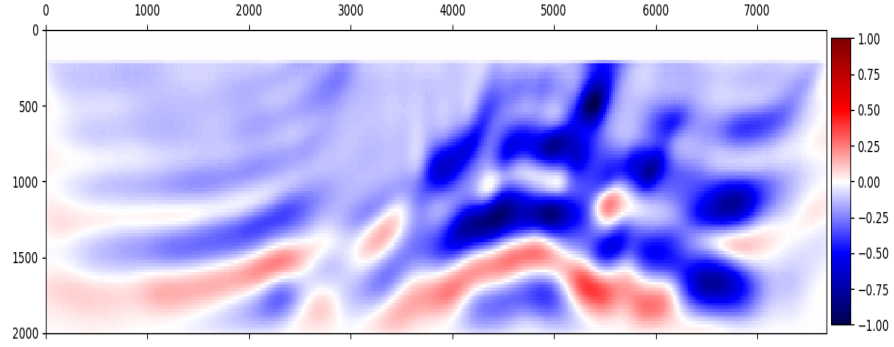
(b) FDEWD tomographic gradient

Figure 3.15: Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 4Hz

3.2 Gradient Computation in FWI with Finite Difference Explicit Wavefield Decomposition



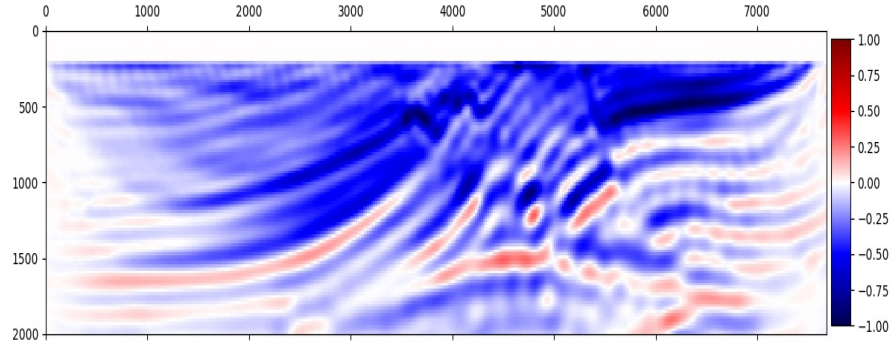
(a) Conventional FWI gradient



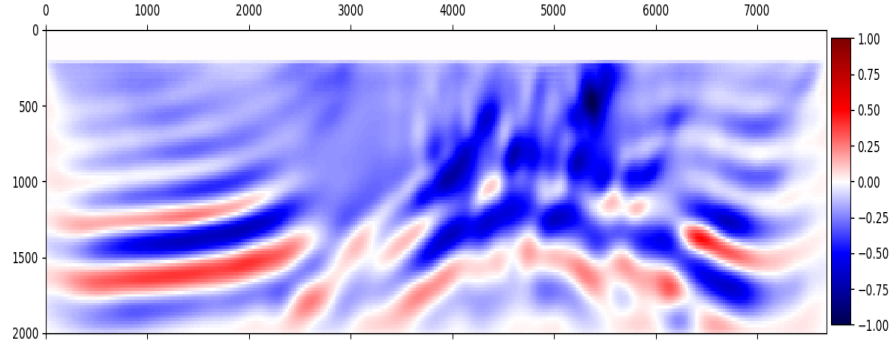
(b) FDEWD tomographic gradient

Figure 3.16: Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 6Hz

3.2 Gradient Computation in FWI with Finite Difference Explicit Wavefield Decomposition



(a) Conventional FWI gradient



(b) FDEWD tomographic gradient

Figure 3.17: Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 8Hz

3.3 Applications of FDEWD

The availability of the tomographic and migration components of the FWI gradient allows for various application of the methods that guarantees such separation. While the tomographic component is used in various FWI schemes to update the low wavenumber components of the model, the migration component can be used to update the high wavenumber component of the model or even applied (with some modifications) in an RTM scheme. FDEWD, therefore, can be potentially applied in RTM and FWI.

3.3.1 FDEWD and the PML Absorbing Boundary Condition

The perfectly matched layer, PML (Berenger, 1994), is the preferred absorbing boundary condition for simulation of wave equation (Collino and Tsogka, 2001; Komatitsch and Tromp, 2003; Liu and Tao, 1997). The split PML method, described in this section, naturally requires the splitting of the wavefield, a condition which is met by the FDEWD.

Starting from a plane wave solution of the homogeneous wave equation (equation 3.37), if the solution is adjusting by adding a decay factor, it will result in equation 3.38.

$$p = A \exp[i\omega(t - \frac{r\hat{n}}{c})] \quad (3.37)$$

Where p is pressure, A is the wave amplitude, ω is the frequency, t is time, r is the distance from source, $\hat{n}$ is the unit vector in the direction of wave propagation and c is the velocity.

$$p = A \exp[i\omega(t - \frac{r\hat{n}}{c}) - \alpha \frac{ct}{d}] \quad (3.38)$$

where α is some arbitrary damping constant and d is the PML distance over which the solution is damped.

Equation 3.38 can be differentiated as:

$$\frac{\partial p}{\partial t} = (i\omega - \frac{\alpha c}{d})p \quad (3.39)$$

$$\begin{aligned} \frac{\partial^2 p}{\partial t^2} &= (i\omega - \frac{\alpha c}{d})^2 p \\ &= (-\omega^2 + \frac{\alpha^2 c^2}{d^2} - 2i\omega \frac{\alpha c}{d})p \end{aligned} \quad (3.40)$$

and

$$\nabla^2 p = -\frac{\omega^2}{c^2} p \quad (3.41)$$

Equation 3.41 can be inserted into equation 3.40, resulting in equation 3.42.

$$\frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = \nabla^2 p + (\frac{\alpha^2}{d^2} - 2i\omega \frac{\alpha}{cd})p \quad (3.42)$$

Using equation 3.39

$$2i\omega \frac{\alpha}{cd} p = 2 \frac{\alpha}{cd} \frac{\partial p}{\partial t} + 2 \frac{\alpha^2}{d^2} p \quad (3.43)$$

Equation 3.42 becomes

$$\frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = \nabla^2 p - 2 \frac{\alpha}{cd} \frac{\partial p}{\partial t} - \frac{\alpha^2}{d^2} p \quad (3.44)$$

In the finite difference implementation of the split PML absorbing boundary condition, the wavefield is split into orthogonal components at the boundary. In 2D, this is written as

$$p = p_x + p_z$$

Equation 3.44 is thus written as

$$\frac{1}{c^2} \frac{\partial^2(p_x + p_z)}{\partial t^2} = \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial z^2} - 2 \frac{\alpha}{cd} \frac{\partial(p_x + p_z)}{\partial t} - \frac{\alpha^2}{d^2} (p_x + p_z) \quad (3.45)$$

which is then decomposed into

$$\frac{1}{c^2} \frac{\partial^2 p_x}{\partial t^2} = \frac{\partial^2 p}{\partial x^2} - 2 \frac{\alpha}{cd(x)} \frac{\partial p_x}{\partial t} - \frac{\alpha^2}{d(x)^2} p_x \quad (3.46)$$

and

$$\frac{1}{c^2} \frac{\partial^2 p_z}{\partial t^2} = \frac{\partial^2 p}{\partial z^2} - 2 \frac{\alpha}{cd(z)} \frac{\partial p_z}{\partial t} - \frac{\alpha^2}{d(z)^2} p_z \quad (3.47)$$

Within the computation domain (Figure 3.18), the decay constant is zero and thus equations 3.46 and 3.47 reduce to

$$\frac{1}{c^2} \frac{\partial^2 p_x}{\partial t^2} = \frac{\partial^2 p}{\partial x^2} \quad (3.48)$$

and

$$\frac{1}{c^2} \frac{\partial^2 p_z}{\partial t^2} = \frac{\partial^2 p}{\partial z^2} \quad (3.49)$$

which are the homogeneous forms of equations 3.17 and 3.18.

3.3.2 Gradients Mixing, Scaling and the Subspace Method

The FDEWD method yields model parameters of different physical dimensions. It is important that these model parameters be mixed in a balanced way, at every iteration, to speed up convergence and guarantee a better result compared to conventional FWI method. A framework that utilises FDEWD method can thus be designed in a subspace fashion (Kennett and Williamson, 1988; Oldenburg and Li, 1994), with 2-dimensional subspace model space. Unlike the conventional

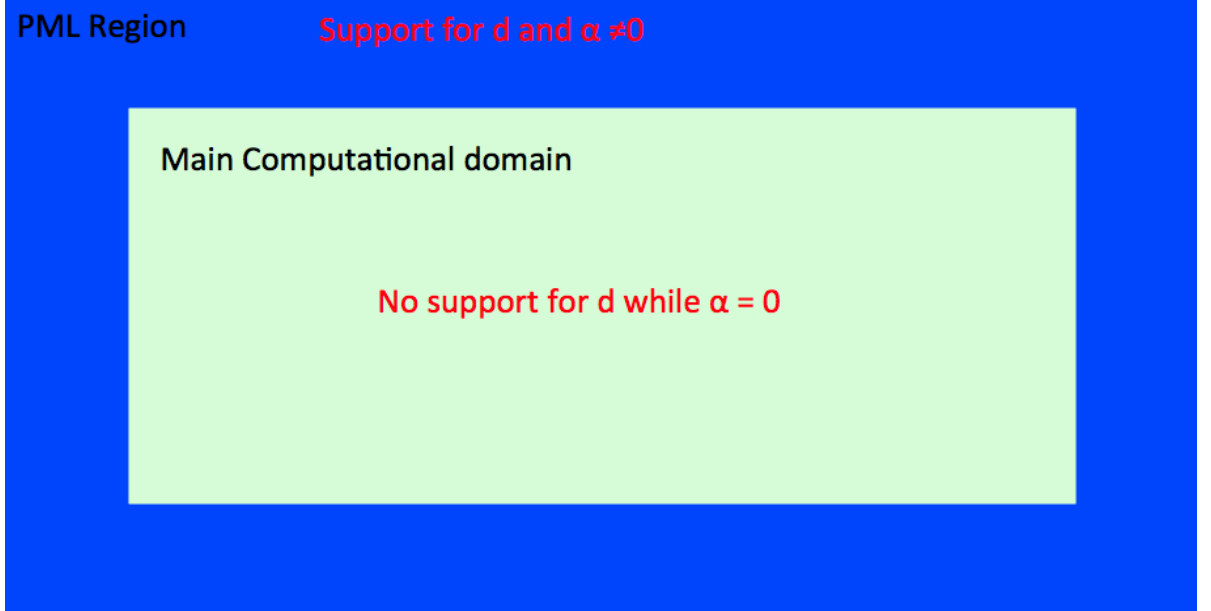


Figure 3.18: The computational and the PML domain, showing supports for decay constant, α and damping distance, d

method where the search for the model update is performed along a single direction, in the subspace method (Rawlinson and Sambridge, 2003), the search is carried out simultaneously along multiple search directions that when combined, span a subspace of the model space.

In the general subspace method, the optimization is restricted to a relatively small n -dimensional subspace of a larger M -dimensional model space at each iteration. The model perturbation to the current model is then computed as a linear combination of the scaled basis function of each subspace. If δm represents the model perturbation, and a^j represents a basis function, then

$$\delta \mathbf{m} = \sum_{j=1}^n \mu_j \mathbf{a}^j = \mathbf{A} \boldsymbol{\mu} \quad (3.50)$$

3.3.3 FDEWD and Conventional FWI

In the conventional FWI approach, we proceed by first building the low wavenumber components of the model, and continuing in some wavenumber continuation scheme (known commonly as multiscale FWI) to update the higher wavenumbers of the model.

3.3.4 FDEWD and Reflection FWI

In chapter two, an overview of reflection FWI was presented. It was shown that reflection FWI is highly depended on the ability to separate the FWI gradient into a tomographic and a migration component (Xu, Chen, Lambaré, Zhang, Wang et al., 2012).

The ability of FDEWD to cleanly separate the FWI gradient into a tomographic and migration component makes it a good scheme for performing reflection FWI.

3.3.5 FDEWD and RTM

It is a well established knowledge that when the migration velocity include strong impedance contrasts, RTM suffers from some low wavenumber artefacts (Zhang and Sun, 2008). Several schemes have been suggested by different authors to suppress these low wavenumber noises (Laplace filtering (Youn and Zhou, 2001); Poynting vector method (Hu and McMechan, 1987; Liu et al., 2011; Yoon et al., 2004)).

3.4 Conclusion

The theory of finite difference explicit wavefield decomposition is shown here. It allows us to generate up to sixteen unique gradient components which are grouped into tomographic updates, migration updates and noise. These gradients have many application in seismic data inversion and imaging. The migration updates

3.4 Conclusion

can be combined with the tomography update to carry out reflection FWI. The scheme could also be adapted to run the code in a reverse time migration setting. The scheme can be adapted to run transmission FWI without the need for any data muting.

Chapter 4

Numerical Testbed for Seismic Tomography

4.1 Introduction

One of the main objectives of the thesis was to design and implement a numerical testbed for studies in seismic tomography (and imaging). Hence, the entire code used in this project was developed during the course of the project, save for the visualization and plotting tools. The code is written entirely in Fortran ([Brainerd et al., 2009](#); [Chapman, 2008](#); [Chivers and Sleightholme, 2006](#); [Markus, 2012](#); [Metcalf et al., 2011](#)).

The initial code was written to Fortran 95 standard, largely because compilers didn't support the 2003 and 2008 standards. However, it was later refactored to incorporate some aspects of the Fortran 2003 and 2008 standards, especially as common compilers now support most of the 2003 standard and a good amount of the 2008 standard.

The codebase has been designed and written as a testbed for research in tomography and migration, with modularity and ease of testing new ideas key foci of its design. While speed was not a primary focus of the design, efforts were made to write functions and subroutines to encourage auto-vectorization and paral-

lization by the compiler. OPENMP (Chandra, 2001; Chapman et al., 2008; Dagum and Menon, 1998) constructs were also employed to parallelise workloads and improve performance. The current state of the code is written to solve 2D tomography problems, and highly biased towards same. However, extension to 3D tomography (and migration) problems should involve some not too complex additions/changes.

The architectural framework of the code is component-based (Bass, 2007; Bernholdt et al., 2006; Crnkovic, 2001; Garlan and Shaw, 1993; Heineman and Council, 2001; Kemball et al., 2008; Krosche and Hautefeuille, 2007; Quintino, 2008; Szyperski et al., 1999) , where every functional and/or logical unit that contributes to the solution of the tomography problem forms a component and communicates with other components via well-defined interfaces. Each component represents a modular, replaceable, portable and reusable collection of explicit functionalities with associated interfaces. This framework comes with the advantage that new ideas can be quickly implemented in any aspect of the tomography solution by implementing some generic interface in a new component or as part of an existing component.

This chapter discusses the design, implementation, benchmarking and refactoring steps taken while writing the software used in this project.

4.2 Design

The design of the tomography code involves an analysis and development of the various geophysical and mathematical aspects of the problem, and a careful and methodological design of the computer program. Investigations in FWI can be decomposed into a number of inter-related geophysical and mathematical areas, including, but not restricted to :

- Wave Propagation and Scattering
- Seismic Tomography and Inversion
- Source Injection

- Numerical Dispersion
- Numerical Solutions to PDEs
- Problem Formulation/Definition
- Gradient Computation
- Hessian Approximation
- Step-length Computation
- Global Problem Type
- Physics Approximation

A good testbed should therefore allow for implementation and testing of components, implemented within modules, subroutines and functions, in each of the above research areas. The code base developed in this research balances cohesion and loose coupling requirements to handle each study. High level design decisions include designing modules to handle any of the aforementioned research areas. It also includes how the various modules are associated, aggregated and composed. Typically, each module is composed of subroutines and functions, overloaded with a number of generic interfaces.

Following a component-based software development paradigm, the codebase currently has the following components, and their associated interfaces:

Figure 4.6 shows the relationships that exist among all the components. The utilities component provides interfaces for linear algebra, file IO, time series, and other tasks as shown in Figure 4.4. Hence, it is understandable how every other component has some form of dependency on it. New functionality can be added to a corresponding component, and the necessary interface provided for such functionality.

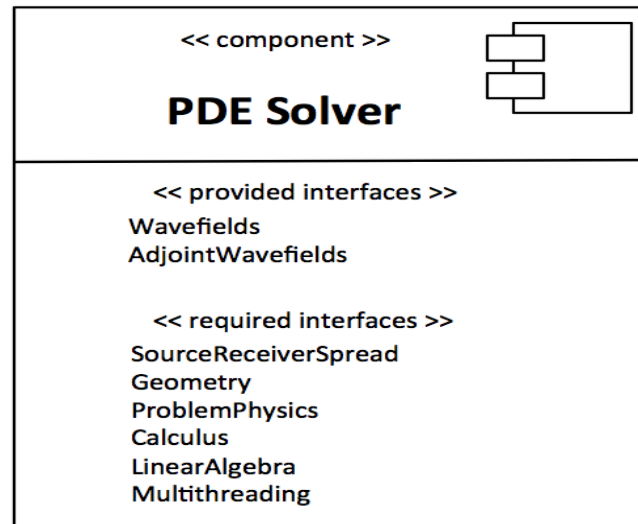


Figure 4.1: PDe Solver Component.

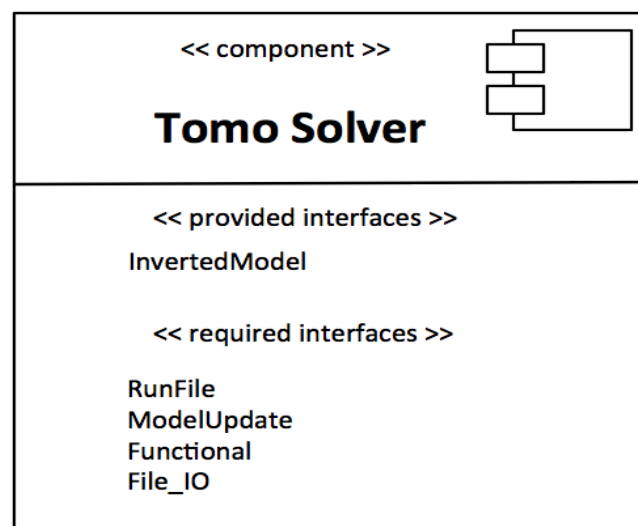


Figure 4.2: Tomography Component.

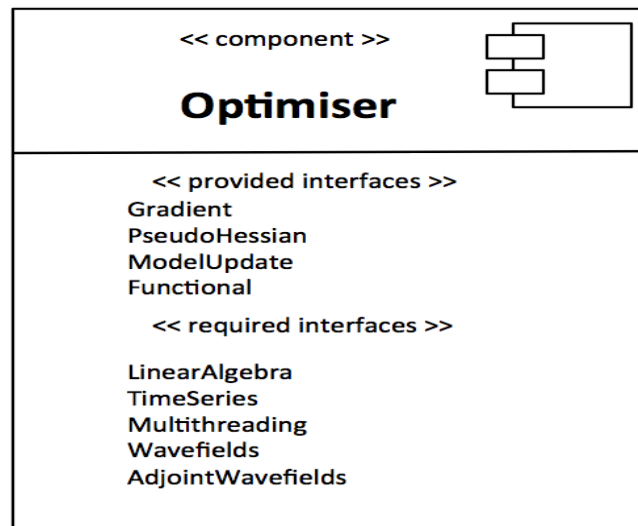


Figure 4.3: Optimiser Component.

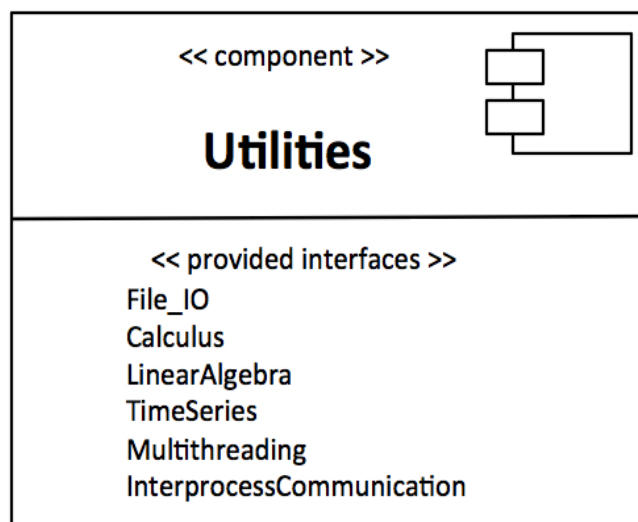


Figure 4.4: Utils Component.

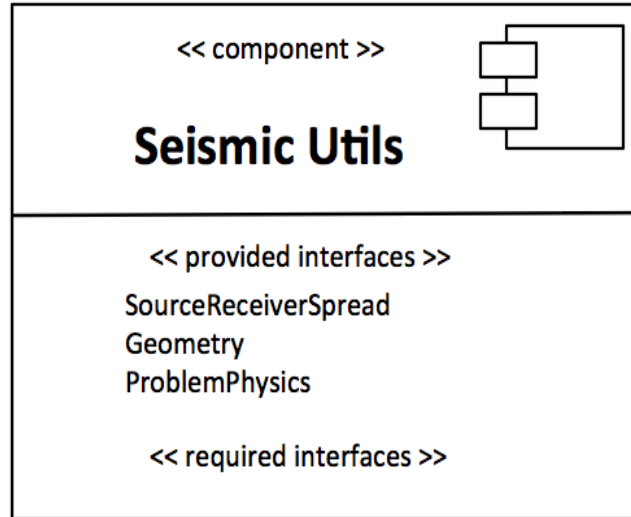


Figure 4.5: Seismic Utils Component.

4.3 Implementation

The code has been implemented in both a procedure and an object oriented paradigm. Both versions of the code have been tested and applied to invert various synthetic datasets. This section discusses the implementation of these versions of the code.

4.3.1 Procedural Implementation

Principles and Strategies

A guiding principle in the implementation of the testbed is the enforcing of a loose coupling among the various components. In procedural Fortran programming, extensive use of Fortran `MODULE` and `GENERIC INTERFACE` blocks have been employed to facilitate a design that allows for easy switching of components. `GENERIC INTERFACE` blocks ensure that new *SUBROUTINE/FUNCTION* could be inserted into the code base with little or no editing to the driver code.

Use of `ELEMENTAL` and `PURE` procedures are employed to suggest to the

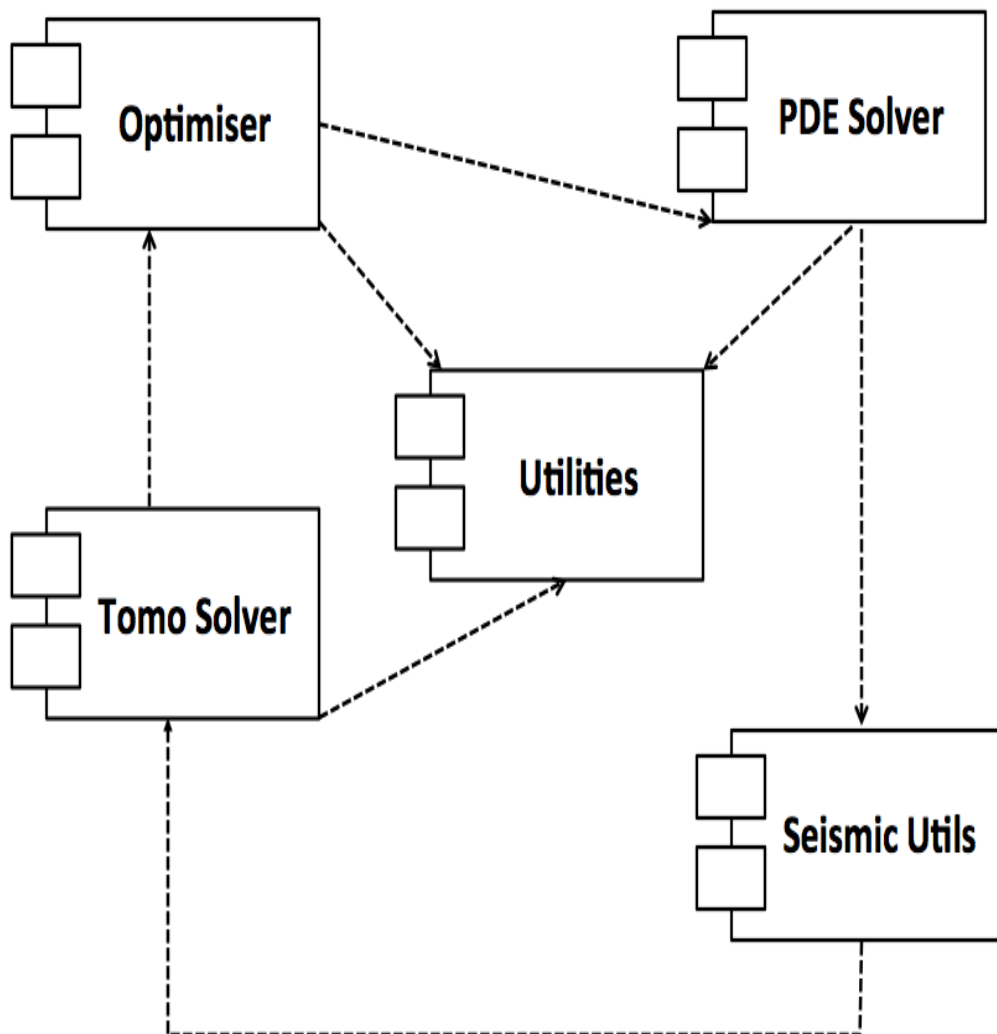


Figure 4.6: Relationships amongst all components.

compiler to optimise subroutines/functions. The optimisation flags are turned on during compilation and such procedures are strong candidates for compiler optimisation.

Named BLOCK constructs are used to name computation blocks to ensure clarity of the code.

Errors are handled as much as possible to give information above failing subroutines and functions and why the errors occurred (where possible). SUBROUTINES and FUNCTIONS that are likely to raise errors are written to return *ERROR_CODE* and/or *ERROR_MESSAGE*. Issues of thread safety requires careful handling in every multi-threaded code. In the design of the testbed, the use of global data has been minimised. File units have also been handled such that a new and available UNIT is provided for each new file access.

Tomography and Synthetics Drivers

A separate driver code is provide for synthetic and tomography runs. Listing 4.1 shows a dummy code for the tomography driver. The various dependencies that exist amongst the various components start to become more obvious. Also, the purpose of the driver is clear and the lines of code involved are kept to a manageable minimum.

Listing 4.1: Main tomography driver

```
1 program tomography
2   use utilities , only : logger
3   use kind_parameters
4   use optimiser , only : compute_gradient , &
5       compute_step_length , update_inversion_params
6   use global_data_module , only : model , step_length , &
7       model_update , inversion_params
8   implicit none
9
10      call get_inversion_params(inversion_params)
11
12      do
13          model_update=compute_gradient(inversion_params)
```

```

14         step_length=compute_step_length(inversion_params,&
15                                         model_update)
16         model = model - step_length * model_update
17
18         call update_inversion_params()
19         call logger%log (inversion_params%status)
20         if (stopping_criteria_met()) exit
21     end do
22
23 contains
24
25     function      stopping_criteria_met() result(action)
26     implicit none
27     logical      :: action
28
29     if (inversion_params%num_iter_done == &
30         inversion_params%total_iters) then
31         action = .true.
32     end if
33 end function
34
35 end program

```

Uniform Data Kinds and Precision

Benchmarking codes imposes a consistency in the output of the code. There could be some inconsistency in the results of the same code when run on different machines. To ensure consistency of results, a module definition for kind parameters is provided. This module defines the precision for every variable used in the program. Subsequently, every module, subroutine and function in this code base makes use of this kind_parameters module (Listing 4.2).

Listing 4.2: Kind parameters used in the code

```

1 module kind_parameters
2 implicit none
3 private
4 public      :: sp, dp, qp
5 integer, parameter :: sp = selected_real_kind(6, 37)

```

```

6 | integer, parameter :: dp = selected_real_kind(15, 307)
7 | integer, parameter :: qp = selected_real_kind(33, 4931)
8 | end module kind_parameters

```

Generic Interface

Generic interfaces were used in this code base as an approach to ensuring minimal adjustments to procedure names with the addition of new procedures to the code base. Initially, the testbed implemented a first order velocity-stress formulation to the solution of the acoustic wave equation. When the second order pressure only solution is implemented, the new procedure is placed into the forward propagator interface. This approach makes it easy to compare of performance of diifferent

4.3.2 Object-Oriented Implementation

In an object-oriented Fortran, interfaces are approximated via the use of ABSTRACT DERIVED TYPES, with Deferred procedure binding. This leaves the implementations of the precedures to child classes. The loose coupling requirement for good software development provides two guiding principles:

- Favor object composition over class inheritance.
- Program to an interface, not an implementation.

Abstract Data Type

The idea of an abstract data type (ADT) encourages thorough thinking about how softwares are designed. Abstraction allows for focus on the big picture of software design, without getting lost in the details.

***Listing 4.3:** Fortran Skeleton Abstrtact Data Type*

```

1 | module skeleton_ADT
2 |   use, only :
3 |   implicit none

```

```

4  private
5  public  :: skeleton
6
7  type  skeleton
8      private
9      contains
10     procedure , public ::
11     final  :: clean_up
12 end type
13
14 interface skeleton
15     procedure instantiate_object
16 end interface
17
18 contains
19
20 type(skeleton)  function  instantiate_object()
21     implicit none
22 end function
23
24 elemental subroutine clean_up(this)
25     implicit none
26     type(skeleton), intent(inout) :: this
27 end subroutine
28 end module skeleton_ADT

```

An ADT is a logical description of data and the sets of operations allowed on the data, without regards to how they will be implemented. In other words, an ADT provides an implementation-independent view of data. Usually, there are different implementations of an abstract data type. This allows for an easy switch of the various implementation without altering the way a user interacts with the data.

Listing 4.3 presents an abstract data type for some skeleton data. It shows various dummies to represent how the object is instantiated, destroyed, the operations permitted and the visible properties of the data type.

Listing 4.4: Data Type for Laplacian

```

1  module laplacian_ADT

```

```

2  implicit none
3  private
4  public   :: laplacian
5
6  type laplacian
7    private
8    integer                                :: order
9    logical                                :: free_surface
10   real, allocatable, dimension(:)      :: fd_coeffs
11   real, pointer, dimension(:,:)        :: field
12   real, pointer, dimension(:,:)        :: laplacian_2d
13   contains
14     procedure, public   :: compute_laplacian
15     procedure, private :: second_order_laplacian
16     procedure, private :: fourth_order_laplacian
17     procedure, private :: sixth_order_laplacian
18     procedure, private :: eighth_order_laplacian
19     final                :: clean_up
20 end type
21
22 interface laplacian
23   procedure constructor
24 end interface
25
26 contains
27
28 type(laplacian) function constructor(m_order, free_surf)
29   implicit none
30   integer, intent(in)  :: m_order
31
32   constructor%order = m_order
33   constructor%free_surface = free_surf
34
35   if (mod(order,2) .ne. 0) then
36     write (*,*) "Order_should_be_a_multiple_of_2"
37     write (*,*) "Order_has_been_given_a_default_of_2"
38     order = 2
39   end if
40
41   if (order .eq. 2) then
42     allocate (fd_coeffs(0:1))

```

```

43         constructor%fd_coeffs(0) = -2.0
44         constructor%fd_coeffs(1) = 1.0
45     else if (order .eq. 4) then
46         allocate (fd_coeffs(0:2))
47         constructor%fd_coeffs(0) = -2.6
48         constructor%fd_coeffs(1) = 1.4
49         constructor%fd_coeffs(2) = -0.1
50     else if (order .eq. 6) then
51         allocate (fd_coeffs(0:3))
52         constructor%fd_coeffs(0) = -2.8215452
53         constructor%fd_coeffs(1) = 1.57663238
54         constructor%fd_coeffs(2) = -0.18347238
55         constructor%fd_coeffs(3) = 0.01761260
56     else if (order .eq. 8) then
57         allocate (fd_coeffs(0:4))
58         constructor%fd_coeffs(0) = -2.97581692
59         constructor%fd_coeffs(1) = 1.70664680
60         constructor%fd_coeffs(2) = -0.25959423
61         constructor%fd_coeffs(3) = 0.04618682
62         constructor%fd_coeffs(4) = -0.00533093
63     end if
64 end function
65
66
67 subroutine compute_laplacian(this)
68     class(laplacian) ,intent(in)          :: this
69     integer                                :: dim1, dim2
70
71     dim1 = size(field, 1)
72     dim1 = size(field, 2)
73
74     if (order == 2) then
75         laplacian_2d = this%second_order_laplacian()
76     else if (order == 4) then
77         laplacian_2d = this%fourth_order_laplacian()
78     else if (order == 6) then
79         laplacian_2d = this%sixth_order_laplacian()
80     else if (order == 8) then
81         laplacian_2d = this%eighth_order_laplacian()
82     end if
83

```

```

84  end subroutine
85
86
87  function second_order_laplacian(this)
88      class(laplacian) ,intent(in)          :: this
89      real, pointer, dimension(:, :)        :: fld
90      real, pointer, dimension(:, :)        :: laplac_2d
91      integer                                :: dim1, dim2
92          integer                            :: i, j
93      real                                    :: a0, a1
94
95      a0 = this%fd_coeffs(0)
96      a1 = this%fd_coeffs(1)
97      do j = 2, dim2-1
98          do i = 2, dim1-1
99              laplac_2d(i, j) = 2.0*a0*fld(i, j) + a1*(fld(i-1, j) + &
100                  fld(i+1, j) + fld(i, j-1) + fld(i, j+1))
101          end do
102      end do
103
104      if (this%free_surface) then
105          do i = 2, dim1-1
106              laplac_2d(i, 1) = 2.0*a0*fld(i, 1) + a1*(fld(i-1, 1) + &
107                  fld(i+1, 1) + fld(i, 1))
108          end do
109      end if
110  end function
111
112
113  elemental subroutine clean_up(this)
114      implicit none
115      type(laplacian), intent(inout) :: this
116      if (allocated(this%fd_coeffs)) then
117          deallocate(this%fd_coeffs)
118      end if
119      if (associated(this%field)) then
120          nullify(this%field)
121      end if
122      if (associated(this%laplacian_2d)) then
123          nullify(this%laplacian_2d)
124      end if

```



```
125 |   end subroutine  
126 |  
127 | end module laplacian_ADT
```

Listing 4.4 shows an implementation of the Laplacian (which in a sense is a data type) following the skeleton shown in Listing 4.3.

Clean Code Approach

To allow for seamless collaboration and extension, a clean code approach was employed in the implementation of the testbed. Useful rules like naming all constants, minimising the use of global data, writing code that comments itself, providing global debugging switch, providing global type parameters, avoiding side effects as much as possible, preventing implicit typing etc provides a clean code base that is easy to maintain, extend and collaborate on.

4.3.3 Benchmarking

Each procedure added to the code base is tested and benchmarked against an established acceptance criterion. There is a desire to automate the testing and benchmarking of future procedure added to the code base.

4.4 Refactoring and Improvement

When a procedure is accepted as performing as intended, a minor refactoring is applied when necessary. Also, as more of the 2003 and 2008 specification got supported by compilers, minor refactoring were performed to incorporate some very necessary features. It is the desire to make the code base fully OOP in the future.

Chapter 5

Synthetic Data Examples

5.1 Introduction

This chapter discusses some applications of FDEWD to synthetic datasets. The example examines a reflection dataset lacking in low frequencies (the inversion has been performed for frequencies between 5 Hz and 10 Hz in one experiment, and between 3 Hz and 10 Hz in the other experiment). Two starting models are provided to invert this dataset. The starting models are highly smoothed version of the true model, with some scaling applied to the smoothed model. The first starting model scales the values below the water layer by a factor lower than unity while the second starting model scales the values below the water layer by a factor greater than unity.

The results from these experiments are compared with the results from conventional FWI to show areas of improvements. Accuracy of these results are analysed using sample well logs at selected positions, normalised cross-correlation values, data residuals and model residual plots.

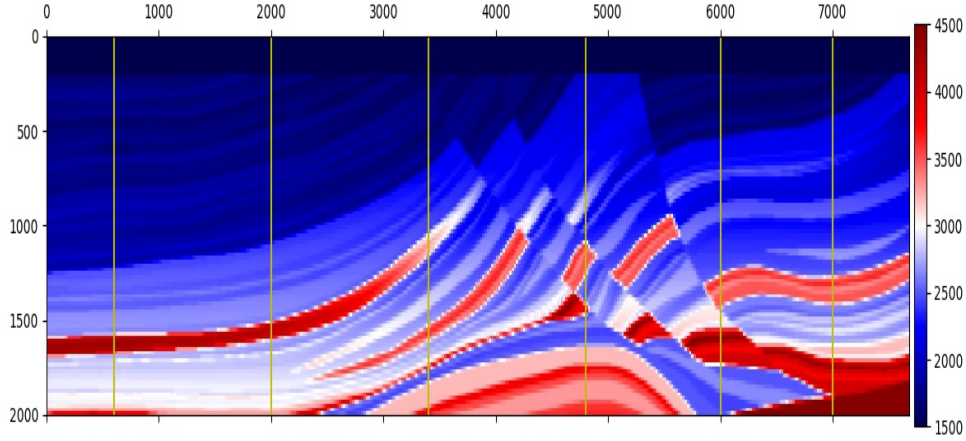


Figure 5.1: The true velocity model with the locations of the well logs. The well logs will be used to analyse the accuracy of the inverted model

5.2 Problem Definition

The Marmousi model (Versteeg (1994)) was designed to be geologically plausible. It contains many reflectors, steep dips and strong velocity gradients both laterally and vertically (figure 5.1). Datasets generated from the Marmousi model contains useful transmission energies (figure 5.2) deep enough to invert the central parts of the model (if large offsets are low frequencies are considered).

In figure 5.2, a sample shot gather for a source propagated through the true Marmousi model, with surface source and surface receivers configuration, is shown. There is a rich supply of reflection, transmission and back scattered energies. For a source propagated through a smoothed version of the model (figure 5.3), there are still present very useful supply of transmission energies. However, the lack of scatterers in the smooth model implies a lack of reflections and back scattered energies. We can expect conventional multiscale FWI, with low frequencies and long offsets, to perform well in inverting this data starting from such smooth model. This is because the initial stages of the inversion will employ the rich sup-

plies of transmission energies and low frequencies in building the low wavenumber components of the model early on in the inversion process.

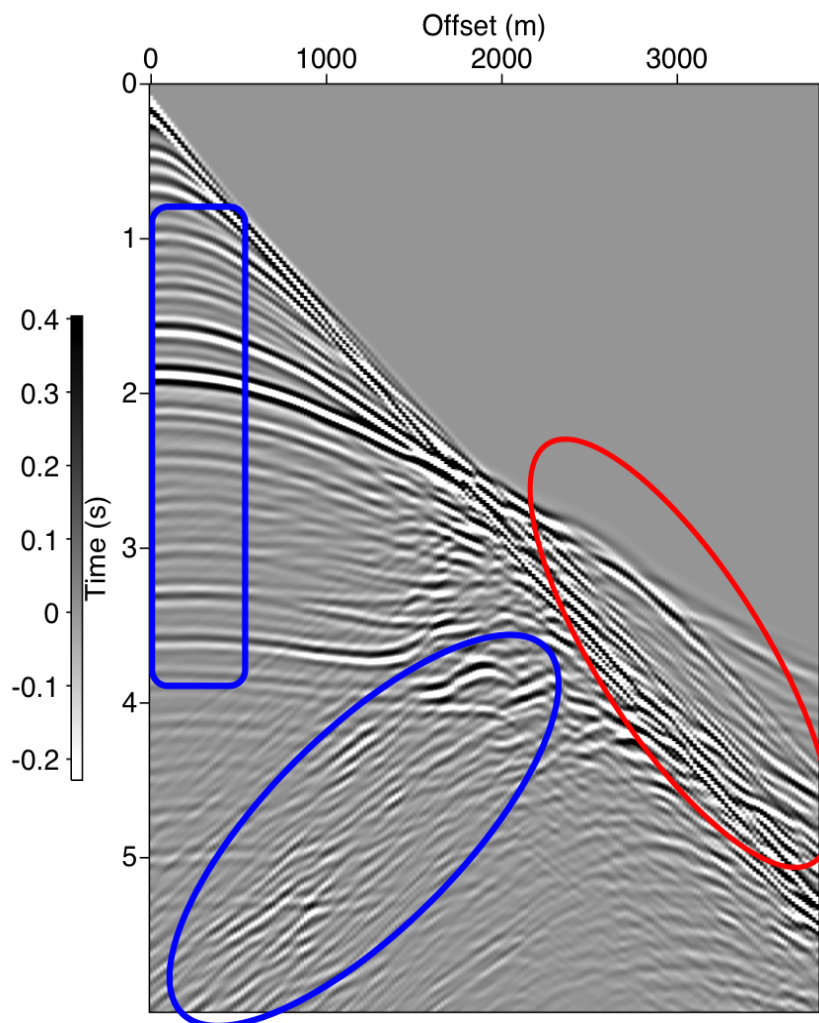


Figure 5.2: A sample shot data generated from the true model. The presence of different energy modes makes this data very friendly for seismic inversion via the multiscale FWI

A much difficult challenge is with a dataset lacking in long offsets and low fre-

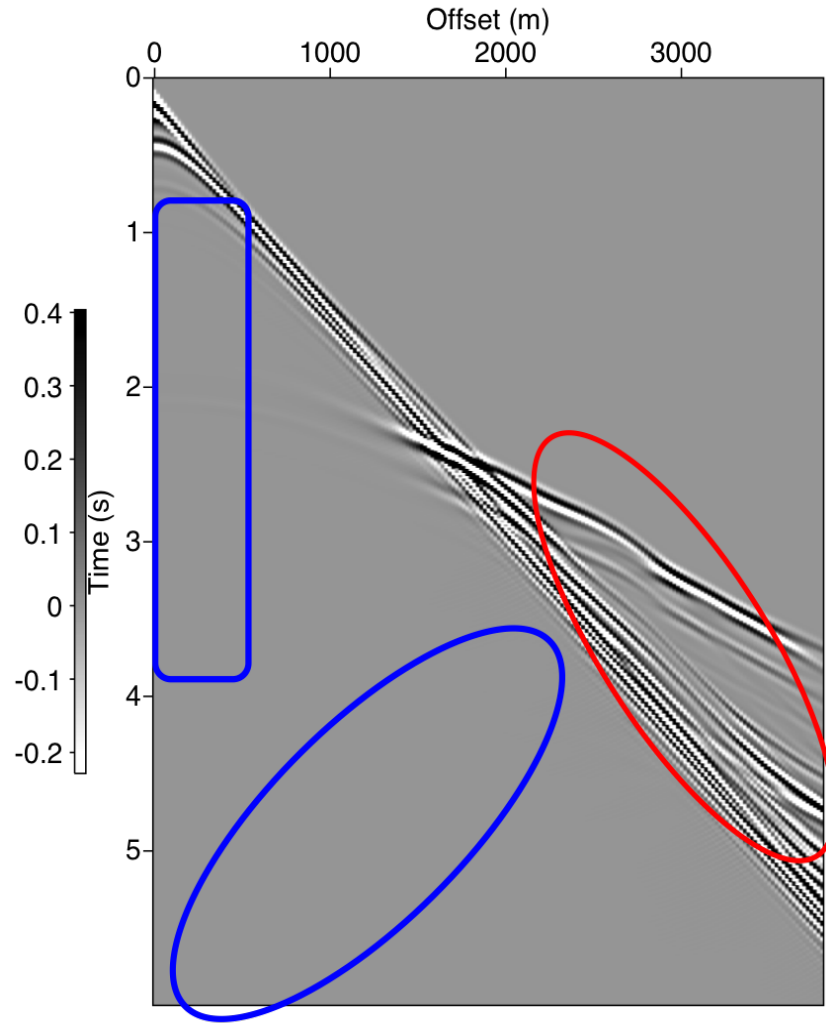


Figure 5.3: A sample shot data generated from a smooth Marmousi model. Though lacking reflections and back-scattered energies, the transmission energies present in the data still make the model a suitable starting model for multiscale FWI.

quencies, but compensated by a reasonable starting model, albeit bad enough for conventional FWI to fail. The chapter analyses this challenge and proffer a novel inversion scheme to overcome such challenges via the FDEWD.

The data

The data to be inverted has been constrained to be reflection dominated. It has been modelled with a 10Hz Ricker wavelet on a 20 m mesh, using an eighth order in space and second order in time finite difference code. The data is sampled at 2 ms, and contains free-surface multiples. Figures [5.4](#) to [5.14](#) show samples of the shot data

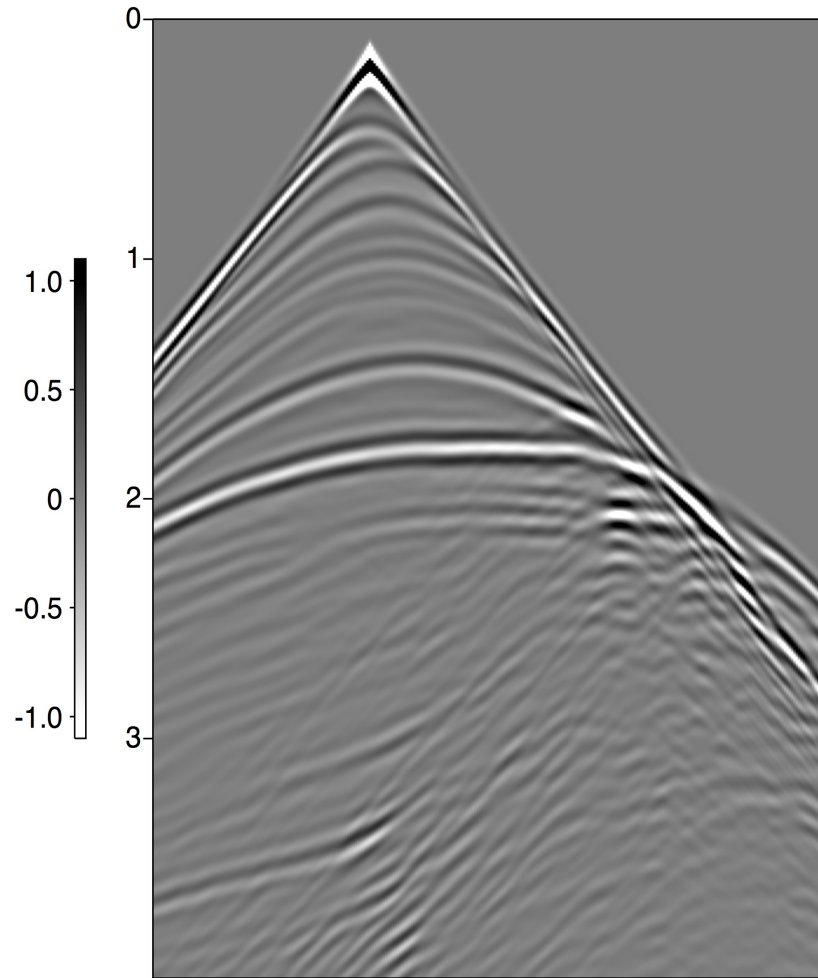


Figure 5.4: *A sample shot data from the true Marmousi model*

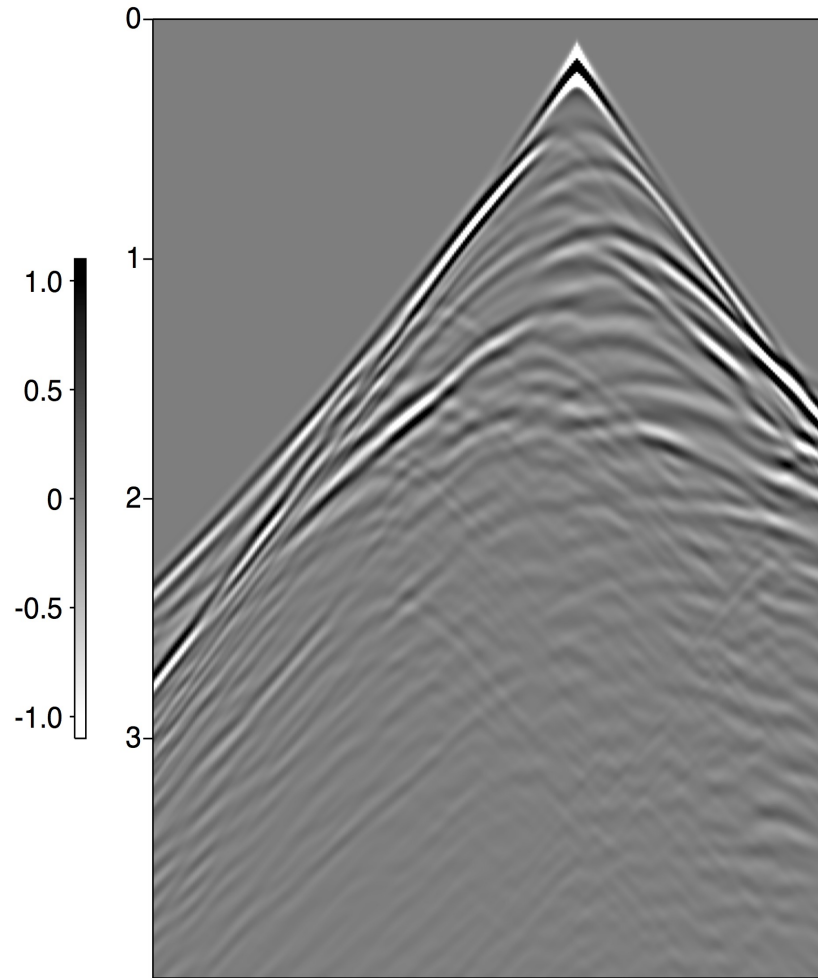


Figure 5.5: *A sample shot data from the true Marmousi model*

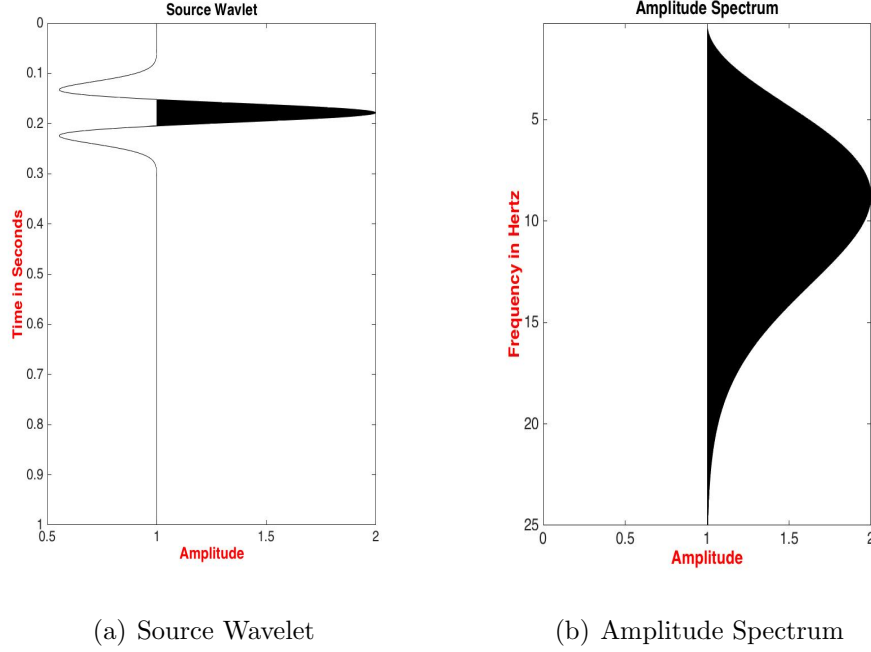


Figure 5.6: Source wavelet and its amplitude spectrum

The source wavelet

The source wavelet is the exact wavelet used in simulating the data. It is a Ricker wavelet centred at 8.5Hz (figure 5.6) and sampled at 2ms.

Acquisition geometry

The receivers in the experiments are fixed for all shots, and placed on every grid node 40 m from the free surface. The minimum offset used in these experiments is 0 m, while the maximum offset is 3840 m. Though the spread of each shot gather changes from end-on in some cases to split-spread in other cases, however, the maximum offset in all cases is limited to 3840 m.

The start models

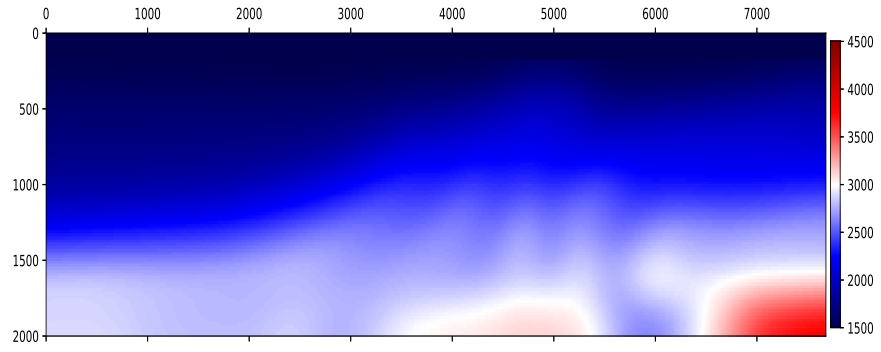
In this section, three (3) velocity models are discussed and compared in presenting the starting models used for the experiments in this chapter.

1. Model-A, a **very good start model**: This represents a model built from the true Marmousi model by gently applying a Gaussian filter on it. The model represents the true background velocity model. It can be thought of as the velocity model that correctly propagates the low frequencies of the source wavelet.
2. Model-B, a **low (slow) background start model**: This represents a model built from the very good start model by harshly applying a Gaussian filter, and a less than one (0.9) bulk shifting to it. The expectation here is to ensure that the low frequencies are cycle-skipped.
3. Model-C, a **high (fast) background start model**: This model is built in a similar way as the low background start model from the very good start model. First a Gaussian filter is harshly applied to the very good start model. Afterwards, a greater than one (1.05) bulk shifting is applied to it. The expectation here also is to ensure that the low frequencies are cycle-skipped.

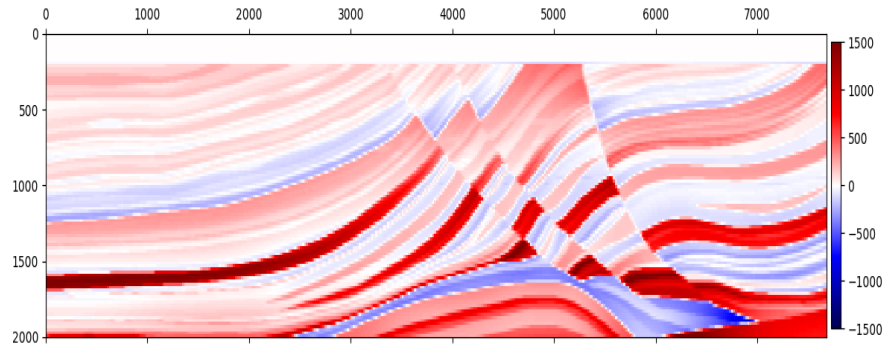
The experiments discussed in this chapter, however, make use of two of the three models above (Model-B and Model-C). As discussed above, the models are highly smoothed version of the true model. Also, they have been scaled such that in the first case (low background start model), the average background model is lower than the true background model by approximately 10% (Figure 5.7) and in the second case (high background start model), the average background model is higher than that of the true background model by approximately 5% (5.8). The models measure 7680 m by 2000 m, and were sampled on a 20m grid.

Figure 5.7b presents the model residual (difference between the true model and the low background start model) for the low background start model. This is computed by subtracting the low background start model from the true model. It can quickly be observed that indeed the bulk velocities for the start model are

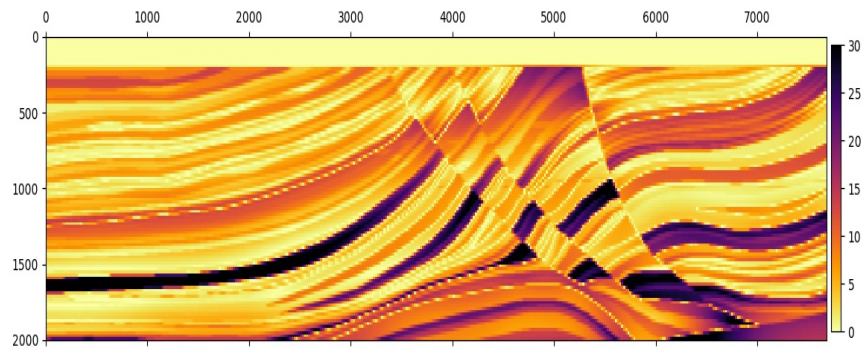
5.2 Problem Definition



(a) Model-B: Low Background Start Model



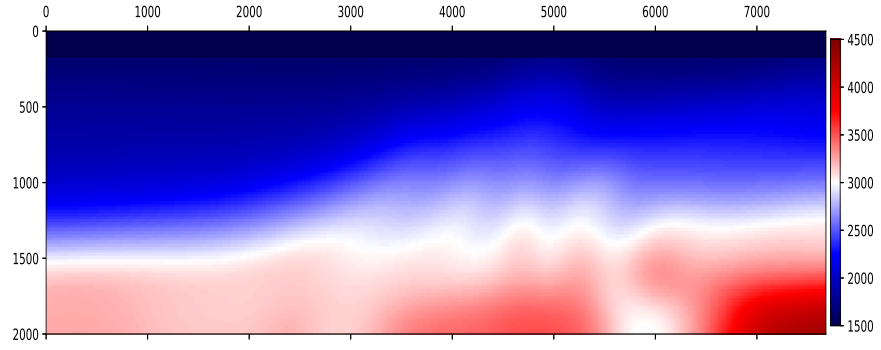
(b) Model Residuals



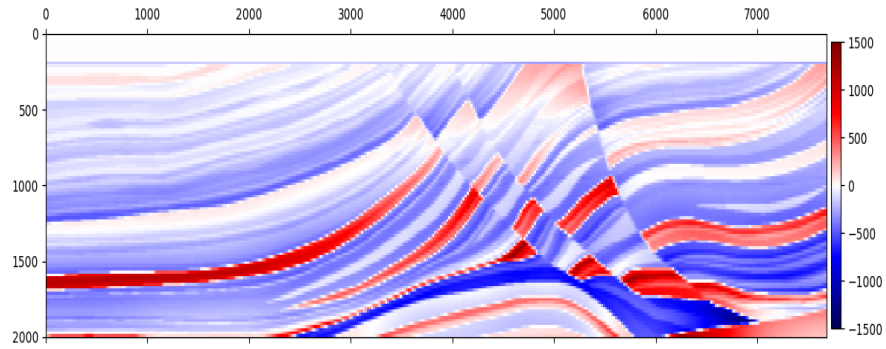
(c) Model Residuals as Percentage

Figure 5.7: *Low Background Start Model (Model-B), and its residuals*

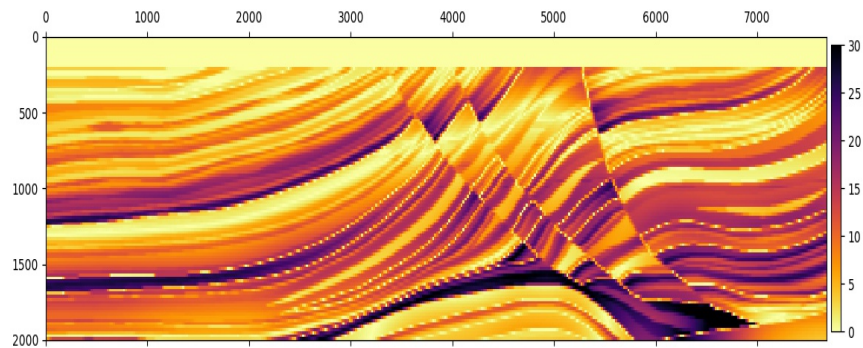
5.2 Problem Definition



(a) Model-C: High Background Start Model



(b) Model Residuals



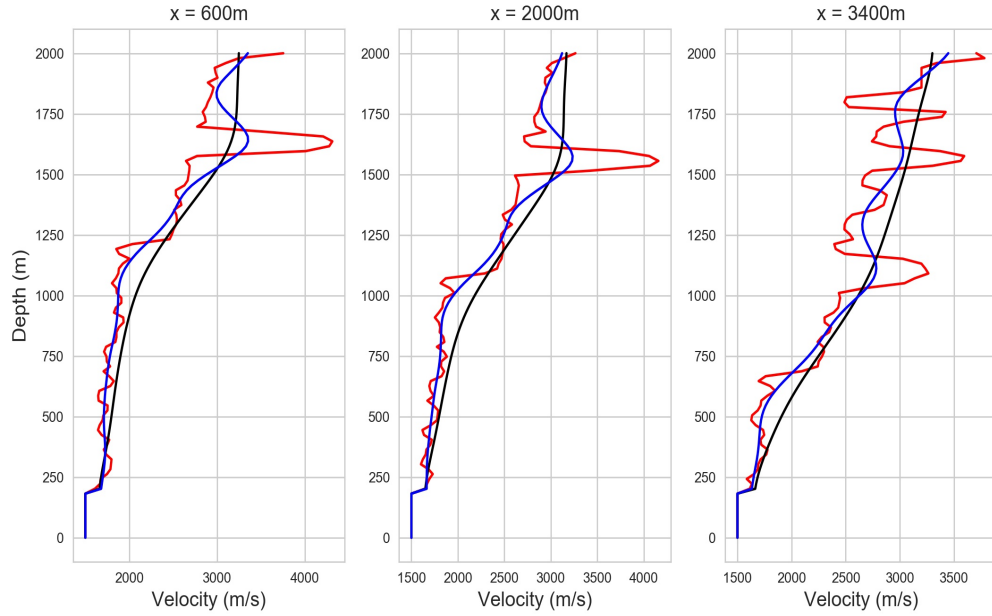
(c) Model Residuals as Percentage

Figure 5.8: High Background Start Model (Model-C), and its residuals

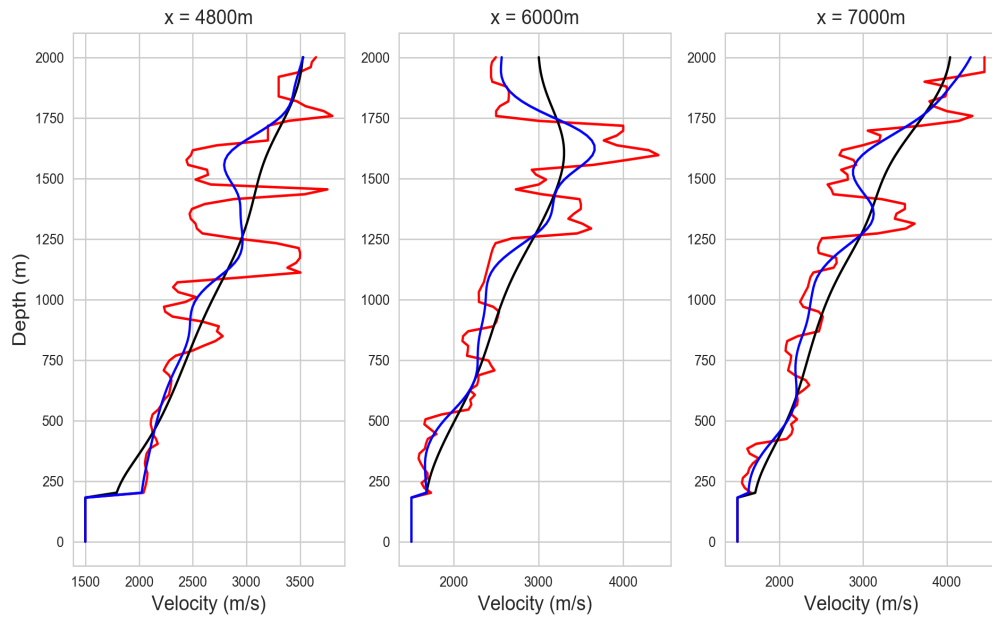
lower than those for the true model. Similarly, a model difference for the high background start model (figure 5.8) is shown in figure 5.8b.

Figures 5.9 and 5.10 show quantitative well log comparisons between each start model and the very good start model. Well logs are taken from the left and right sides of the model for this purpose. From the well logs, it is evident that the start model in each case only contains the very low wavenumbers and is lacking the intermediate wavenumbers necessary for building the higher wavenumbers during the inversion process. The lack of low frequencies in the inversion excludes the application of conventional FWI, as conventional FWI depends largely on the availability of low frequencies and a fairly good start model for successful inversion results.

5.2 Problem Definition



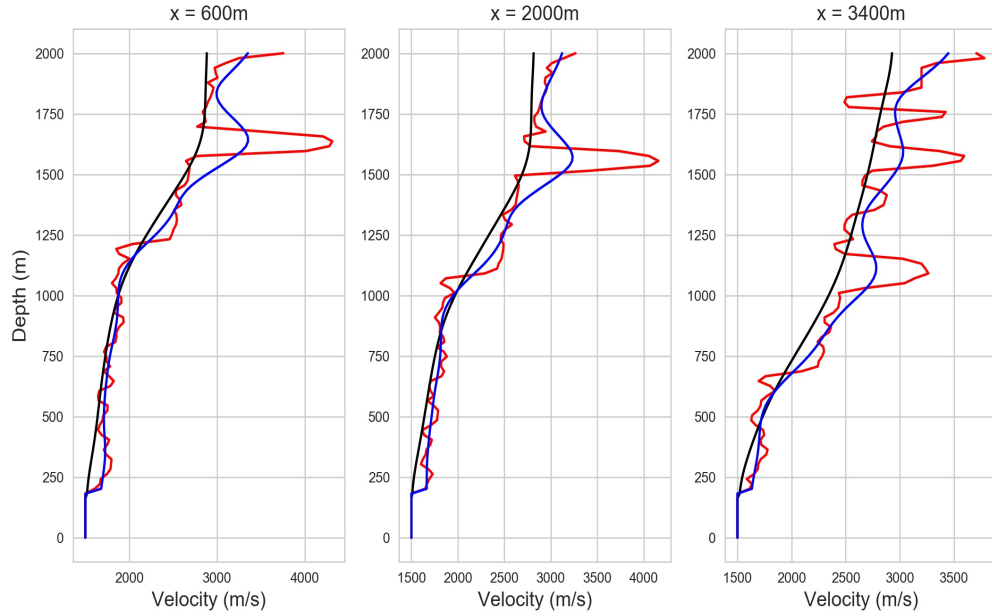
(a)



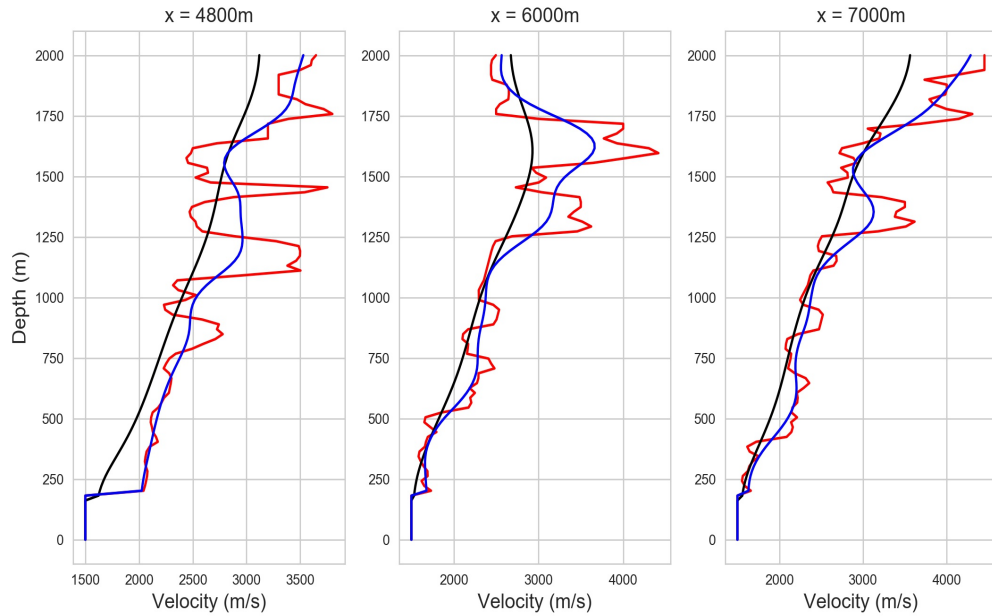
(b)

Figure 5.9: Comparisons of well logs from the true model, true background model (Model-A) and high background start model (Model-C). Red logs represent the true Marmousi model. Blue logs are the very good start model. Black logs represent the high background start model.

5.2 Problem Definition



(a)



(b)

Figure 5.10: Comparisons of well logs from the true model, true background model (Model-A) and low background start model (Model-B). Red logs represent the true Marmousi model. Blue logs are the very good start model. Black logs represent the low background start model.

The start data

Figures 5.11 and 5.13 present sample shot gathers from the high background velocity start model while Figures 5.12 and 5.14 present their respective data residuals. The start data presents very minimal useful energies to kick-start conventional multiscale FWI. The problem clearly gets worse as the experiments discussed in this chapter begin from 3Hz to 5Hz. The data residuals show an abundance of reflections and scattered energies. The inversion will depend on these reflections and scattered energies to invert the model. Building low wavenumbers from reflections is not trivial. The next section discuss some of the schemes proposed in this work to build low wavenumbers from reflections and scattered energies.

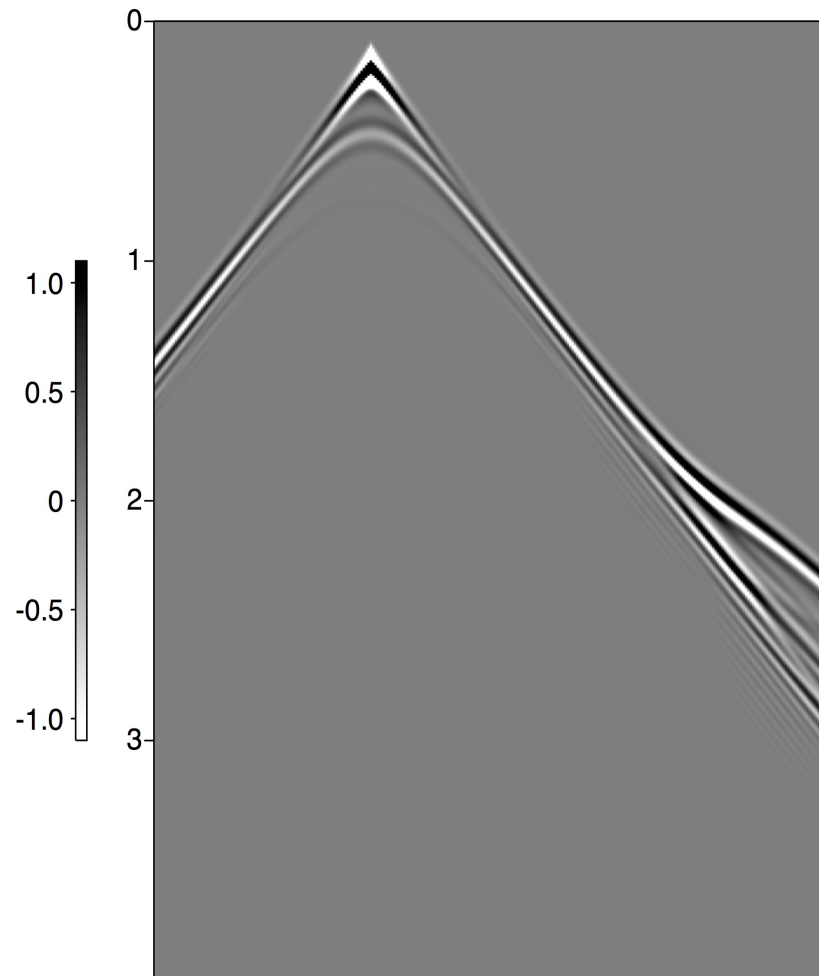


Figure 5.11: *Sample predicted data from the smooth, high background start model*

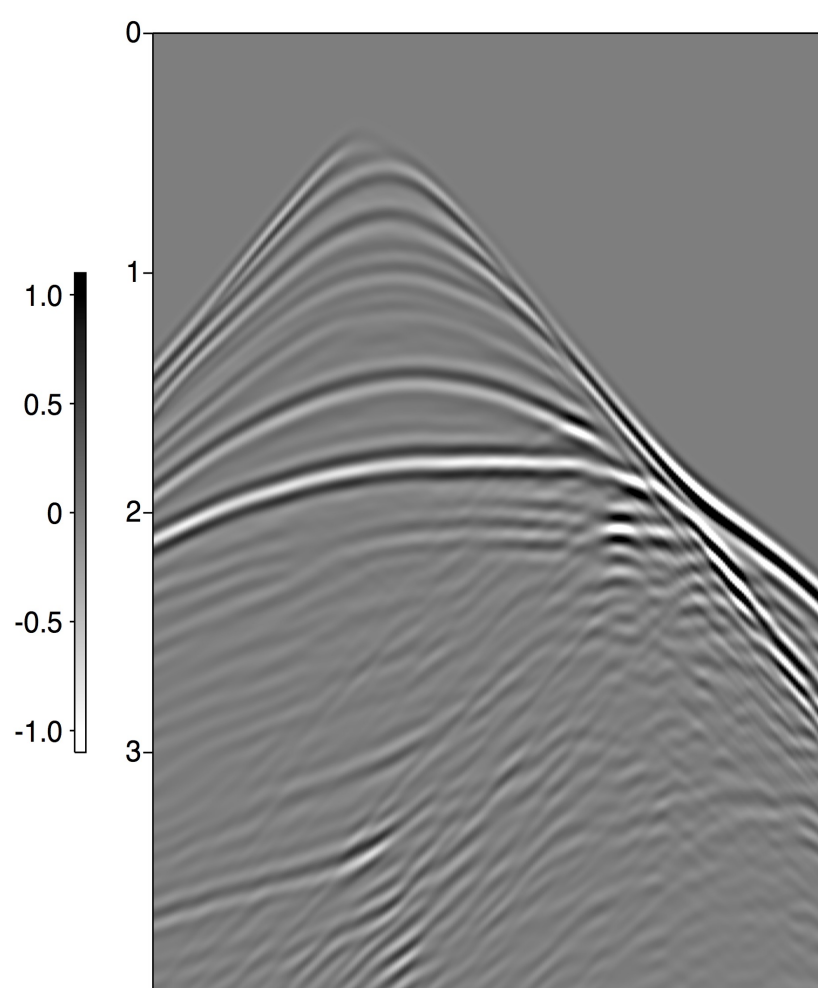


Figure 5.12: *Data residuals for the previous predicted data*

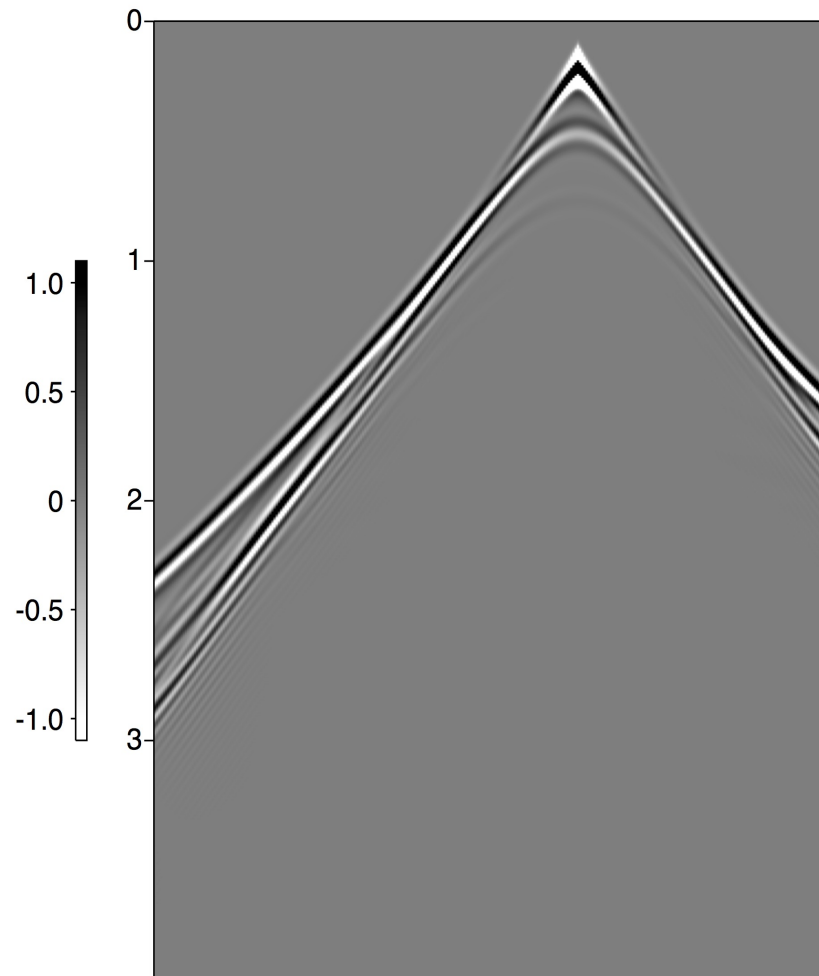


Figure 5.13: Another sample predicted data from the smooth, high background start model

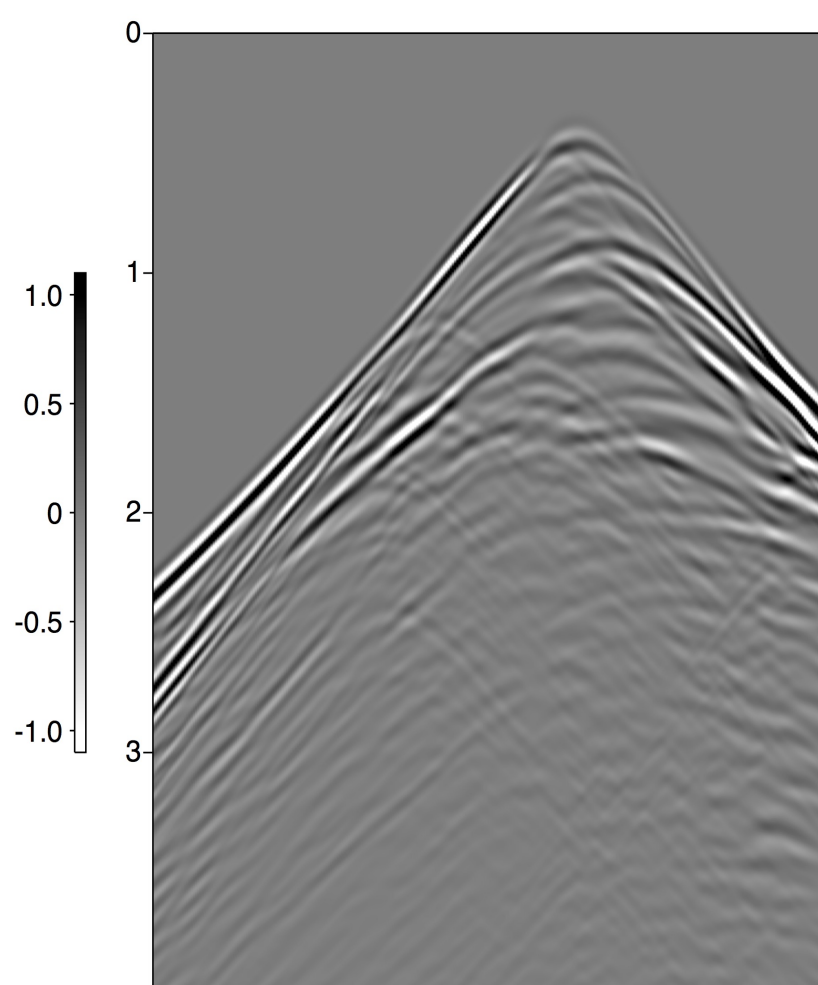


Figure 5.14: Data residuals for the previous predicted data

5.3 Experiment 1 : Reflection FWI with a low background velocity

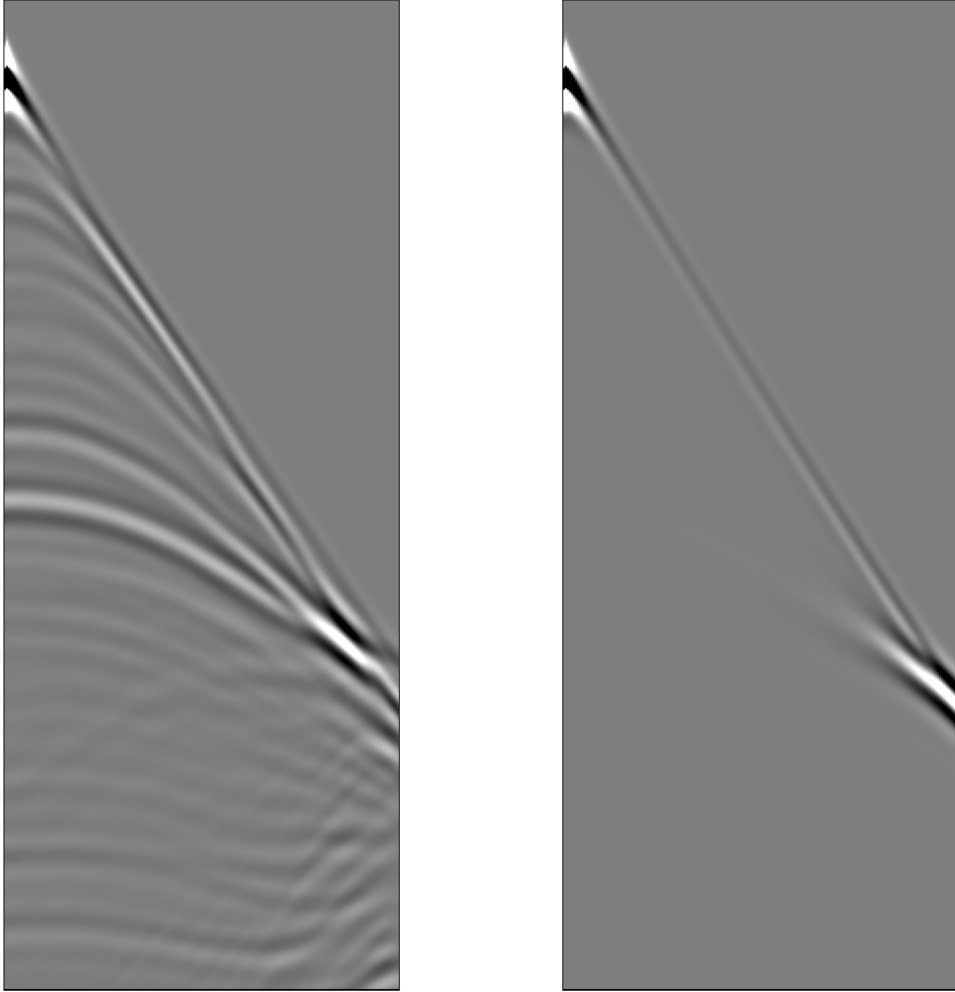
The starting model for this experiment is described in the previous section as a smoothed version of the true model, with a bulk shifting applied using a factor of 0.9.

5.3.1 Inversion procedures and strategies

The inversion strategies and schemes used in these reflection experiments are combinations of the MBTT ([Chavent and Clément \(1993\)](#)) and multiscale ([Bunks et al. \(1995\)](#)) FWI schemes. In the first scheme used in this project, the inversion proceeds by iterating between low wavenumber (tomographic) components and the high wavenumber components of the gradient, generated using ONLY the vertical components of the wavefields. The high wavenumber model updates are initially injected into the model. These are expected to be wrong in amplitudes and positions. Their sole purpose is to generate back-scattered and reflection energies which are used in computing the low wavenumber updates. Two novel additions are introduced in this scheme compared to the conventional implementation of the MBTT scheme ([Xu, Chen, Lambaré, Zhang, Wang et al., 2012](#)). First, we compute the high and low wavenumber updates using only the vertical components of the wavefields. Second, we discard the high wavenumber updates after each frequency run. The inversion procedure is summarised in the scheme below

1. Start from a smooth model
2. Correlate wavefields propagating in opposite directions $\rightarrow$ high wavenumber update
 - One iteration with an exaggerated step length

5.3 Experiment 1 : Reflection FWI with a low background velocity



(a) Observed data

(b) Synthetic data

Figure 5.15: Comparison between shot data

5.3 Experiment 1 : Reflection FWI with a low background velocity

3. Correlate energies propagating in the same direction $\rightarrow$ low wavenumber update
 - Ten iterations with an accurate step length compute
 - Remove rough updates on completion
4. Do one iteration step 1 with true step-length
5. Repeat steps 2 - 4 Increase Frequency
 - Do this for ten frequencies between 5 and 10Hz
6. Apply gentle Gaussian low-pass filter
7. Conventional FWI
 - 60 iterations from 5 to 10Hz

In the second scheme used in this project (Chapter 6), we proceed to initially using the horizontal wavefield components. Because these are mainly turning waves, their depths of penetration are limited to the shallow regions. When the background velocities of the shallow regions have been built using the horizontal components, we proceed to using the vertical wavefields components as described in the scheme above. This scheme is applied to the blind data experiment and detailed in chapter six of this thesis.

5.3.2 Results from the low (slow) background starting model

Initial inversion updates

Figure 5.16a presents the first high wavenumber update used to build scatterers in the model (step 2 in the strategy presented in the previous section), while 5.16b presents the first low wavenumber update of the inversion process (step 3 in the inversion strategy). As described in the strategy, the step length for initially updating the high wavenumber components of the gradient has been exaggerated

5.3 Experiment 1 : Reflection FWI with a low background velocity

(usually by a factor between 2.5 and 3.5 of its true value). This has been done to cause the rough part of the model to be strong enough to generate scattered wavefields required to build the low wavenumber updates in the subsequent iterations at that frequency.

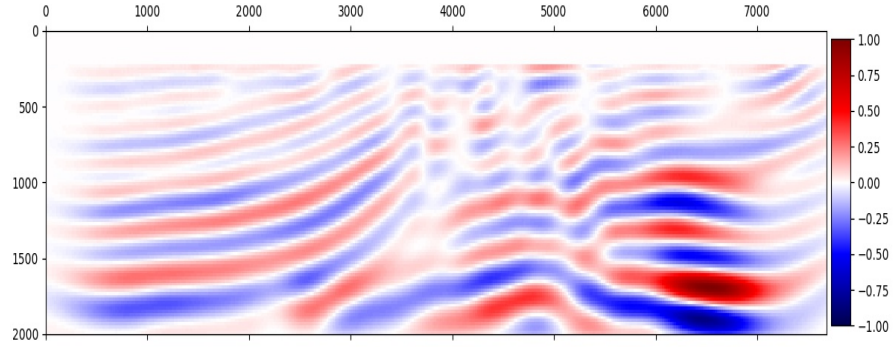
The low wavenumber updates, even at 5Hz, contain only wavenumbers that contribute to building the background velocities of the model. By using only the low wavenumber components of the gradient initially, the inversion systematically builds the background model in a wavenumber continuation fashion, (Alkhalifah, 2016). Thus the method is more robust compared to the convention FWI.

Figure 5.18 presents the updated velocity model after completing the runs of using only the low wavenumbers in the inversion (step 5 of the inversion strategy), discarding any high wavenumber component was initially injected into the model. Quantitative analysis of the model is aided by Figure 5.17, which presents well logs from this updated model. The distribution of wavenumbers present in the well logs suggests that at this stage of the inversion, only the low to intermediate wavenumbers are present in the velocity model. Well logs from the right part of the model (Figure 5.17b) shows a close match between the low to intermediate wavenumbers present in the model and those recovered so far by the inversion.

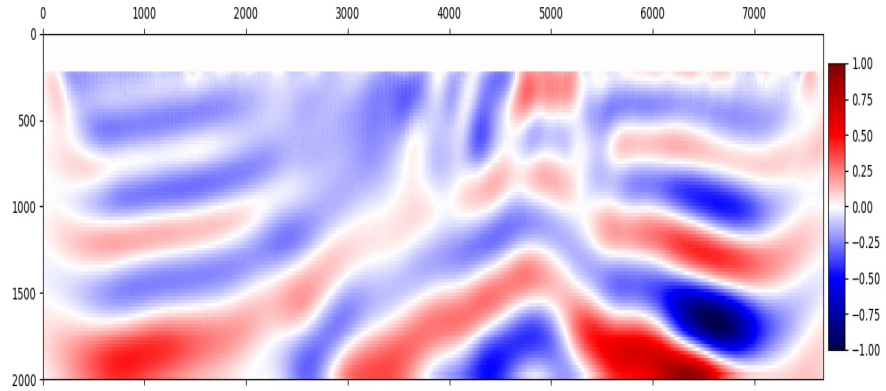
Final inversion results

Figure 5.19 presents the final inverted model, whilst Figure 5.20 present well logs from the final inverted model. Conventional FWI, starting from the model inverted with the tomographic components of the gradient via FDEWD, has successfully recovered the right side of the velocity model. The inversion, however, hasn't performed quite successfully on the leftmost part of the model due to limited energy coverage. The success of the inversion is highly attributed to the ability of FDEWD to preferentially inject only the low to intermediate wavenumbers into the into. This allows for the higher wavenumbers to be accurately placed, together with their correct amplitudes.

5.3 Experiment 1 : Reflection FWI with a low background velocity



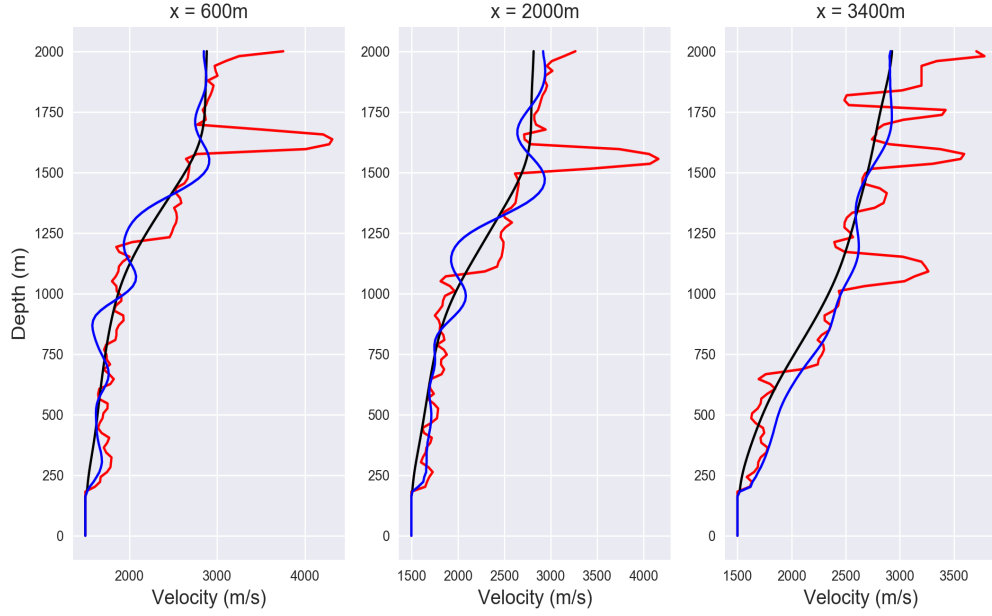
(a) High wavenumber Update



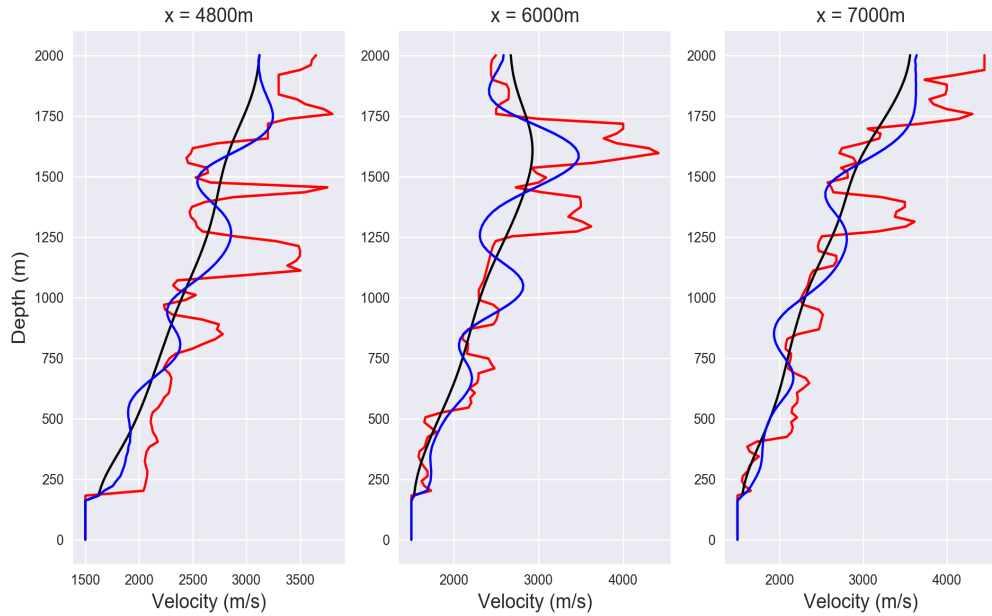
(b) Low wavenumber Update

Figure 5.16: Initial high and low wavenumber updates. The low wavenumber update contains the wavenumbers needed to build the background model. The high wavenumber update is cycle-skipped and does not contribute to building the background model directly

5.3 Experiment 1 : Reflection FWI with a low background velocity



(a) Logs from the left part of the model



(b) Logs from the right part of the model

Figure 5.17: Well logs from the *Intermediate* result, compared with logs from the start model (*Model-B*). Black represents the start model, Red represents the true model, while Blue represents the updated model using the tomographic components of the gradient.

5.3 Experiment 1 : Reflection FWI with a low background velocity

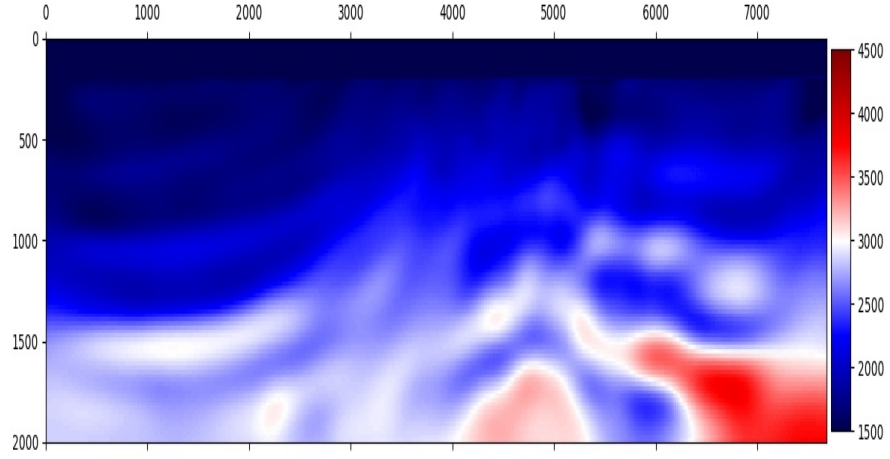


Figure 5.18: Results obtained by doing tomographic updates of the model starting from 5Hz following the scheme outlined in this section

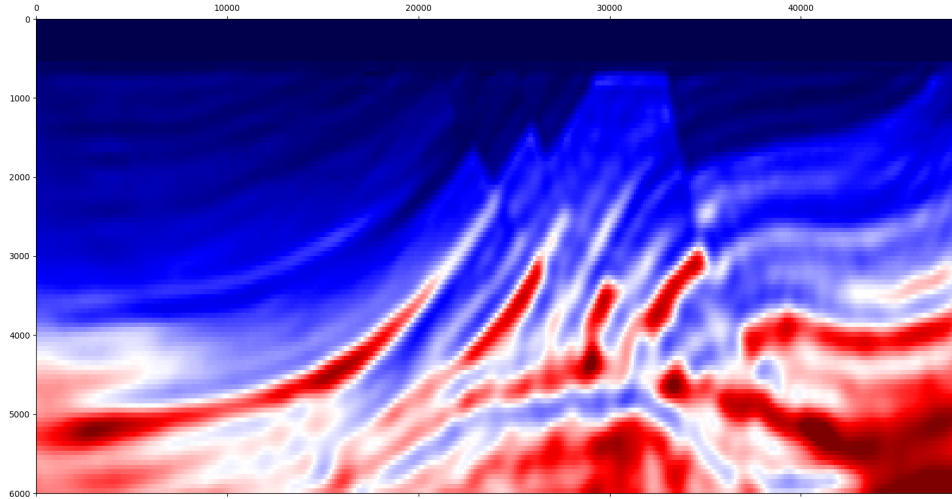
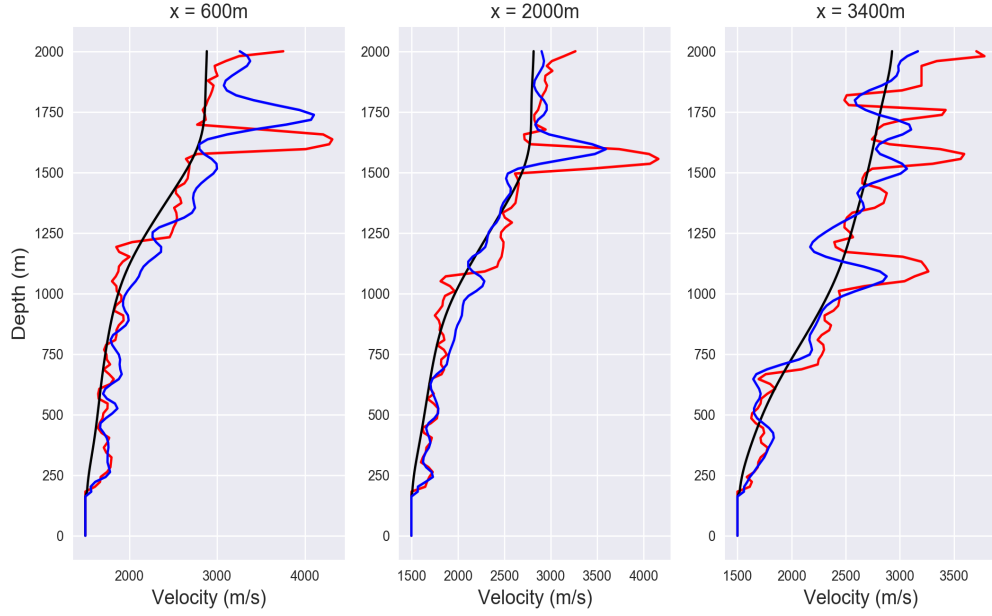
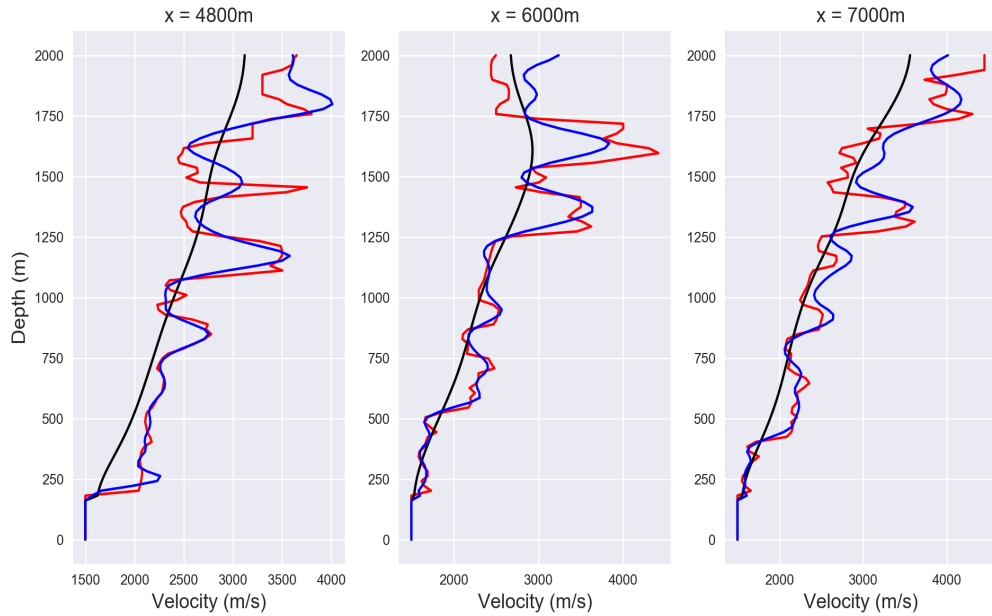


Figure 5.19: Final results obtained using finite difference explicit wavefield decomposition and conventional FWI. Colour scale same as in Figure 5.18

5.3 Experiment 1 : Reflection FWI with a low background velocity



(a) Logs from the left part of the model



(b) Logs from the right part of the model

Figure 5.20: Well logs from the final inversion result, compared with logs from the start model (Model-B). Black represents the start model, Red represents the true model, while Blue represents the final updated model.

5.4 Experiment 2 : Reflection FWI with a high background velocity

The starting model for this experiment is described in the previous section as a smoothed version of the true model, with a bulk shifting applied using a factor of 1.05. This is a slightly more difficult problem compared to the low velocity start model. Hence, the experiment has commenced from 3Hz.

5.4.1 Inversion procedures and strategies

The scheme used in this experiment is similar to the low velocity case.

1. Start from a smooth model
2. Correlate wavefields propagating in opposite directions $\rightarrow$ high wavenumber update
 - One iteration with an exaggerated step length
3. Correlate energies propagating in the same direction $\rightarrow$ low wavenumber update
 - Ten iterations with an accurate step length compute
 - Remove rough updates on completion
4. Do one iteration step 1 with true step-length
5. Repeat steps 2 - 4 Increase Frequency
 - Do this for ten frequencies between 3 and 8Hz
6. Apply gentle Gaussian low-pass filter
7. Conventional FWI
 - 60 iterations from 5 to 10Hz

5.4 Experiment 2 : Reflection FWI with a high background velocity

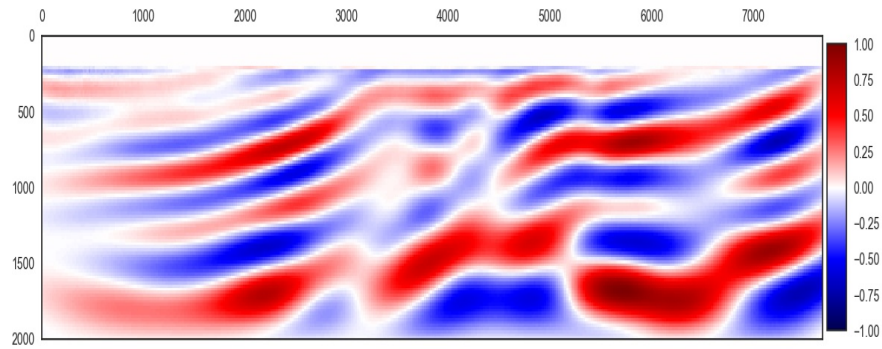
5.4.2 Results from the high (fast) background starting model

Initial results from the high background starting model

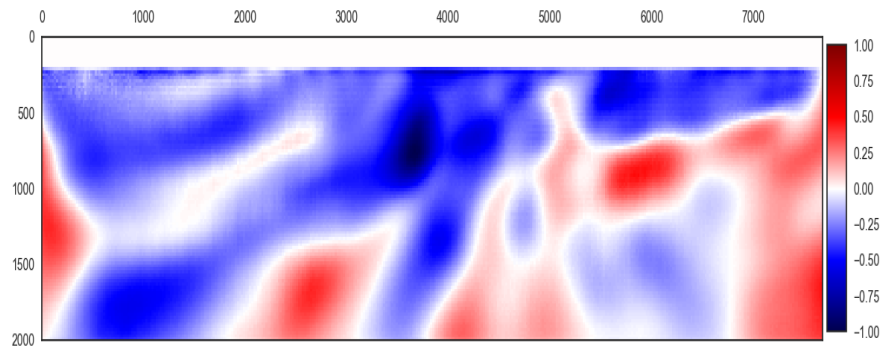
The initial results from the low wavenumber components of the update are shown in Figures 5.22 and 5.23. These results show well logs from the true model, Model-A, Model-C (used as the starting model to drive the inversion) and the recovered velocity model. See section 5.2 for a review of the model definitions.

In Figure 5.22, a comparison of the tomographic result and the start model is shown. In general, the inversion has been dominated by low to intermediate wavenumbers, as evident on the logs. These tomographic updates, injecting the necessary low to intermediate wavenumbers into the model, are necessary for the natural wavenumber continuation recipe required for successful FWI. The accuracy of the positions of these updates can be assessed by comparing the tomographic updates with Model-A (Figure 5.23). The well logs from the left side of the model, dominated by low wavenumbers, match more closely compared to the well logs from the right side of the model. It will be noted that the right side of the Marmousi model has structures, with dips varying in degrees and directions.

5.4 Experiment 2 : Reflection FWI with a high background velocity



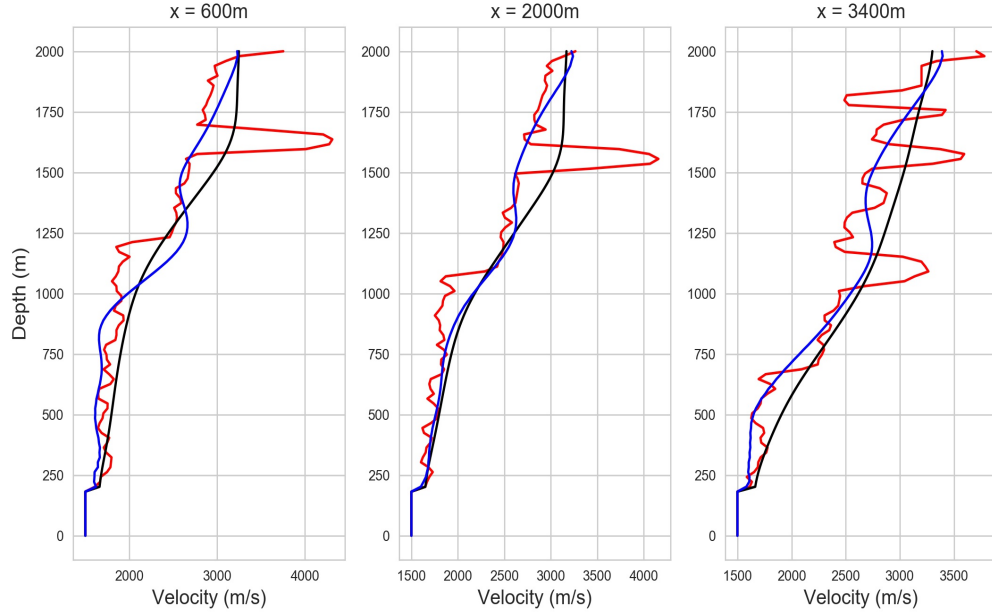
(a) High Wavenumber Update



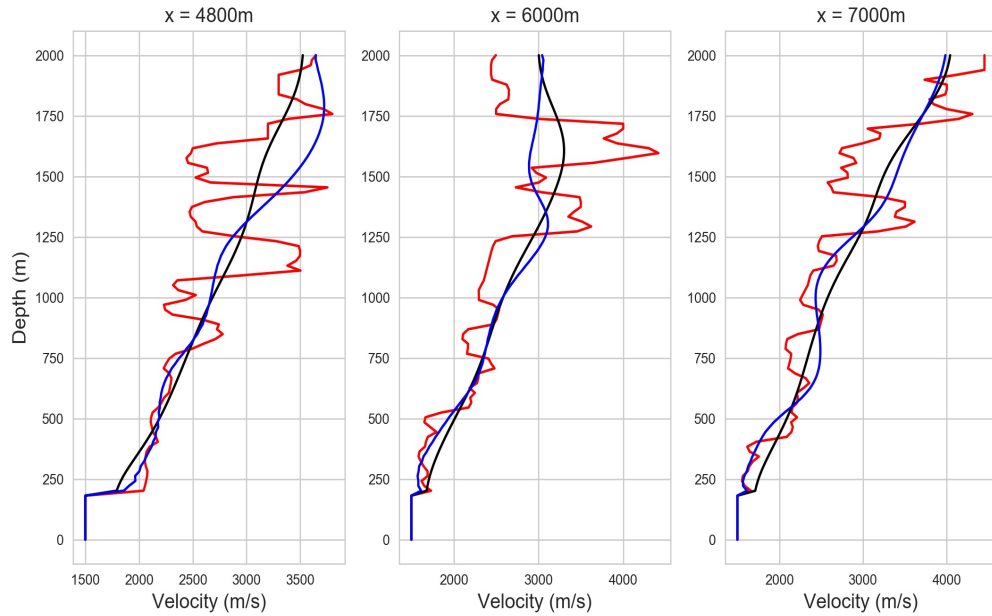
(b) Low Wavenumber Update

Figure 5.21: Initial high and low wavenumber updates. The high wavenumber updates are initially injected into the model. These then generate scattered energies used in building the low wavenumber updates during subsequent iterations.

5.4 Experiment 2 : Reflection FWI with a high background velocity



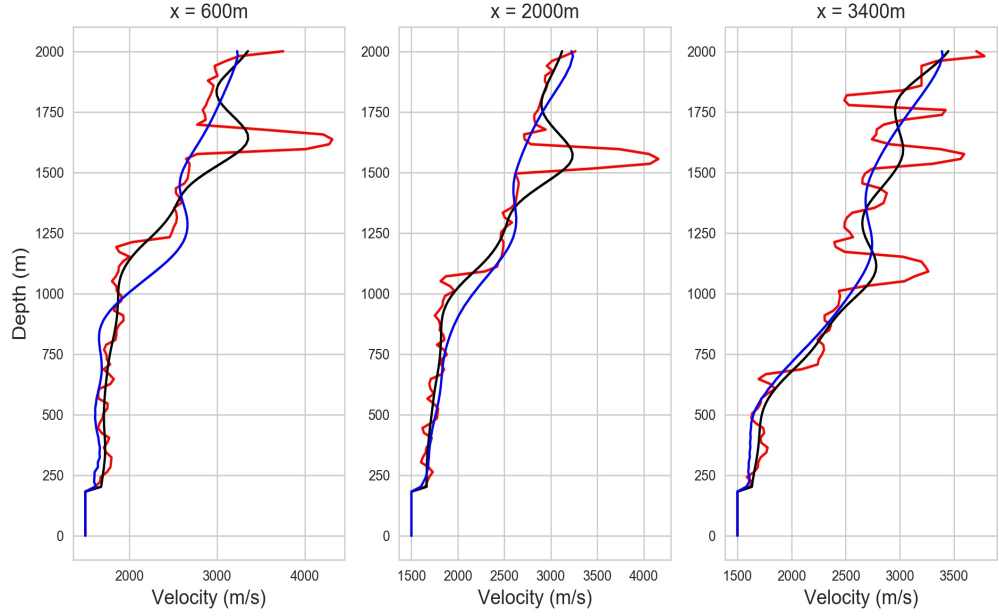
(a) Logs from the left part of the model



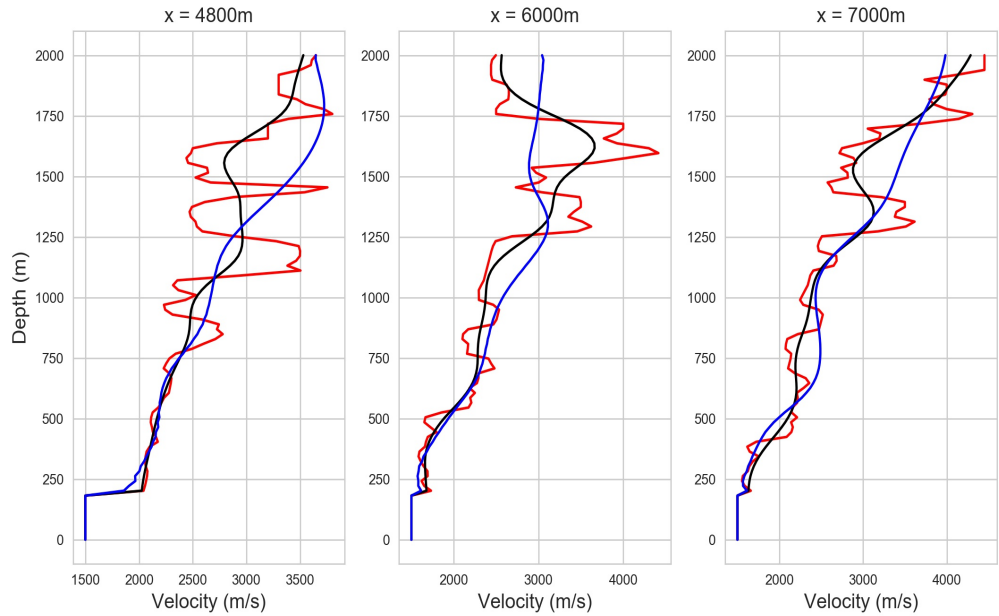
(b) Logs from the right part of the model

Figure 5.22: Well logs from the initial result (3 to 6Hz), compared with logs from the start model (Model-C). Black represents the start model, Red represents the true model, while Blue represents the updated model. The surface location of each log is shown on the log.

5.4 Experiment 2 : Reflection FWI with a high background velocity



(a) Logs from the left part of the model



(b) Logs from the right part of the model

Figure 5.23: Well logs from the initial result (3 to 6Hz), compared with logs from the very good start model (Model-A). Black represents the the very good start model, Red represents the true model, while Blue represents the updated model. The surface location of each log is shown on the log.

5.4 Experiment 2 : Reflection FWI with a high background velocity

Final results from the high background starting model

The final model obtained using the migration-demigration strategy explained in the previous section is shown in Figure 5.25, while well logs comparisons are shown in Figure 5.24. The inverted model shows a recovery of both the residual low wavenumbers missing from the previous stage (Figure 5.26), and the required high wavenumbers. The amplitude of the recovered model also match those of the low wavenumbers. Sample shot data generated from the final model are shown in Figures 5.27, 5.34 and 5.35.

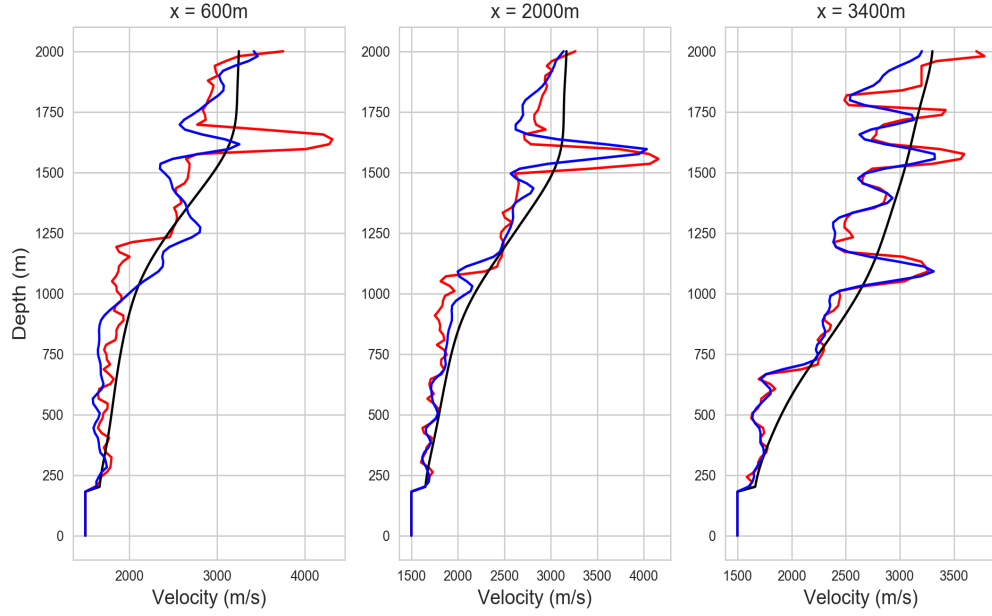
The results obtained show that finite difference explicit wavefield decomposition, combined with full waveform inversion, has recovered both the low wavenumber and the high wavenumber components of the model. The data modelled from the recovered model predicted offset that were never included in the inversion, suggesting the robustness of the method and the accuracy of the results.

Traces from the start, recovered and true models are also compared in Figures 5.28 to 5.33, for a sample shot from the right side of the model; and Figures 5.36 to 5.41, for a sample shot gather from the left side of the model.

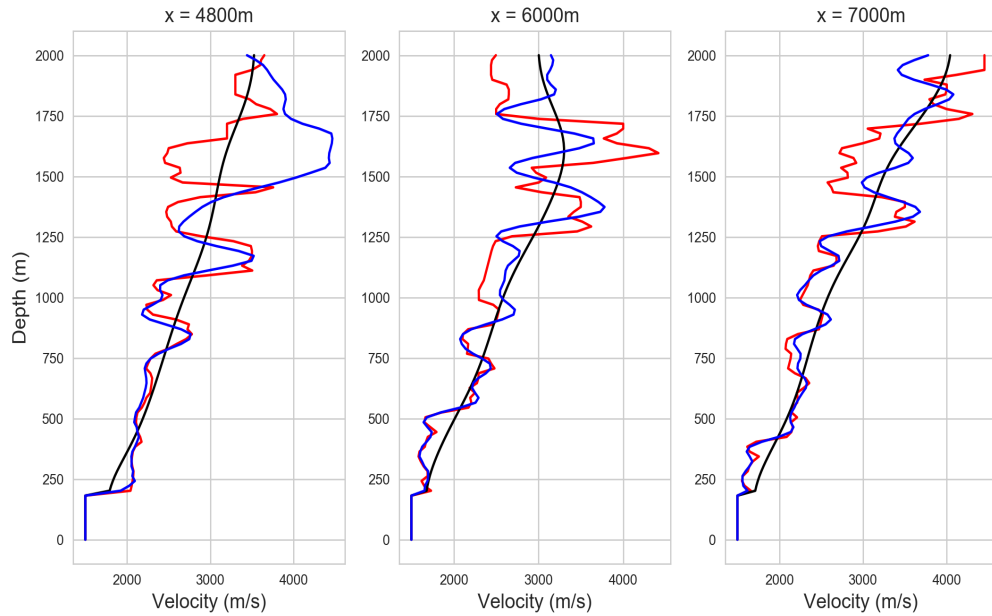
In both scenarios, the near traces show that only the direct arrivals are matched for the start model. However, for the recovered model, all arrivals seem to match in amplitudes and phases. The mid traces show an interesting situation when the phases and amplitudes that were mismatched in the start model are subsequently matched in the recovered model.

Finally, the near traces also show a close and good match for the recovered and true model traces.

5.4 Experiment 2 : Reflection FWI with a high background velocity



(a) Logs from the left part of the model



(b) Logs from the right part of the model

Figure 5.24: Well logs from the final inversion result, compared with logs from the start model (Model-C). Black represents the start model, Red represents the true model, while Blue represents the final updated model.

5.4 Experiment 2 : Reflection FWI with a high background velocity

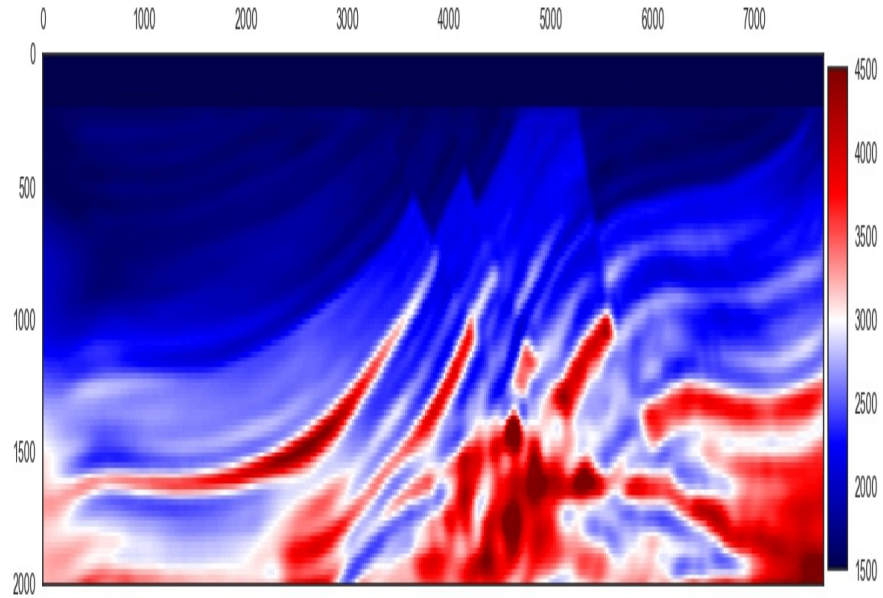
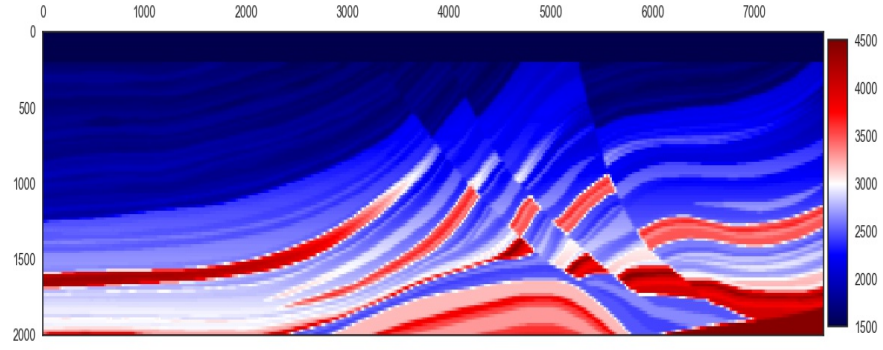
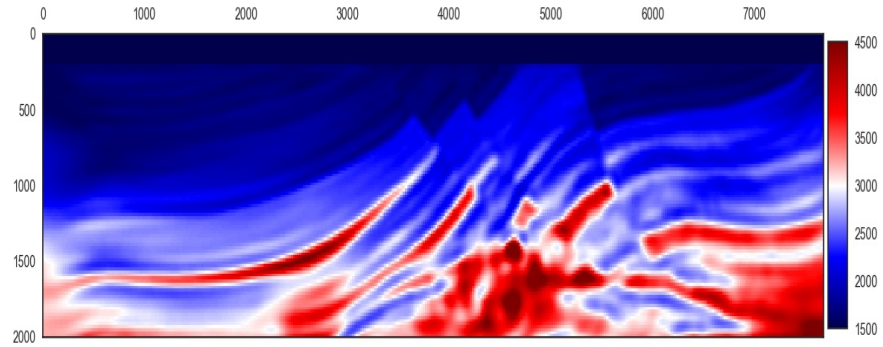


Figure 5.25: Final results obtained using finite difference explicit wavefield decomposition and conventional FWI

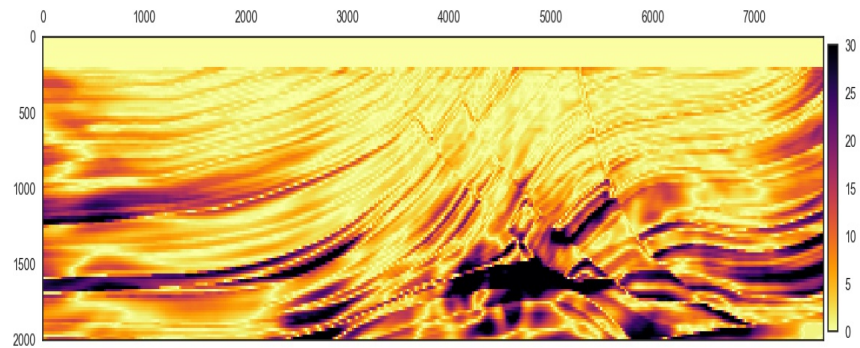
5.4 Experiment 2 : Reflection FWI with a high background velocity



(a) True Model



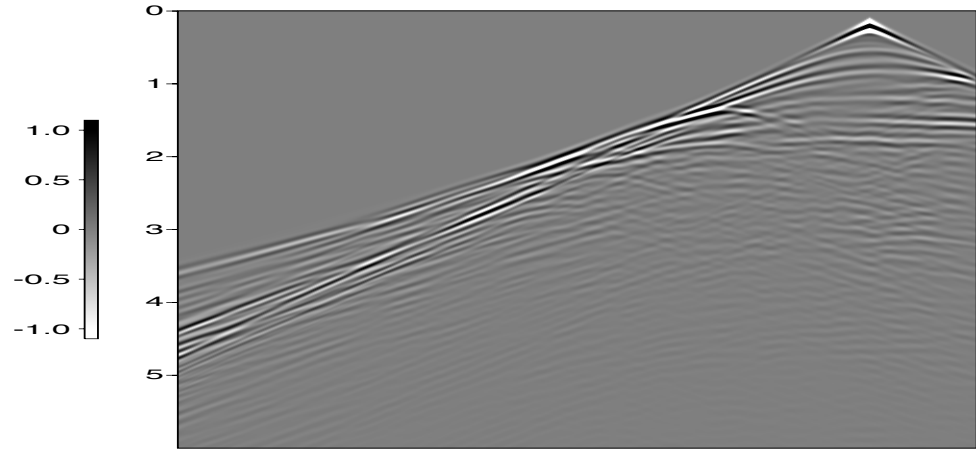
(b) Inverted Model



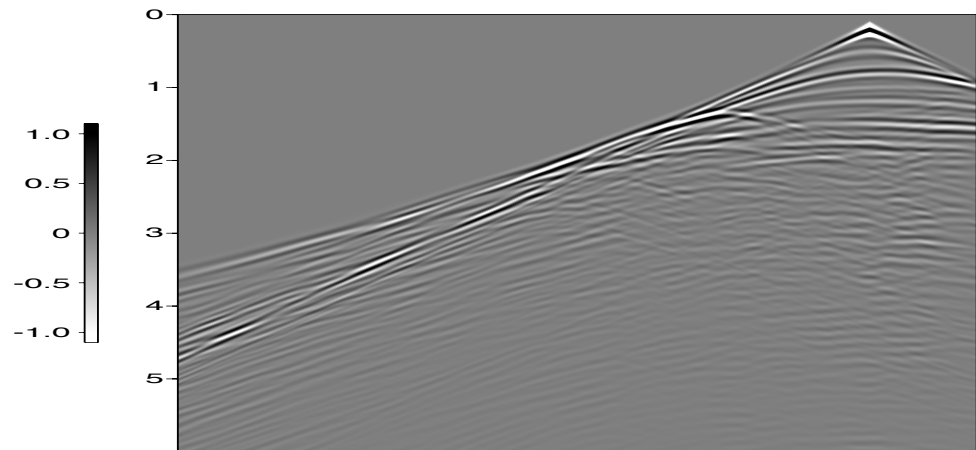
(c) Model Residuals as Percentage

Figure 5.26: *Inverted models and its residuals*

5.4 Experiment 2 : Reflection FWI with a high background velocity



(a) Shot from inverted model



(b) Shot from true model

Figure 5.27: Sample shot data from the right side of the model, computed using the inverted model

5.4 Experiment 2 : Reflection FWI with a high background velocity

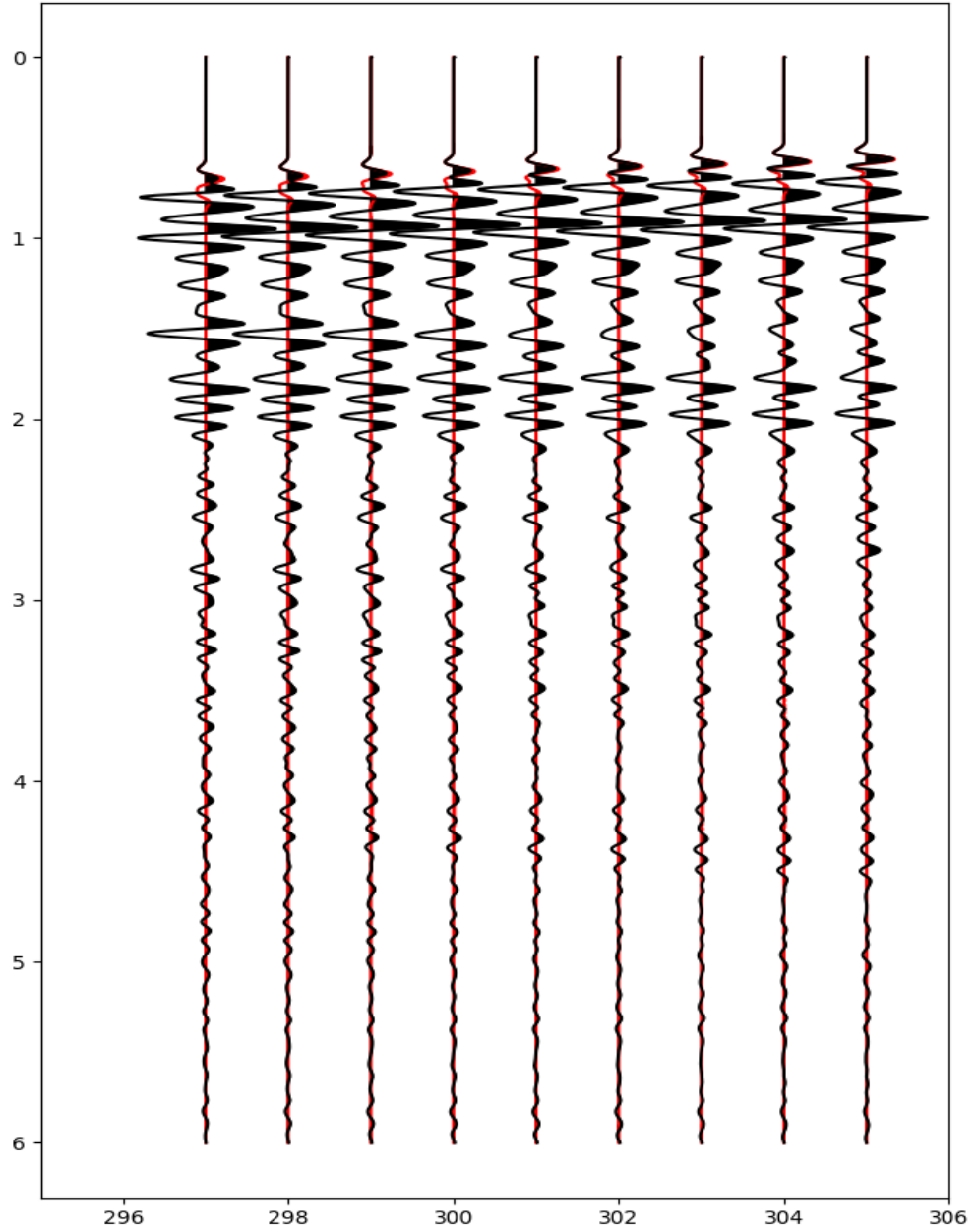


Figure 5.28: Comparison of near traces from the shot gather in Figure 5.27. The red traces represent those from the start model, while the black are from the true model.

5.4 Experiment 2 : Reflection FWI with a high background velocity

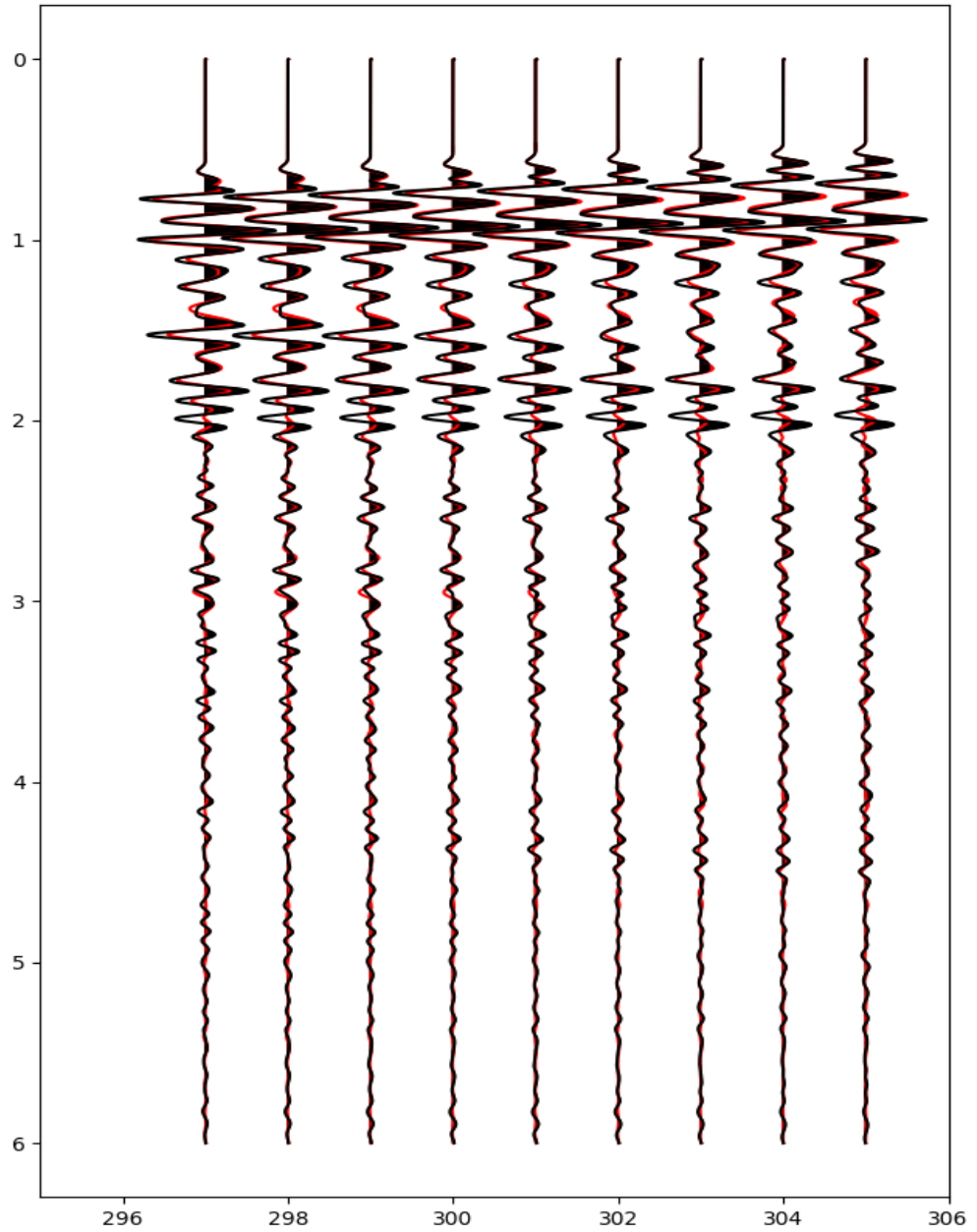


Figure 5.29: Comparison of near traces from the shot gather in Figure 5.27. The red traces represent those from the recovered model, while the black are from the true model.

5.4 Experiment 2 : Reflection FWI with a high background velocity

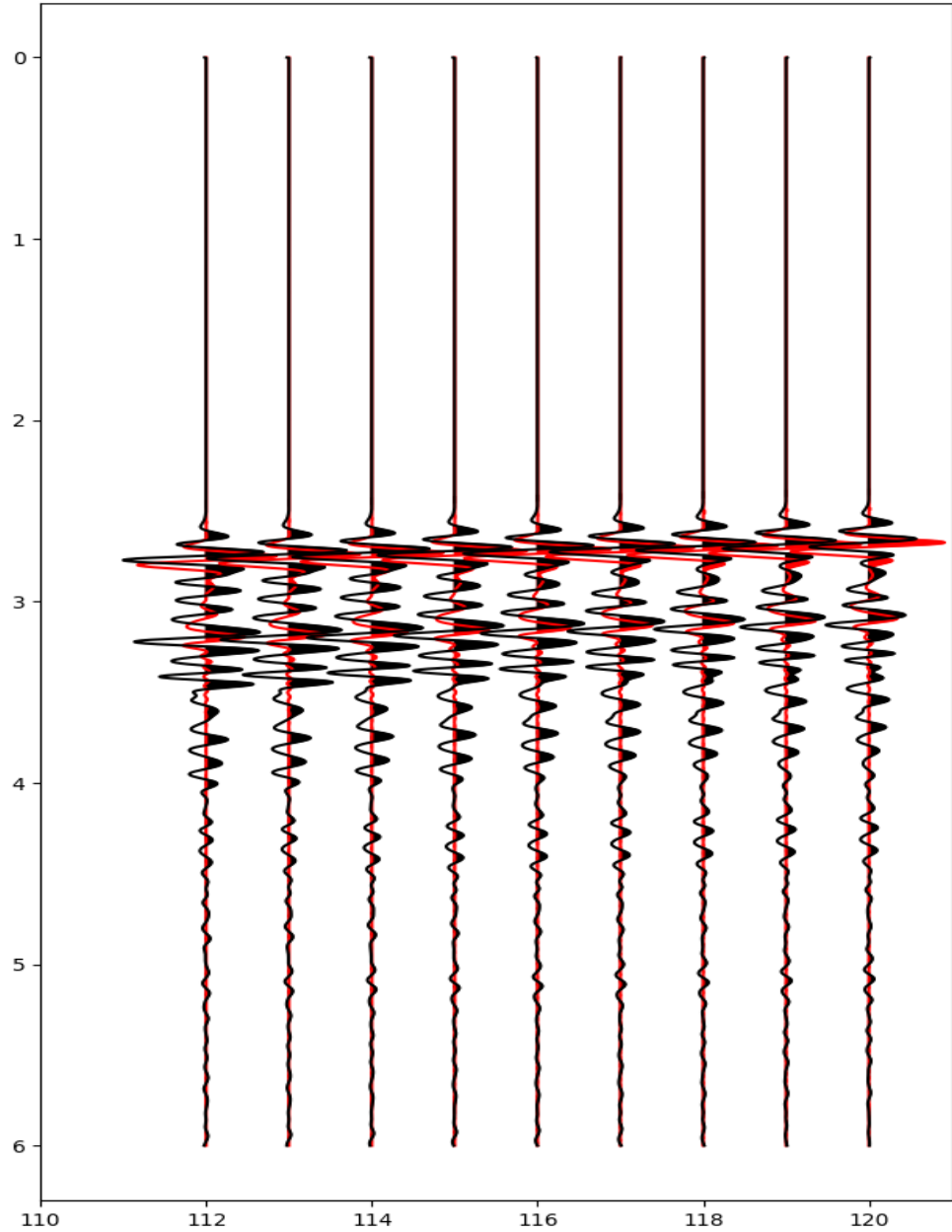


Figure 5.30: Comparison of mid traces from the shot gather in Figure 5.27. The red traces represent those from the start model, while the black are from the true model.

5.4 Experiment 2 : Reflection FWI with a high background velocity

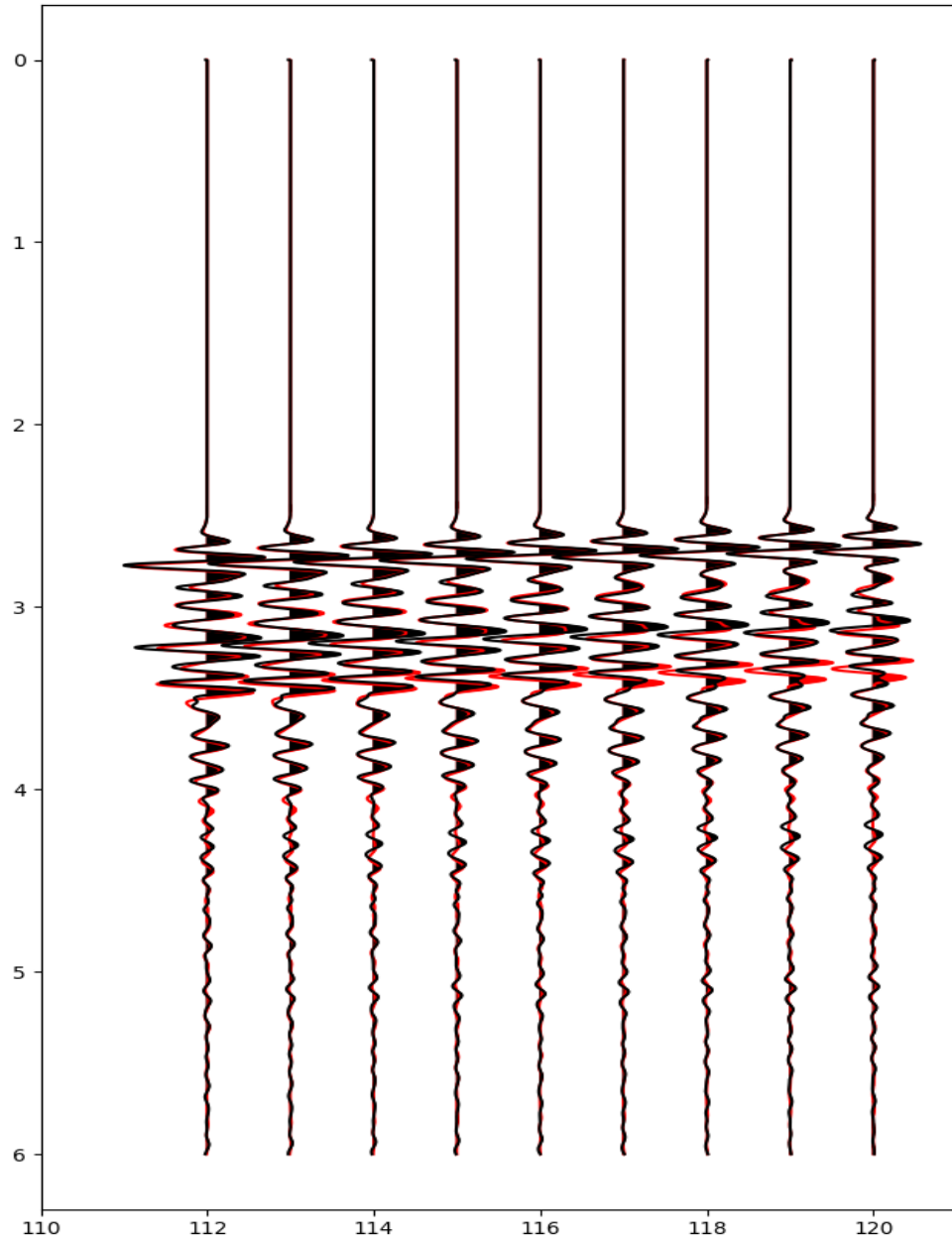


Figure 5.31: Comparison of mid traces from the shot gather in Figure 5.27. The red traces represent those from the recovered model, while the black are from the true model.

5.4 Experiment 2 : Reflection FWI with a high background velocity

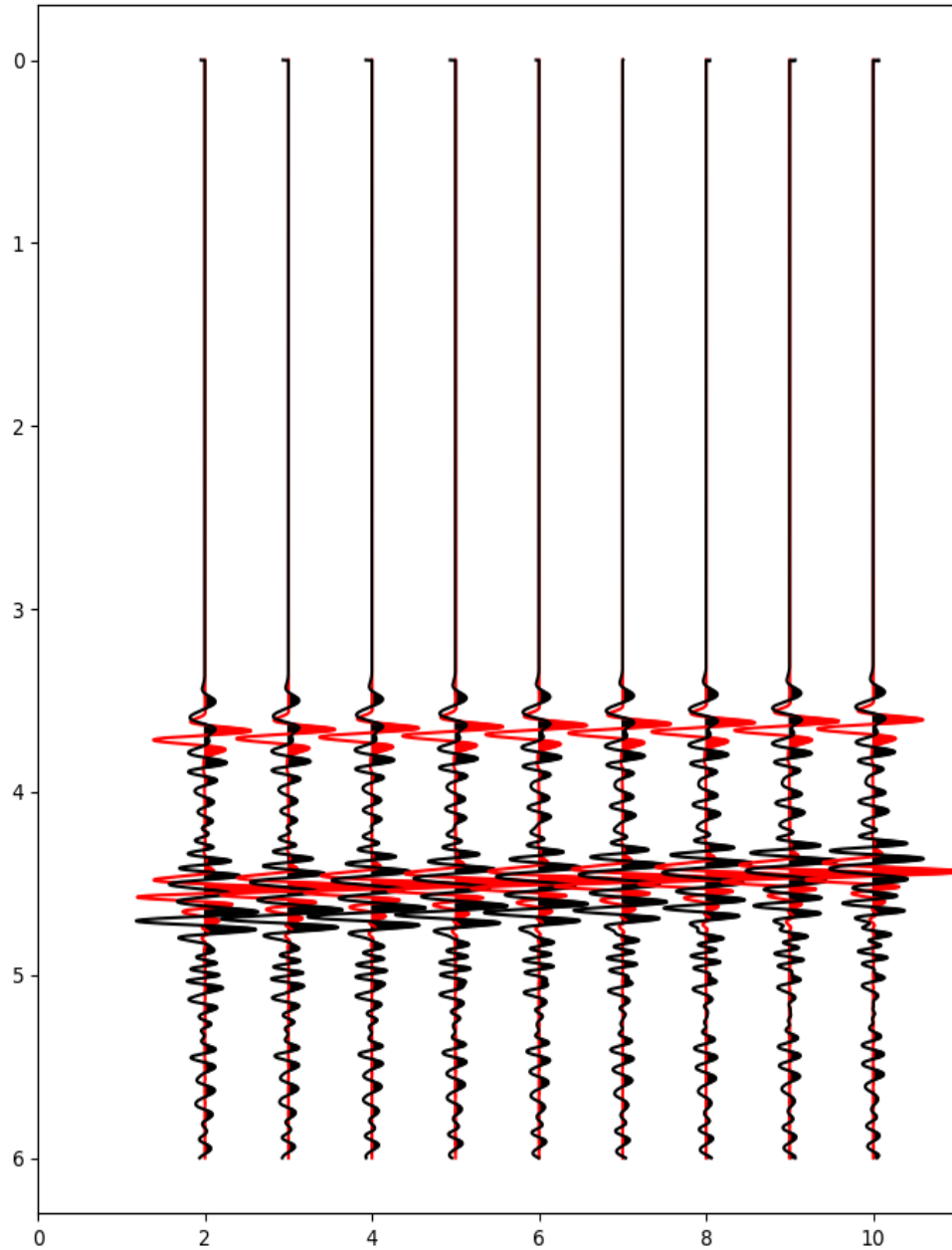


Figure 5.32: Comparison of far traces from the shot gather in Figure 5.27. The red traces represent those from the start model, while the black are from the true model.

5.4 Experiment 2 : Reflection FWI with a high background velocity

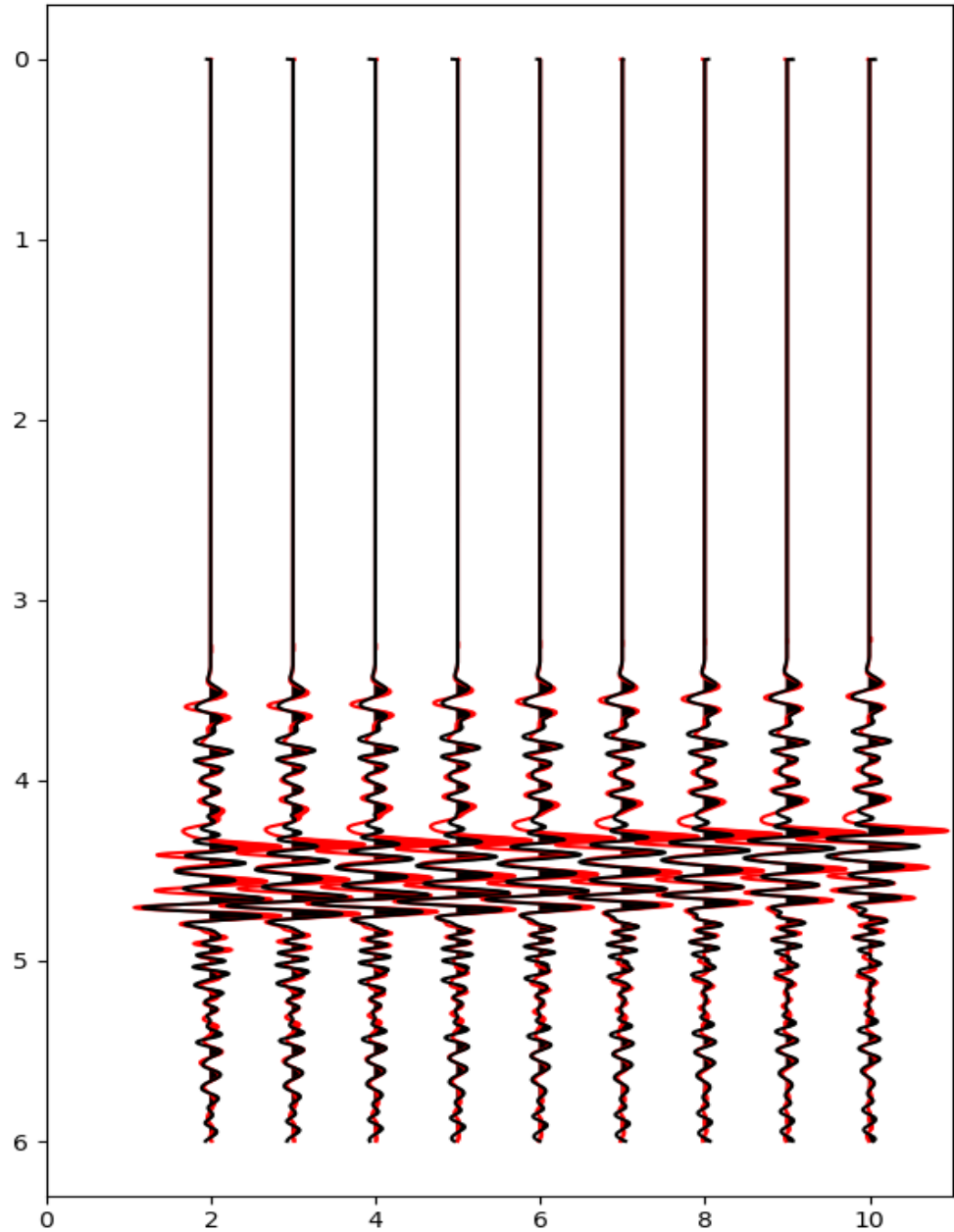
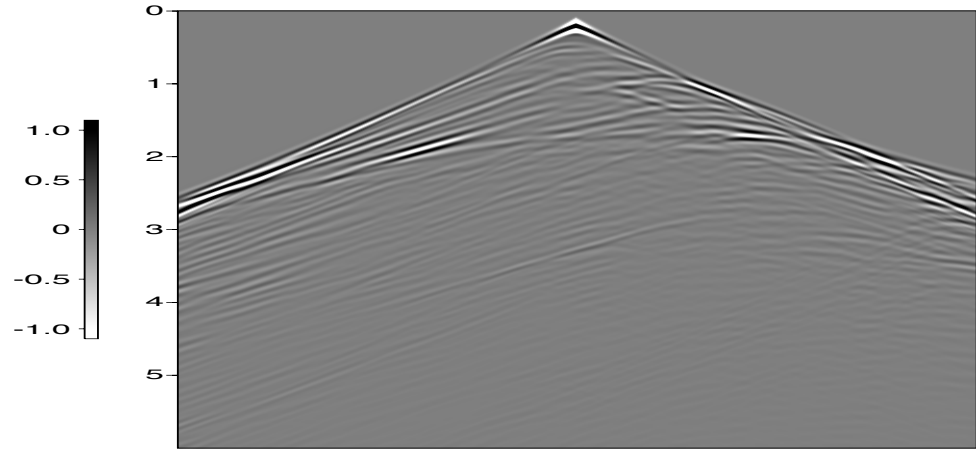
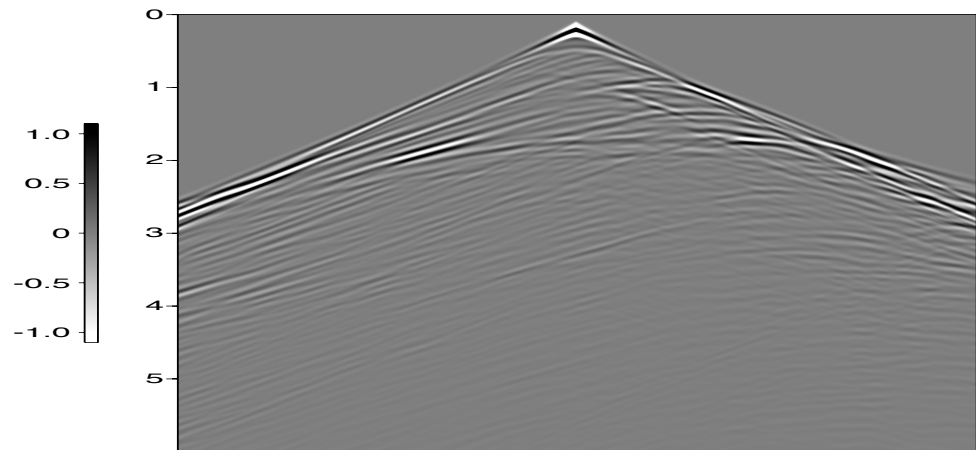


Figure 5.33: Comparison of far traces from the shot gather in Figure 5.27. The red traces represent those from the recovered model, while the black are from the true model.

5.4 Experiment 2 : Reflection FWI with a high background velocity



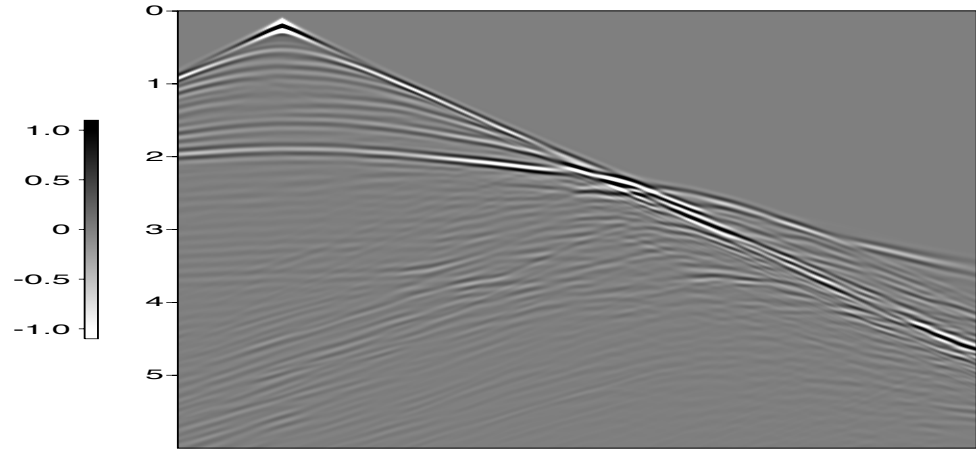
(a) Shot from inverted model



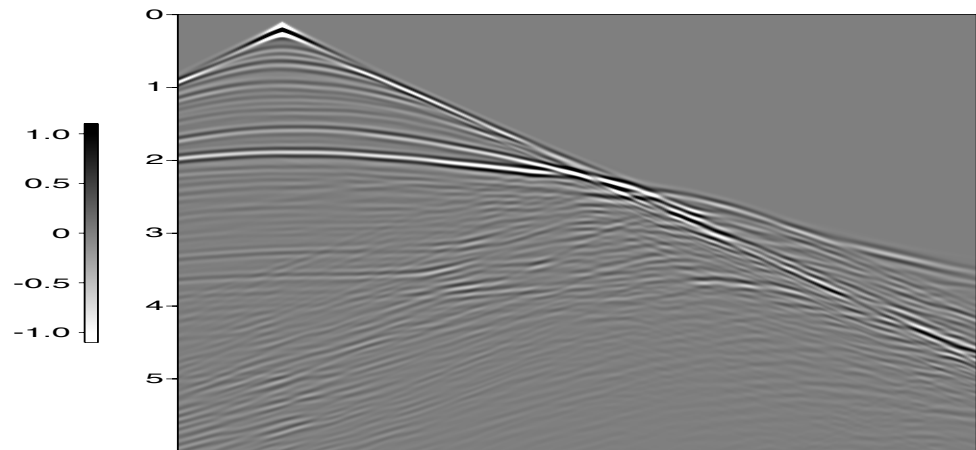
(b) Shot from true model

Figure 5.34: Sample shot data from the center of the model, computed using the inverted model

5.4 Experiment 2 : Reflection FWI with a high background velocity



(a) Shot from inverted model



(b) Shot from true model

Figure 5.35: Sample shot data from the left side of the model, computed using the inverted model

5.4 Experiment 2 : Reflection FWI with a high background velocity

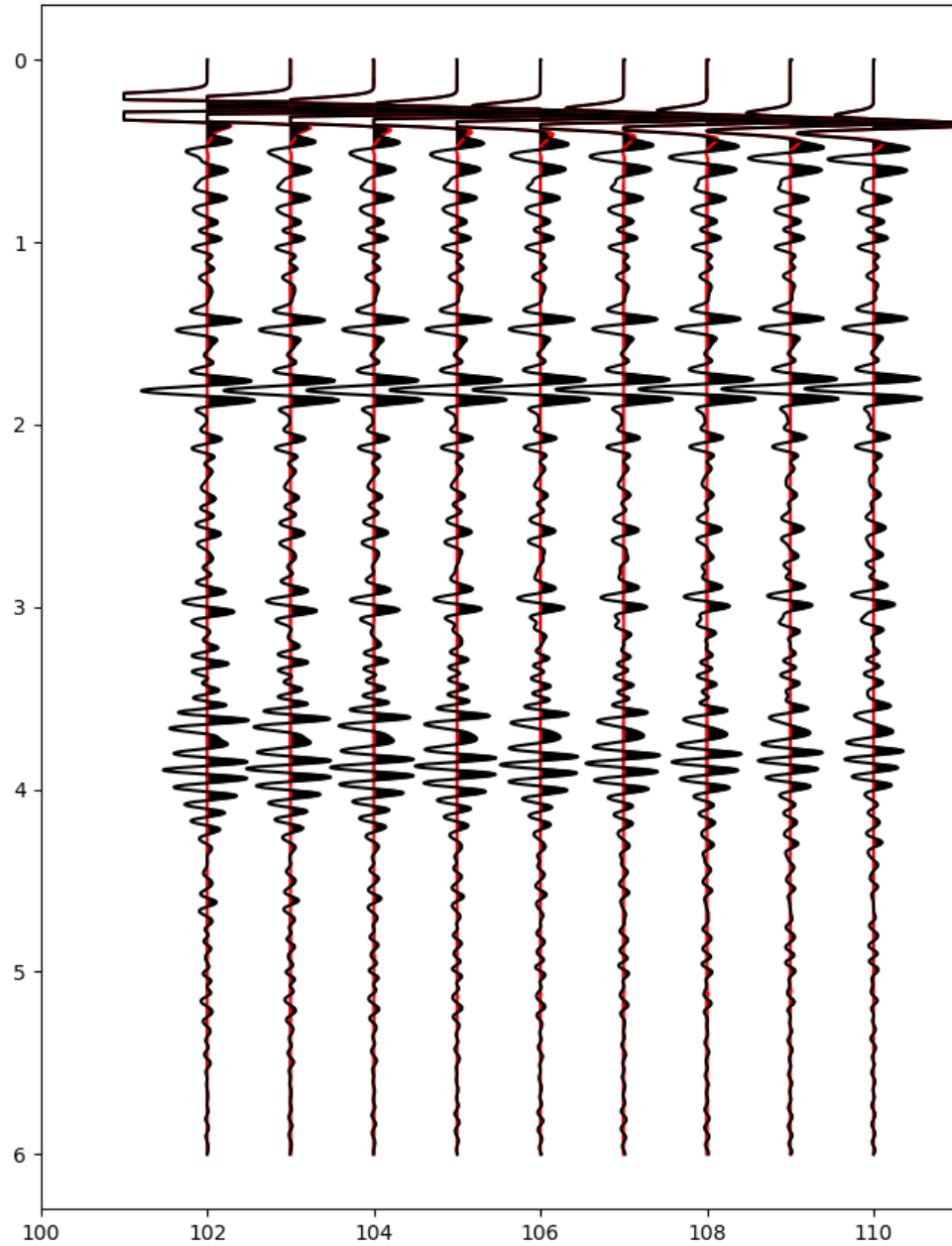


Figure 5.36: Comparison of near traces from the shot gather in Figure 5.35. The red traces represent those from the start model, while the black are from the true model.

5.4 Experiment 2 : Reflection FWI with a high background velocity

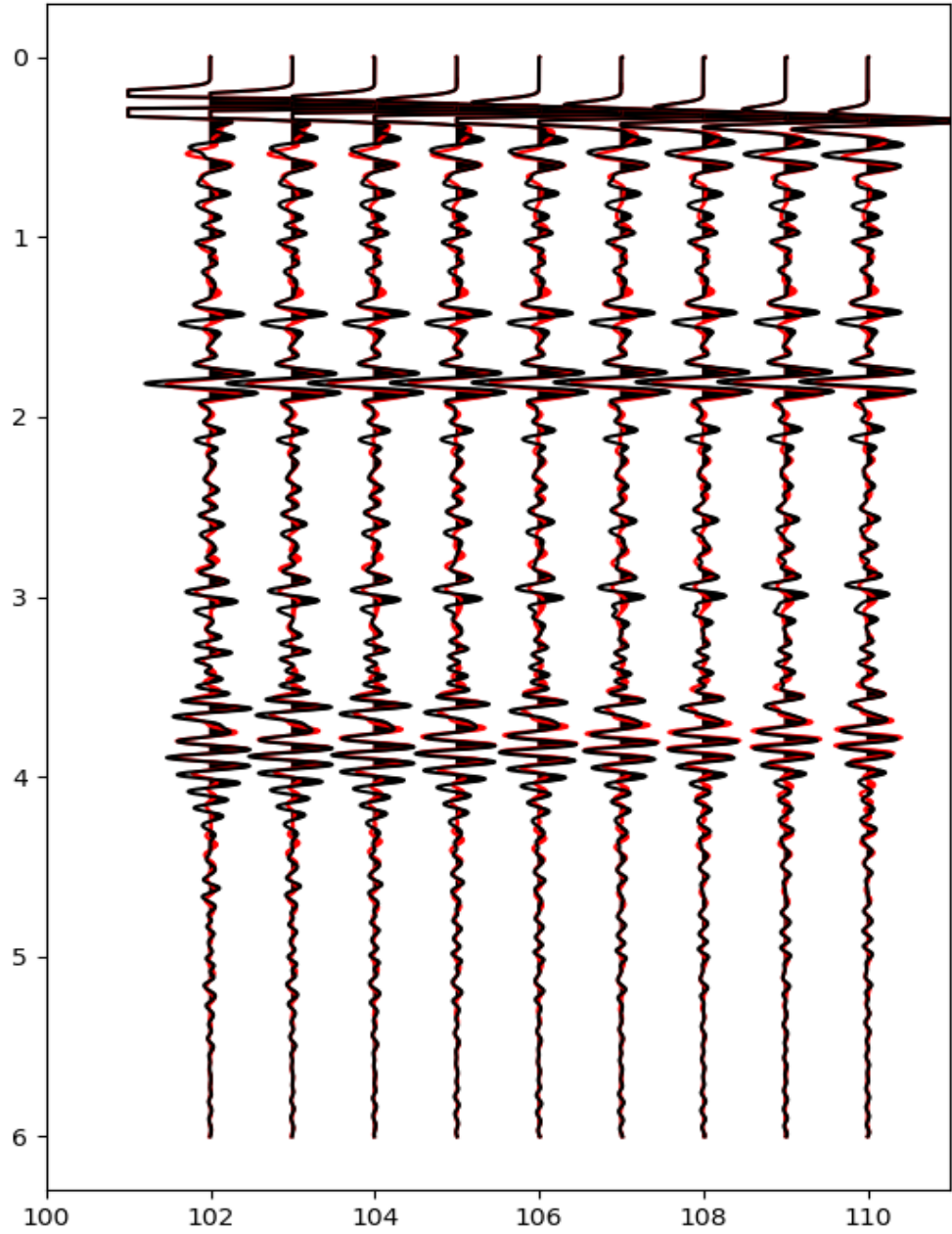


Figure 5.37: Comparison of near traces from the shot gather in Figure 5.35. The red traces represent those from the recovered model, while the black are from the true model.

5.4 Experiment 2 : Reflection FWI with a high background velocity

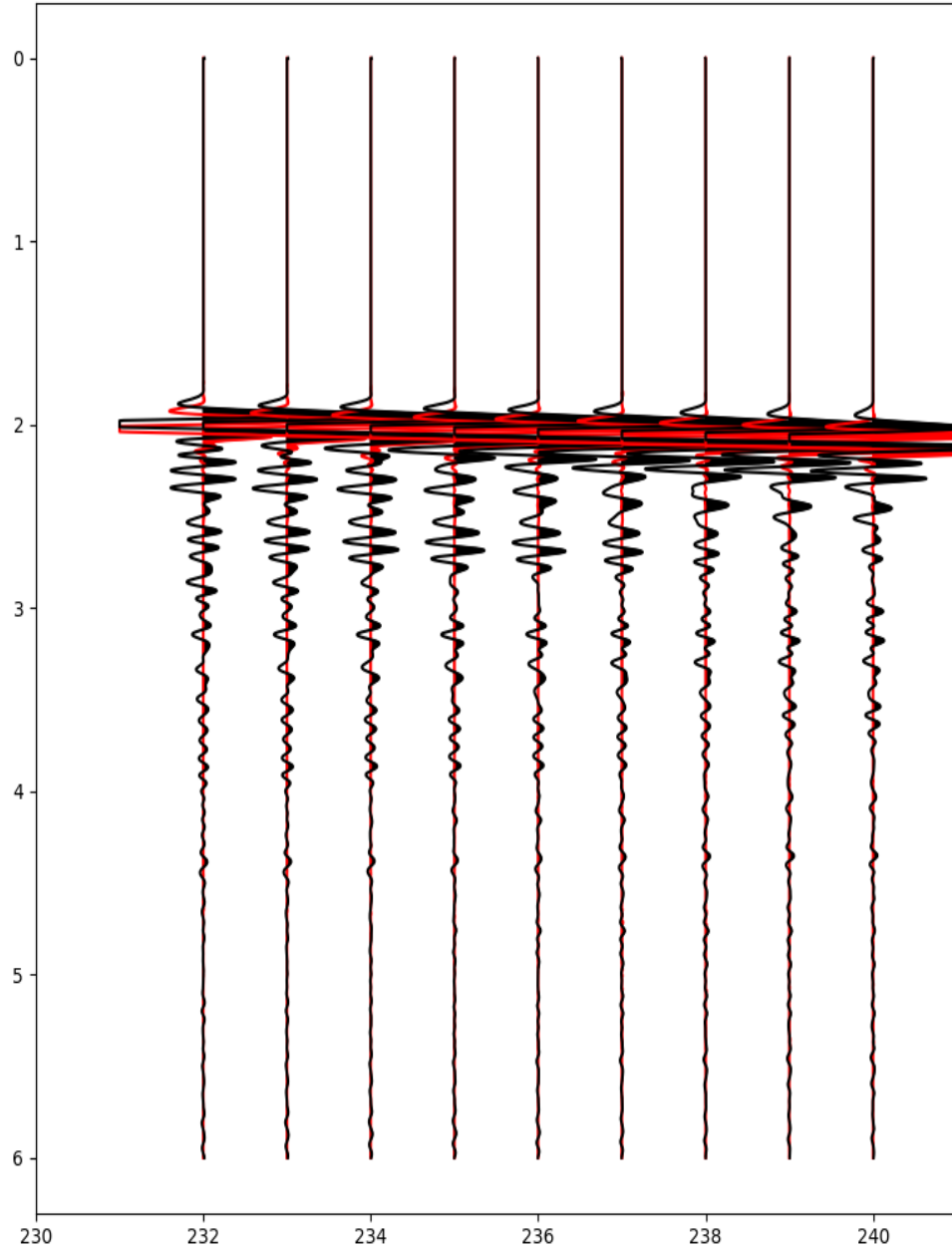


Figure 5.38: Comparison of mid traces from the shot gather in Figure 5.35. The red traces represent those from the start model, while the black are from the true model.

5.4 Experiment 2 : Reflection FWI with a high background velocity

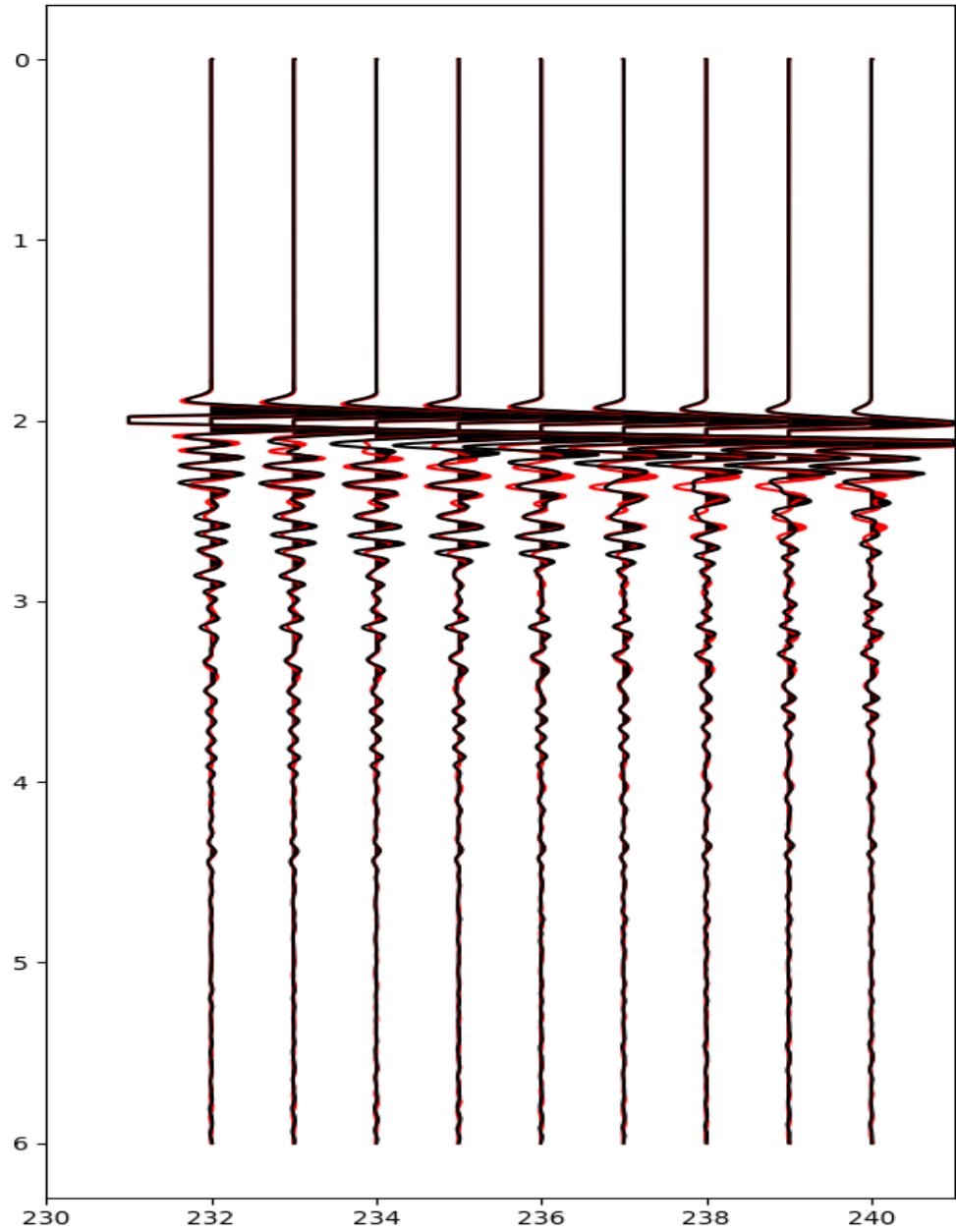


Figure 5.39: Comparison of mid traces from the shot gather in Figure 5.35. The red traces represent those from the recovered model, while the black are from the true model.

5.4 Experiment 2 : Reflection FWI with a high background velocity

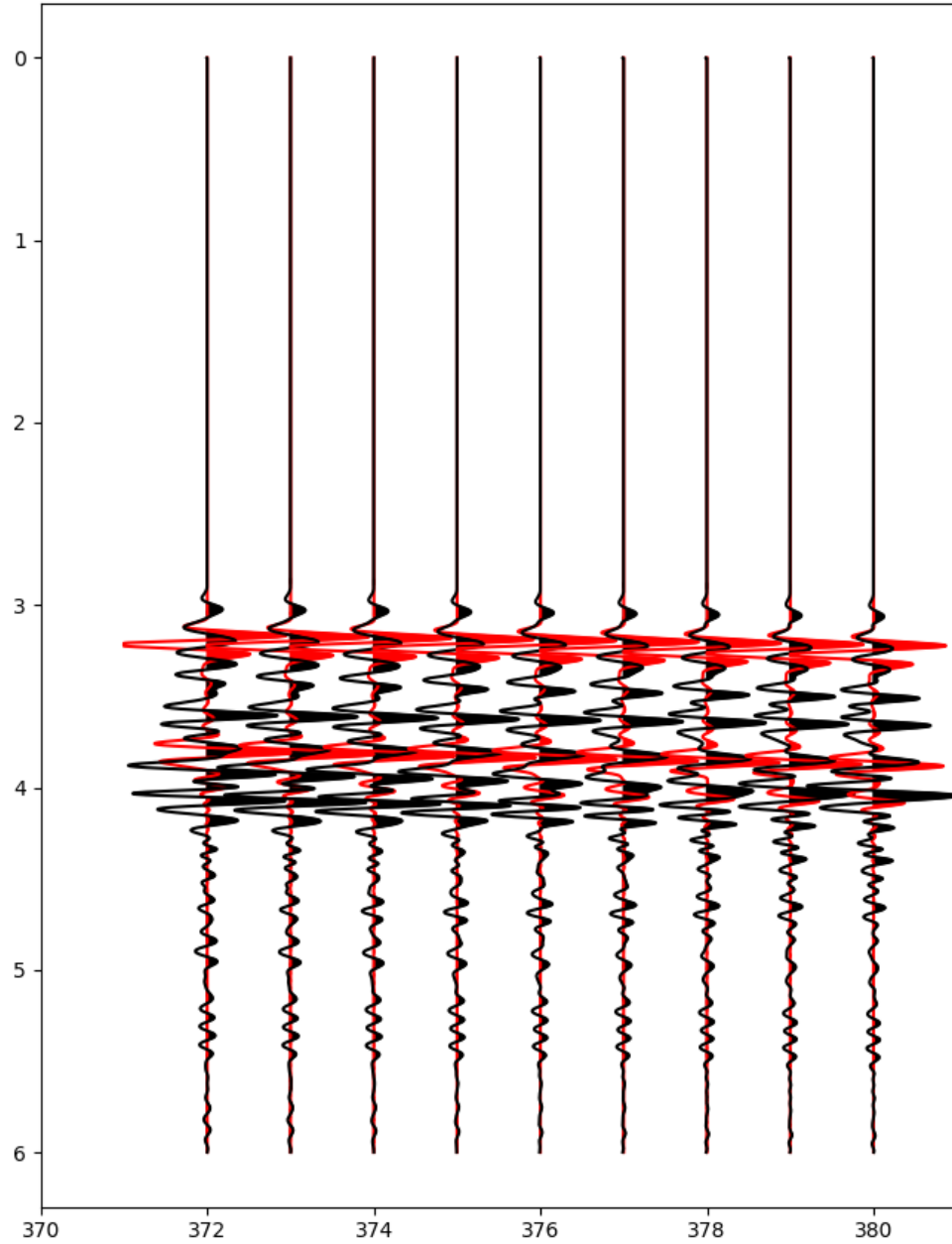


Figure 5.40: Comparison of far traces from the shot gather in Figure 5.35. The red traces represent those from the start model, while the black are from the true model.

5.4 Experiment 2 : Reflection FWI with a high background velocity

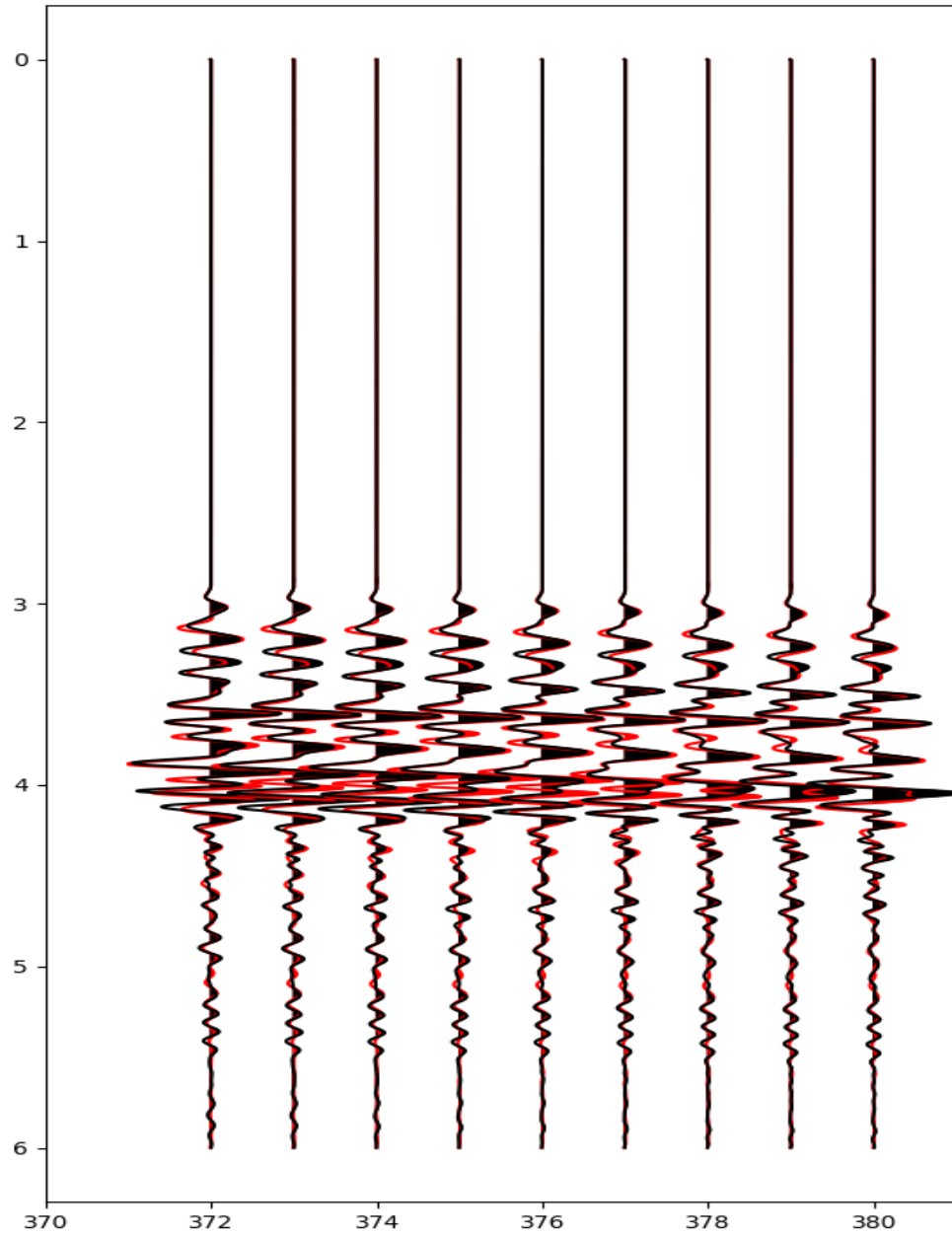


Figure 5.41: Comparison of far traces from the shot gather in Figure 5.35. The red traces represent those from the recovered model, while the black are from the true model.

Chapter 6

Blind Data Example

This chapter discusses an application of the inversion scheme developed in this thesis to a blind synthetic dataset. The chapter begins with a description of the raw dataset, the initial model and a sample well data provided. It continues with a review of the basic pre-processing steps applied to the data. The inversion scheme applied is then presented. A discuss of the inversion results and their evolution through the iteration steps then follows. Finally, the chapter concludes with a summary of the work done with the synthetic blind data.

6.1 Problem Set-up

The problem is the inversion of a 6 km deep velocity model using a dataset with a maximum offset of 8km. The problem is further compounded by the presence of a velocity inversion layer in the model. The dataset is courtesy Chevron Corporation (hereafter Chevron). The dataset distributed in the summer of 2014, was called "2D MARINE ISOTROPIC ELASTIC SYNTHETIC SEISMIC BENCHMARK". Chevron is yet to release the true model, so the dataset as at the time of writing this report could be considered "real". There has been some publications based on this dataset, and we hope to match the results here with some of those results.

6.1 Problem Set-up

A source wavelet and a 1-D velocity model (with the sea bed accurately place) are provided to start the inversion. Finally, Chevron provided a well log from a depth of 1000m to a depth of 2450 m (figure 6.1). This well log shall not be used to constrain the inversion in any way. However, it will serve as a quantitative verifier of the inversion results around its location.

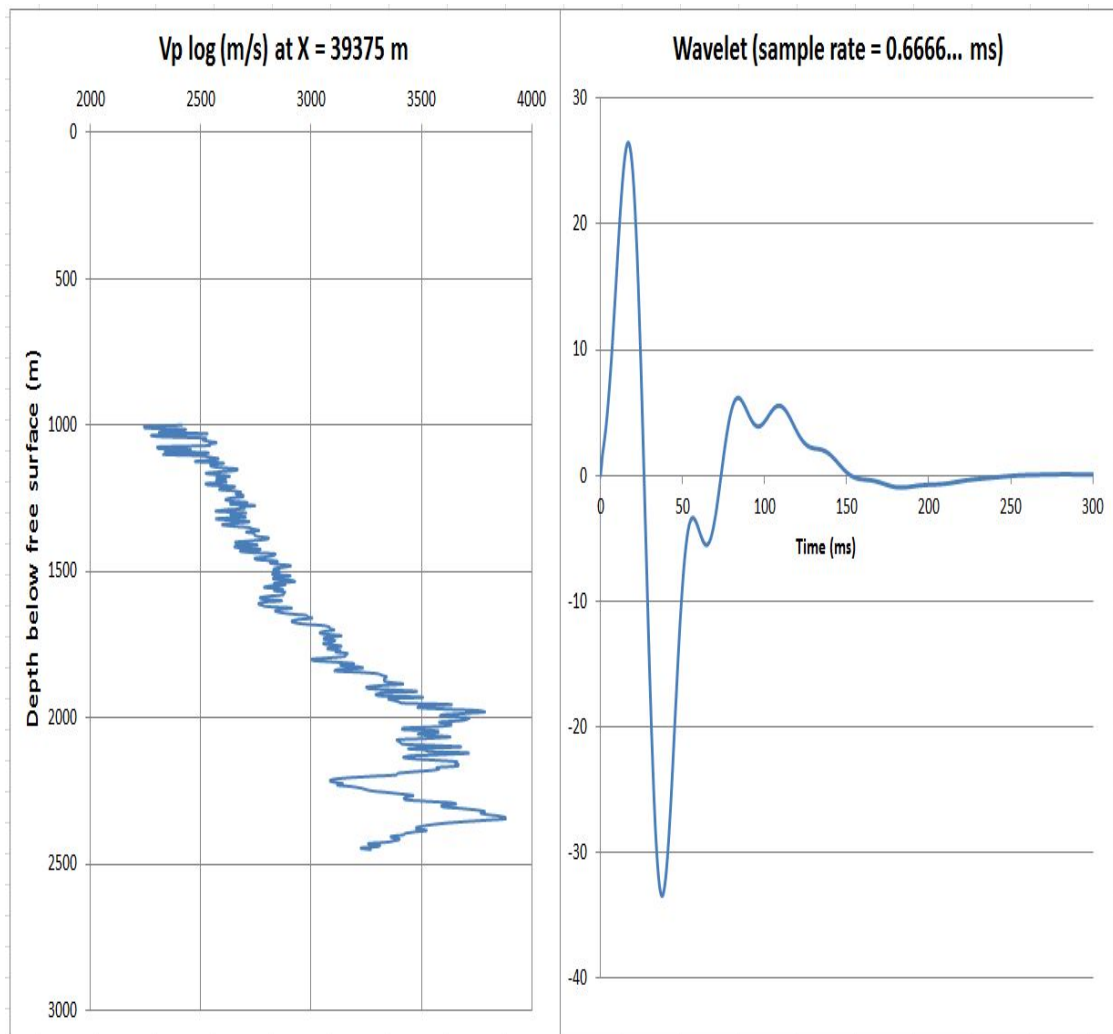


Figure 6.1: Well log and source wavelet provided by Chevron

6.1.1 Data

The dataset provided comprises one SEG-Y formatted file containing the observed seismic data. The observed data was modelled using an isotropic elastic forward modelling program, with the free surface multiples left in the data. The water velocity was fixed at 1510 m/s.

An exploratory analysis of the a sample shot gather (Figure 6.2) from the dataset reveals that the data is limited in offset and it is dominated by reflections and multiples. There is limited amount of turning waves. These can be used for building the very shallow parts of the model.

6.1.2 Start Model

Figure 6.3 shows the start velocity model provided, which measures 48km by 6km. The model is essentially a 1-D layer model with a velocity inversion layer at about 2.2km. The seabed geometry has been provided in the start model.

A plot of the values of the starting velocity at the well location shows that it is relative faster than the true velocity at the well location (figure 6.5).

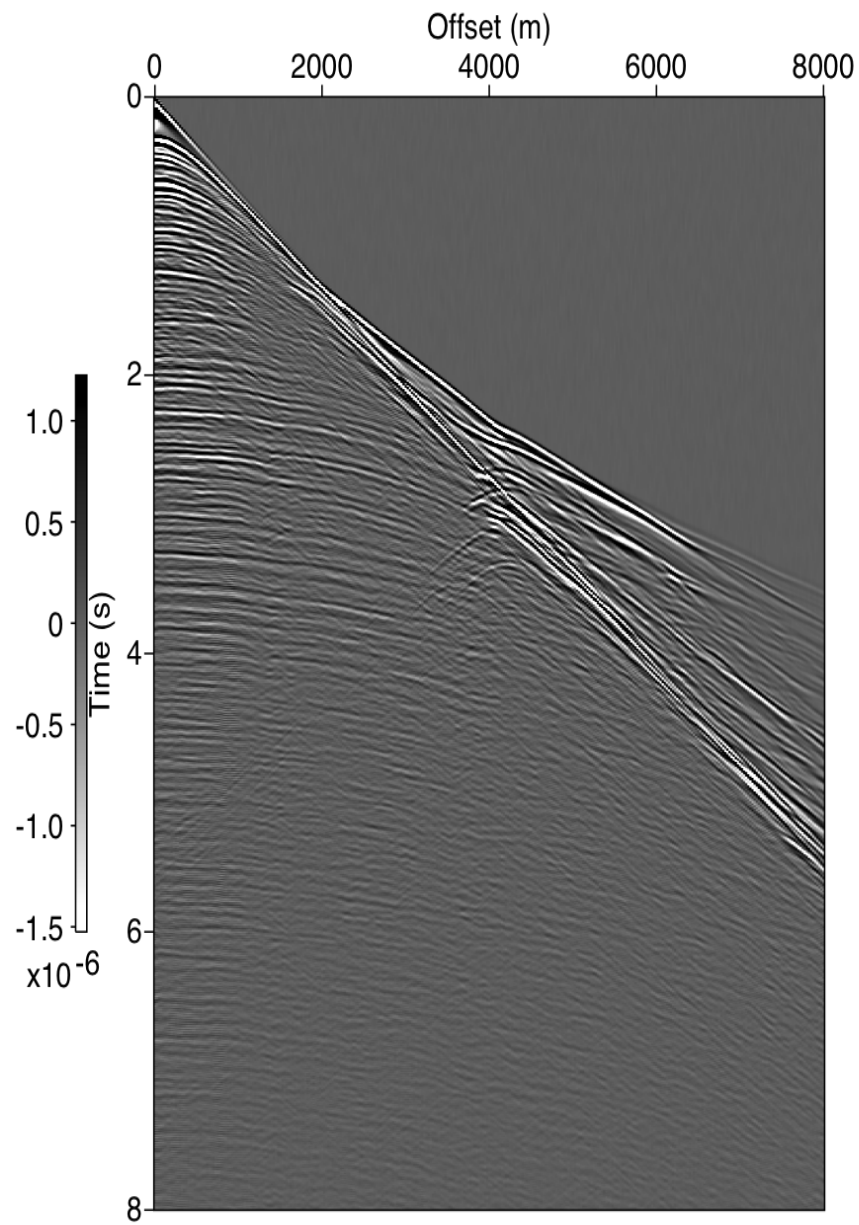


Figure 6.2: *Sample Shot Data (with full bandwidth)*

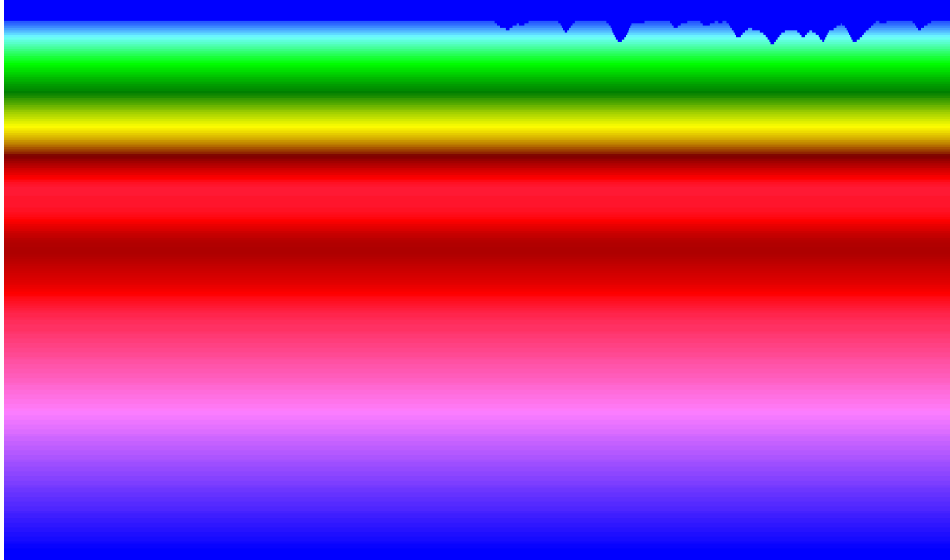


Figure 6.3: Initial model provided. Notice the well defined nature of the sea bed

6.1.3 Source Wavelet

The source wavelet given is a farfield wholespace (no ghost) source wavelet file. The original source wavelet and its frequency spectrum are shown in figure 6.6. The original source wavelet was sampled at 0.6 ms, and was therefore re-sampled to 2 ms.

6.1.4 Geometry

The following geometry described the acquisition and recording of this dataset.

In this experiment, we maintained the original geometry and did no form of reciprocity or some other adjustments to the original geometry

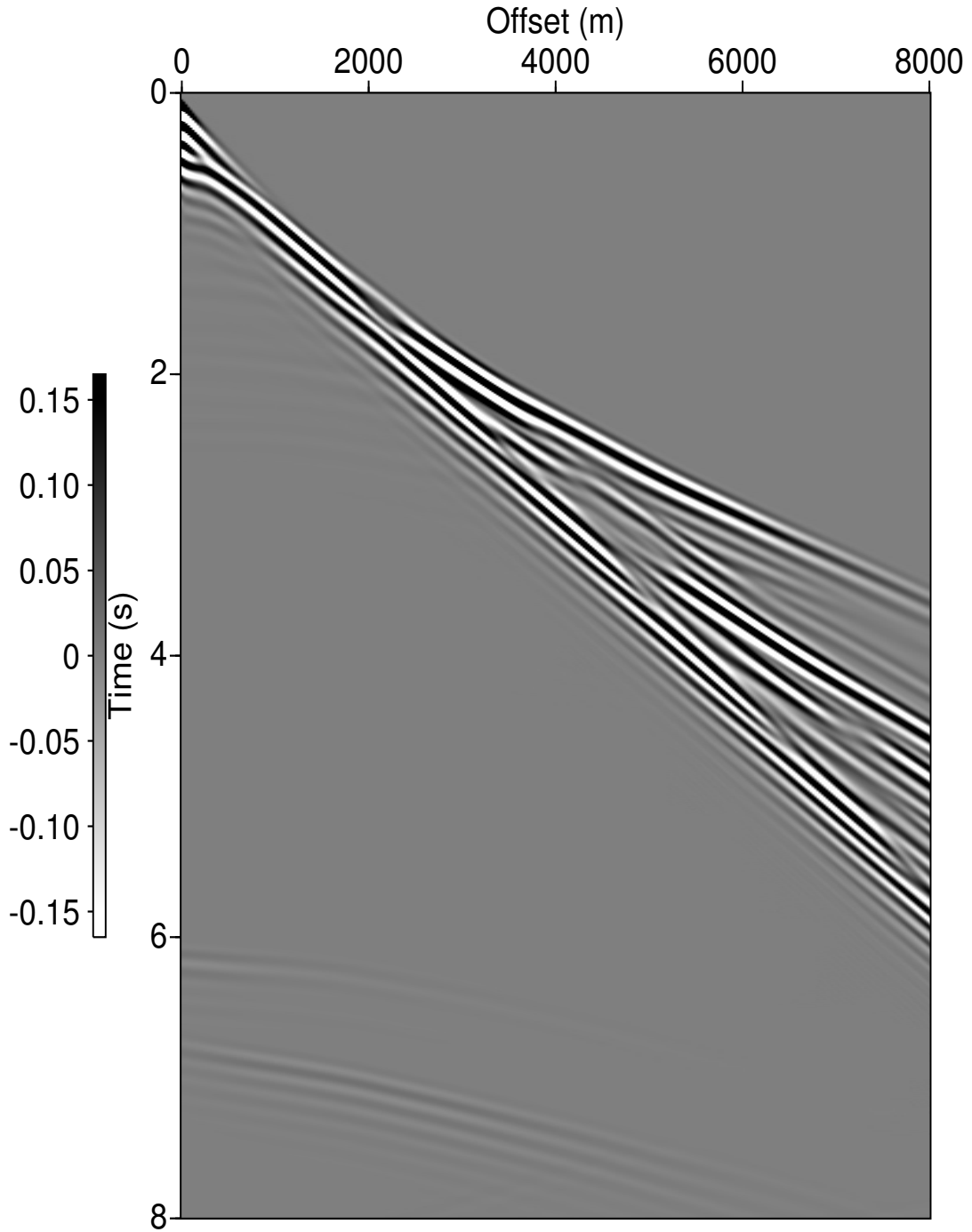


Figure 6.4: Sample Shot Data for a low pass filtered source propagated through the start model. There is a rich supply of transmission energies in this data, which can help in building the overburden via transmission FWI

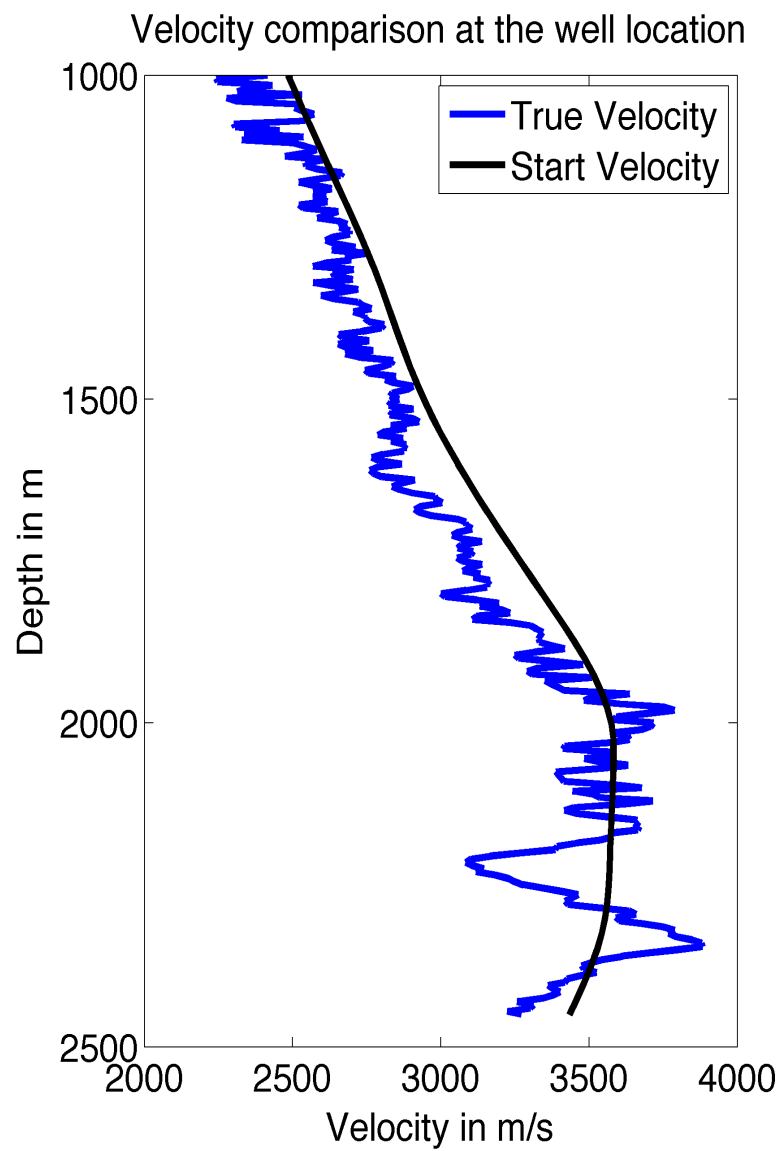


Figure 6.5: Well data comparison between the initial model and the true model

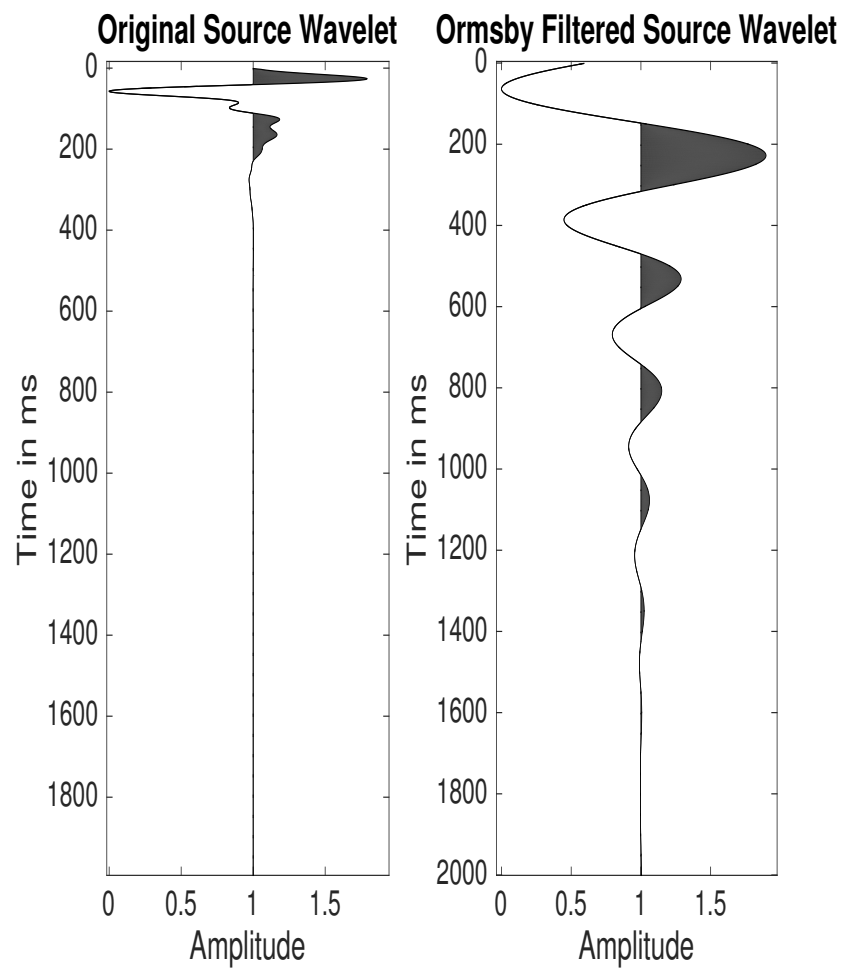


Figure 6.6: Original source wavelet and a low pass version (Ormsby : 2.5 - 4.5 - 6.6 - 9.8)

Acquisition Parameter	Value
Source Depth	15 m
Cable Depth	15 m
Shot Spacing	25 m
Receiver Spacing	25 m
Minimum Offset	0 m
Cable Length	8000 m
Number of Channels	321
Near Channel Number	1
Record length	8.0 s
Sample Interval	4ms

Table 6.1: *A summary of the acquisition parameters*

6.2 QC and Pre-processing

The very minimal pre-processing was applied to the data and source wavelet in form of a low pass filter, to ensure that only permissible frequencies as dictated by the computational mesh are allowed into the source wavelet and the data. Figure 6.6 shows the source wavelet and a low pass version of it. An Ormsby filter with edge frequencies 2.5, 4.5, 7.5, 9.5 Hz was used as the low pass filter.

It is pertinent to stress here that the wave propagation via FDEWD is highly sensitive to spikes around time $t=0$, compared to the conventional finite difference modeling. A workaround to this is to zero pad the wavelet around $t=0$, and gently damp the amplitudes from $t=0$ towards the negative time axis. This ensures that the amplitude of the wavelet starts from zero, and gradually grows to its value at $t=0$. Storage of the energies can then begin at $t=0$.

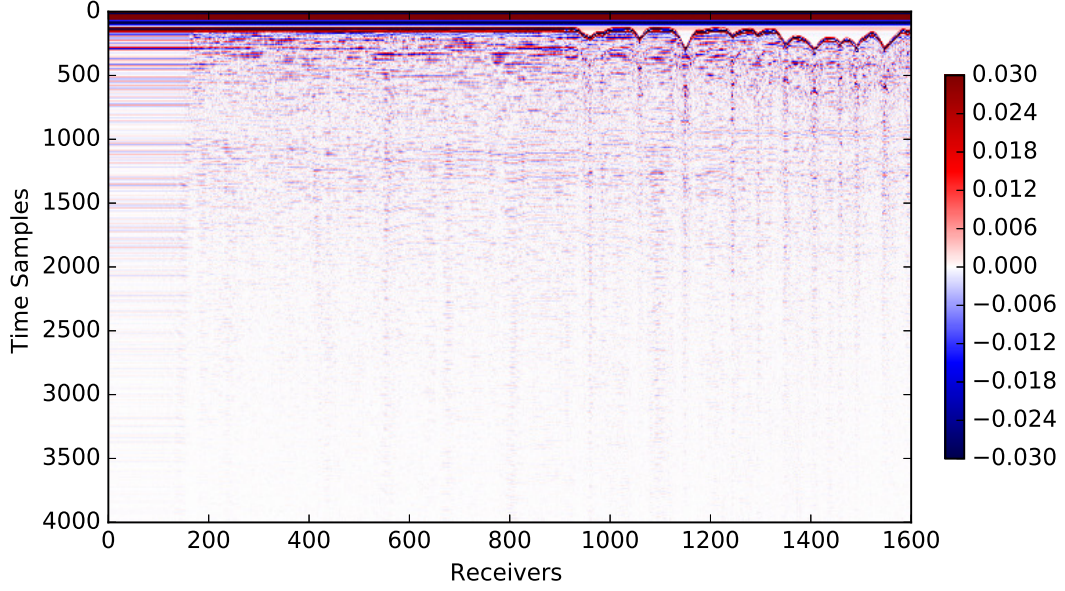


Figure 6.7: Near Trace Plot

6.3 Inversion Strategies

The following strategy was employed in inverting this dataset. (1) The available turning waves are used in building the shallow parts of the model. (2) Reflection FWI (via finite difference wavefield decomposition) is then used to build the macro model for the deeper parts of the model. (3) Conventional FWI then uses the output from process (2) and complete the FWI exercise.

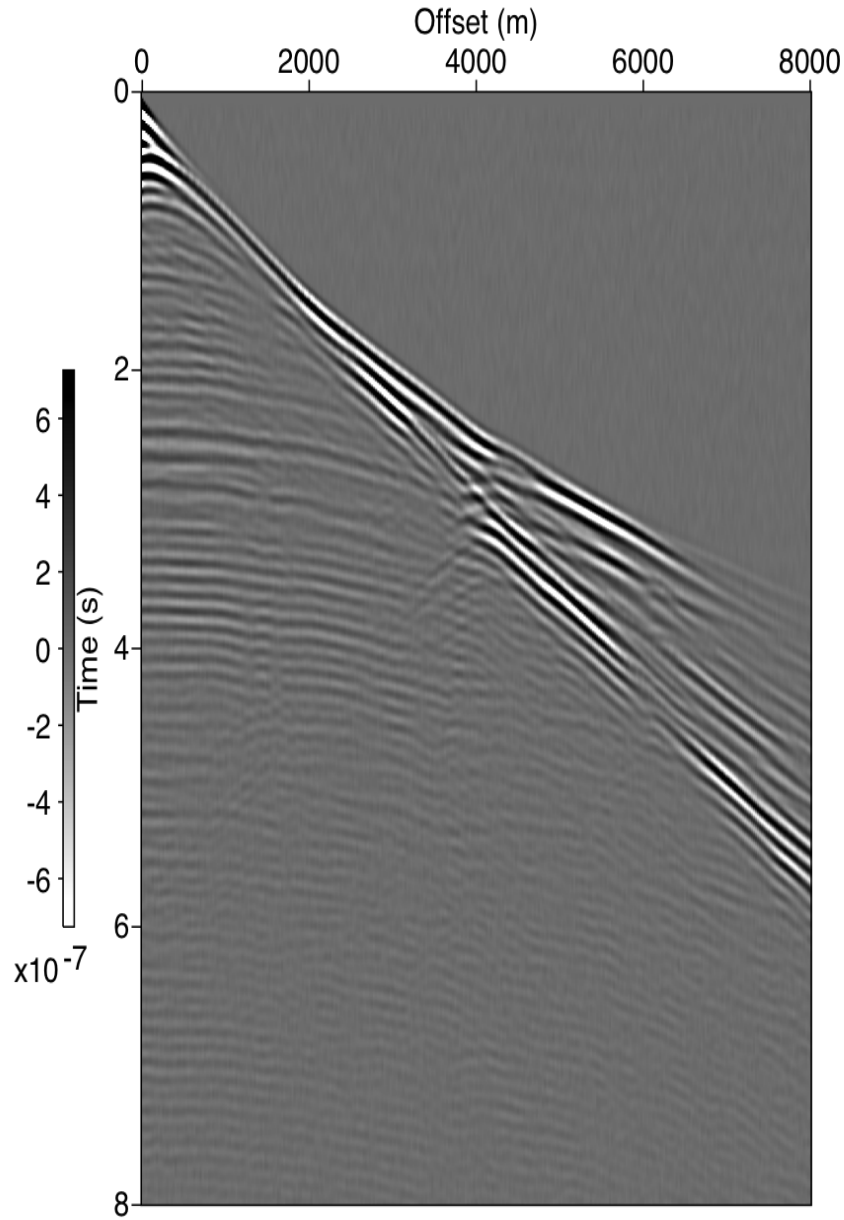


Figure 6.8: Observed Data filtered to acceptable propagation frequency using a 25 m mesh

6.4 Results

6.4.1 Frequency continuation FWI

Conventional FWI using frequency continuation (without data muting) was applied on the data. The results for this variant of FWI are shown in figures 6.9. These results give an idea of where have come from in terms of inverting datasets dominated with reflection energies, as they show how bad things can get.

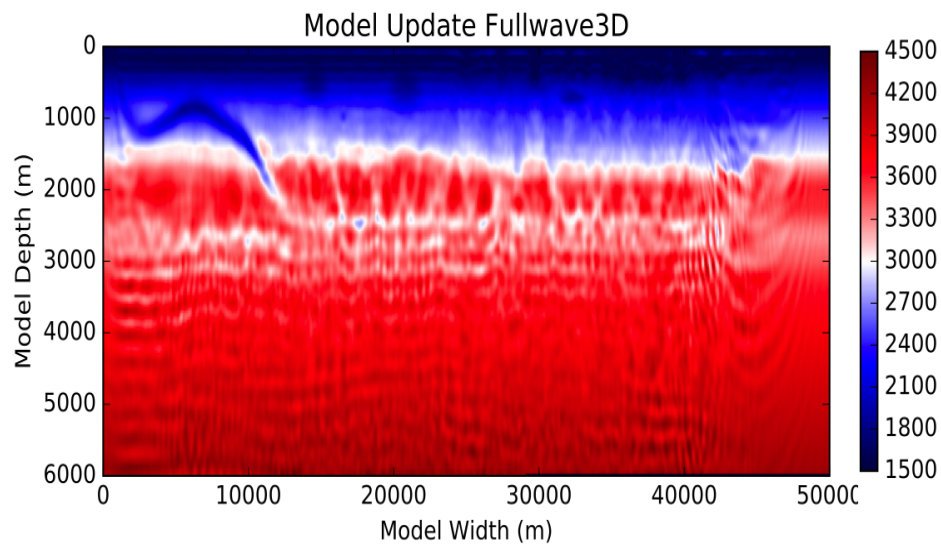


Figure 6.9: *Inverted results from conventional FWI*

6.4.2 Transmission FWI without muting

Traditional methods of doing transmission FWI involve data muting. However, the scheme developed in this thesis allows for transmission FWI on the full datasets, using only the horizontal components of the wavefields. This technique was used to build the shallow parts of the model. The results are shown in Figures 6.10 and 6.11. These results show the performance of the FDEWD in yielding correct gradients for the shallow parts of the model by optimally utilising the horizontal components of the wavefields.

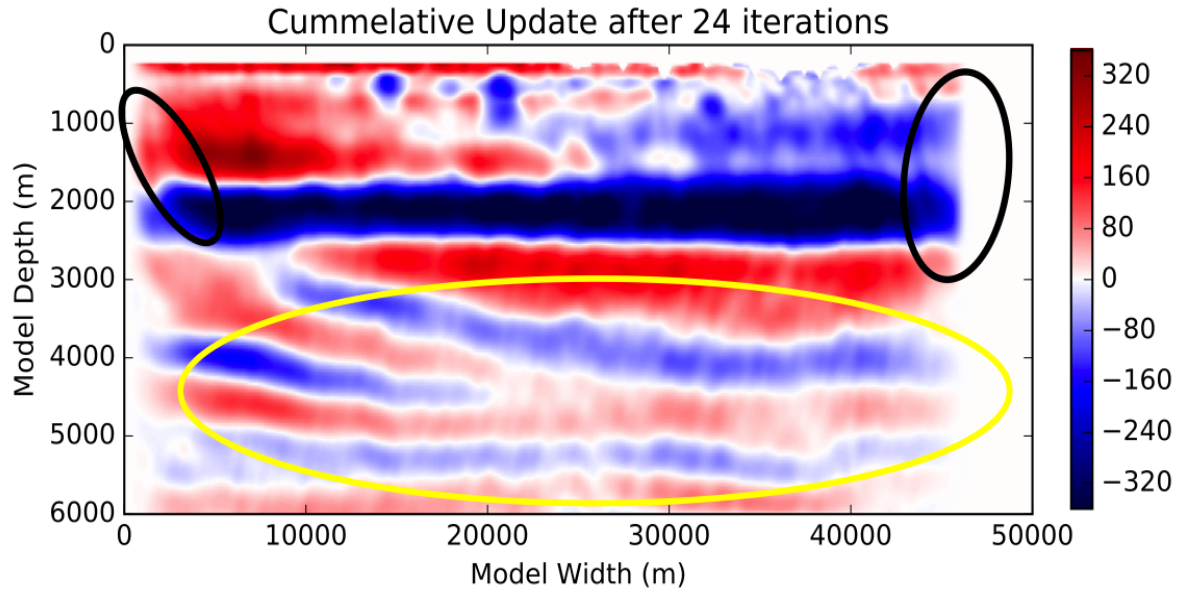


Figure 6.10: Gradients from transmission mode FWI via the use of the horizontal component of the wavefields in FDEWD. Here, horizontal smoothing is applied to the gradient.

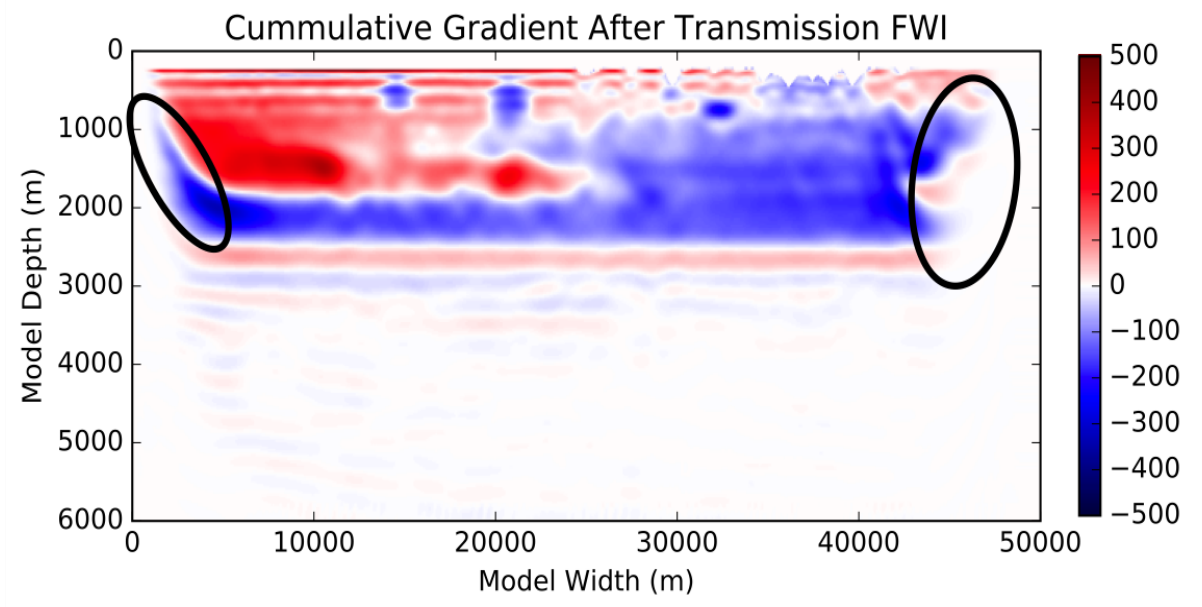


Figure 6.11: Enhanced gradient from transmission mode FWI via the use of the horizontal component of the wavefields in FDEWD. Here, the horizontal smoothing is varied and applied to the gradient. A vertical adaptive weighting factor was applied to enhance shallow parts of the gradient

6.4.3 Reflection FWI via finite difference wavefield decomposition

Reflection FWI benefits from a good starting model in the overburden. The final model of the transmission FWI on this dataset is therefore used as a starting model for the reflection FWI discussed here.

The strategy employed is as described in the previous section. The first migration update is shown in Figure 6.12. It has the necessary high wavenumbers that would act as scatterers in the subsequent build up of the macro model.

The final model recovered via the reflection mode FWI is shown in Figure 6.15. This model is then fed into a conventional FWI engine, and the overall recovered model is shown in Figure 6.16. A discussion of this result and how it compares with results from other variants of FWI is presented in the next chapter.

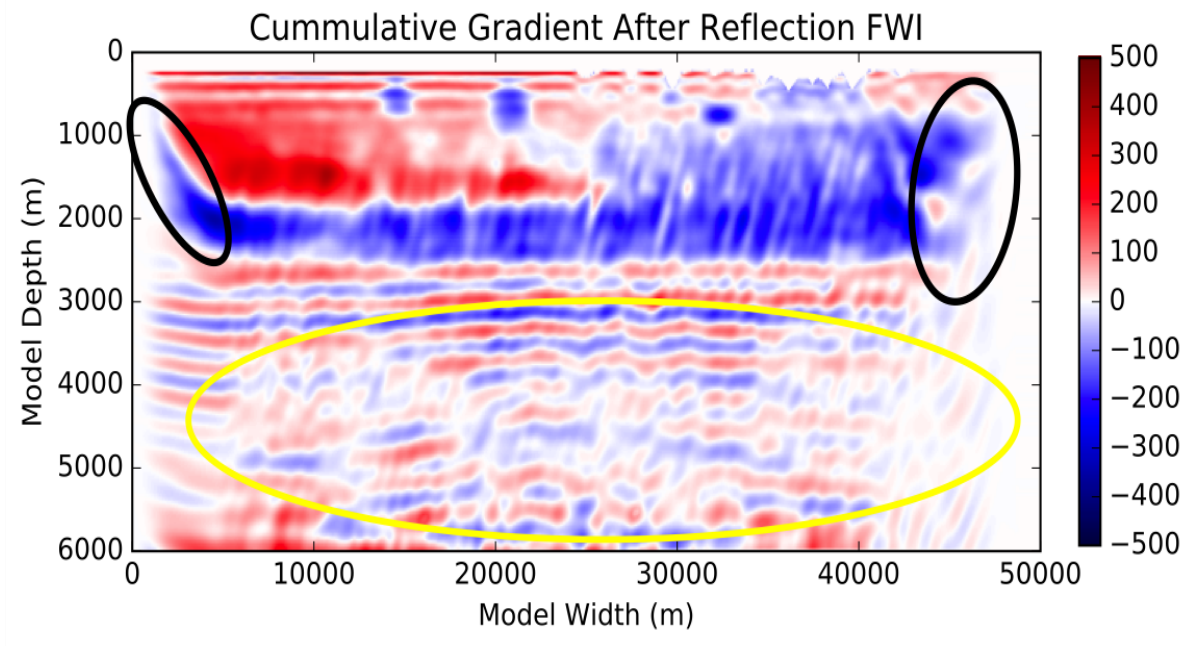


Figure 6.12: Initial gradient model from reflection mode FWI via the use of the vertical component of the wavefields in FDEWD. The start model for this mode was the model derived from the transmission mode FWI

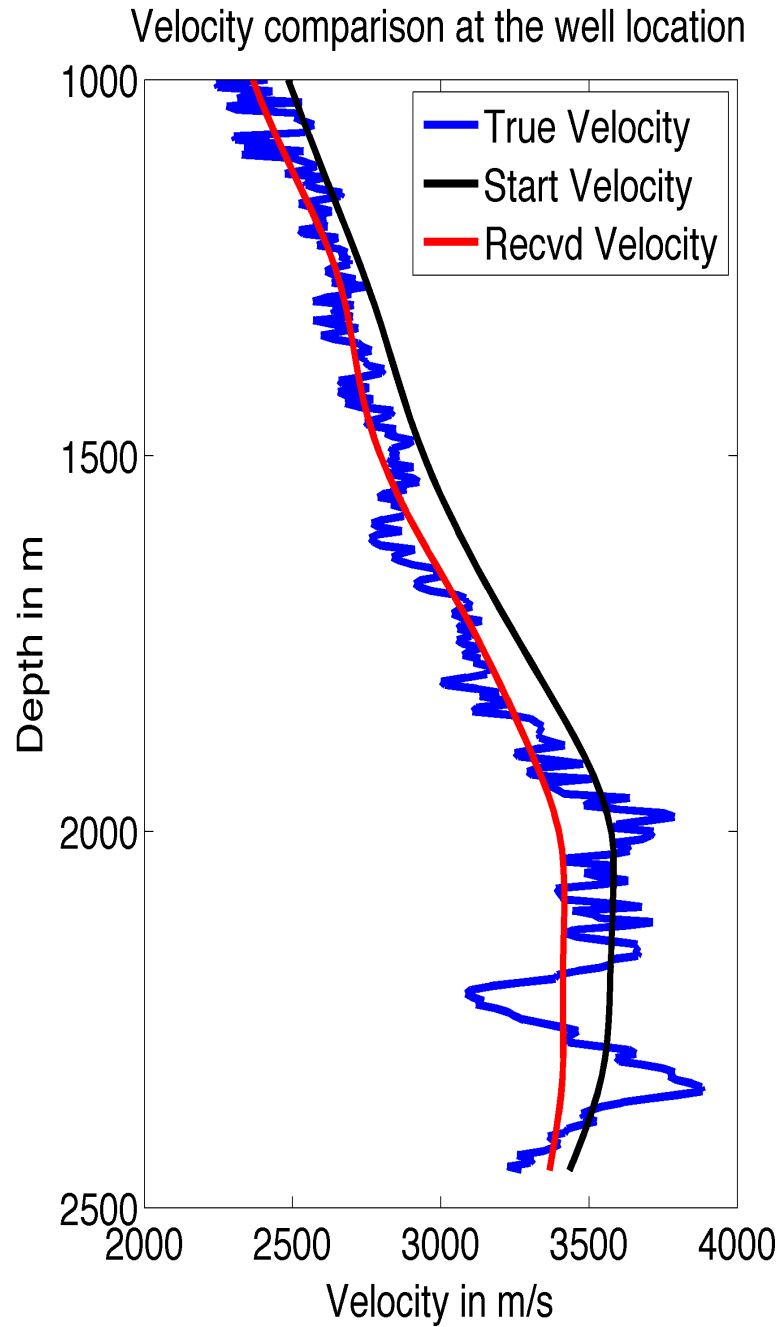


Figure 6.13: Well data comparison between the initial model, the recovered model from transmission FWI and the true model

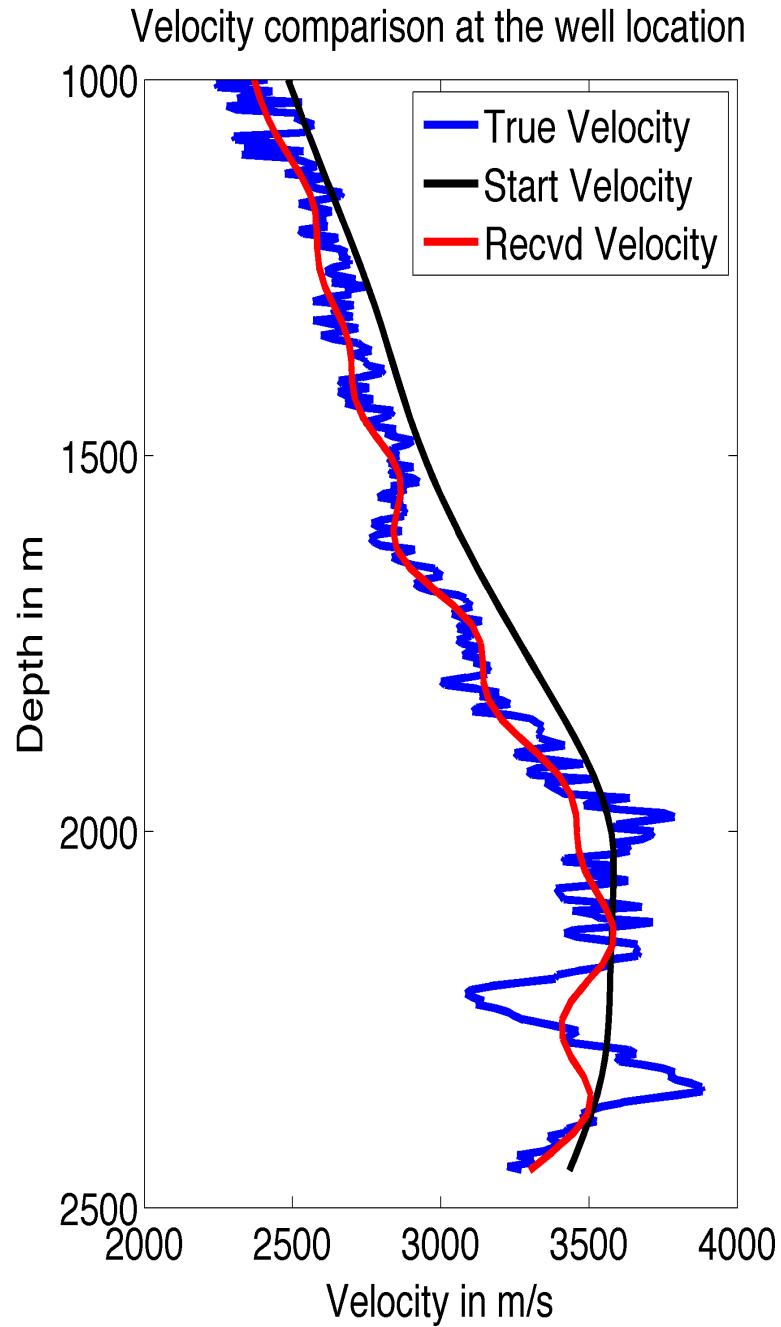


Figure 6.14: Well data comparison between the initial model, the recovered model from recovered FWI and the true model

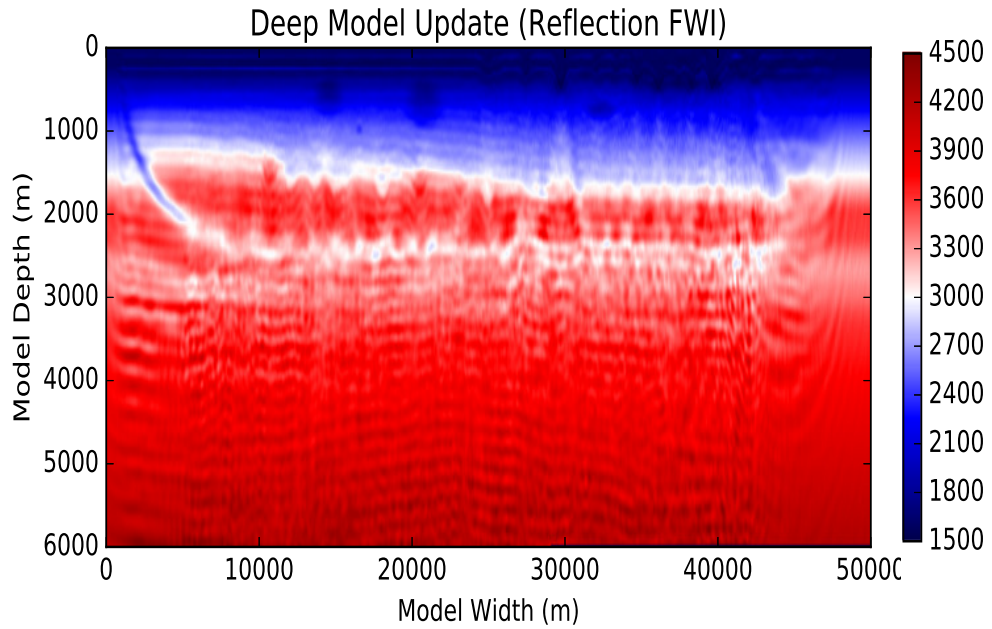


Figure 6.15: Model obtained when FWI is applied on the model obtained from transmission FWI

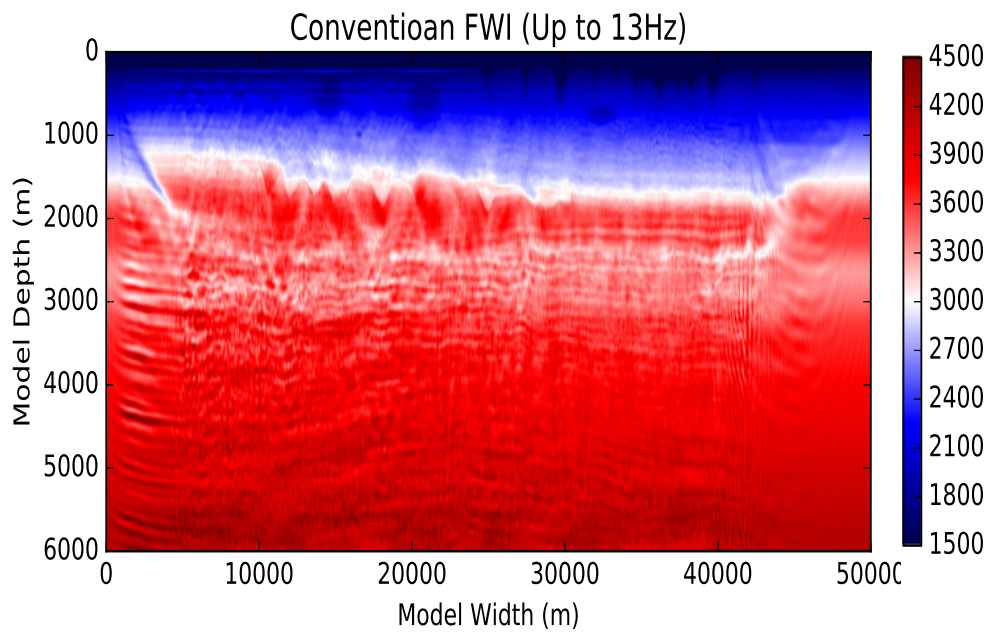


Figure 6.16: Finall results of finite difference explicit wavefield decomposition and subsequent FWI

6.5 Discussion

The results of this blind dataset have been published by a few authors ([Warner et al. \(2015\)](#), [Alkhalifah \(2016\)](#)). While [Warner et al. \(2015\)](#) made use of Adaptive Waveform Inversion (AWI) ([Warner and Guasch \(2014\)](#)), [Alkhalifah \(2016\)](#) used scattering angle filter. These results (figures [6.17](#) and [6.19](#)) match with the results gotten in this thesis (figure [6.20](#)).

A common feature with all the results is that they all built the overburden quite accurately, except for the edges, where AWI out-performs both other methods. Also, none of the methods was used to solely for the inversion of the dataset. Each method builds a model good enough to kick-start conventional FWI. FWI is then used in all cases to complete the inversion.

An advantage of FDEWD method over the other methods compared here is perhaps the minimal amount of pre-processing done on the data before commencing the inversion proper. While various inversion schemes rely of performing various filtering, muting and decimation techniques to tune the input data and make it amenable to the inversion scheme, FDEWD really does little or none of the afore-mentioned pre-processing prior to inversion.

While the potential of the scheme to invert data at reservoir depths has been shown and discussed in this chapter, indeed the scheme still has many unexplored areas and these will be surveyed in the concluding chapter.

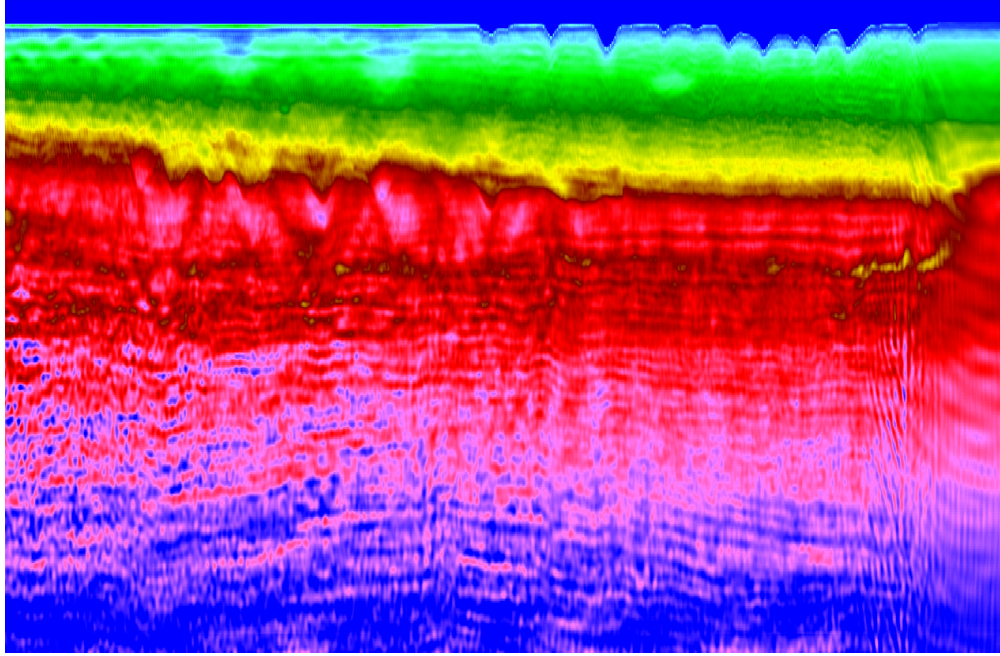


Figure 6.17: Final Inverted Model

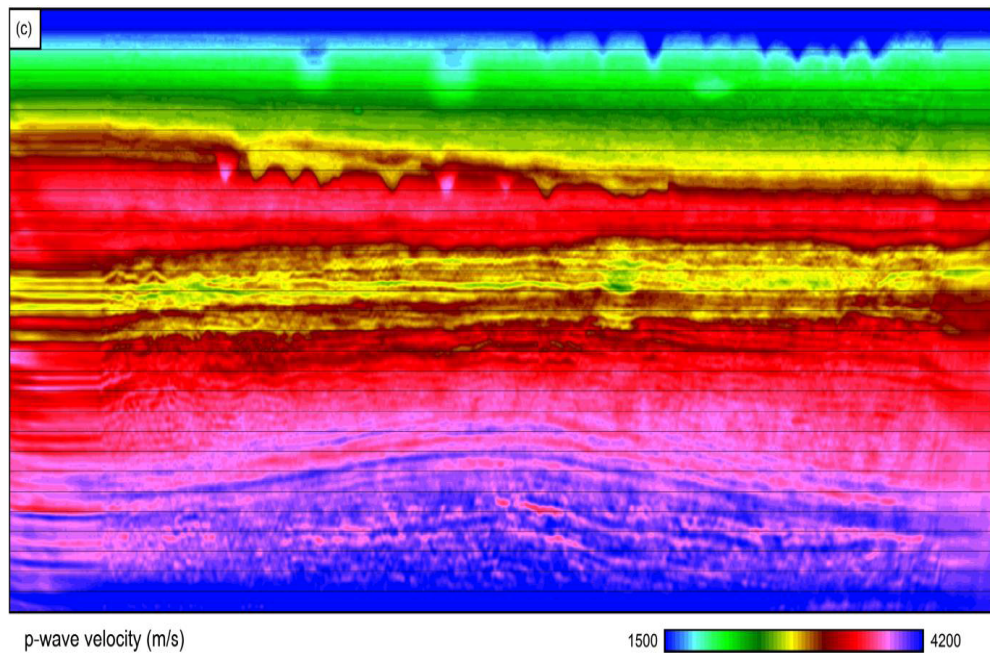


Figure 6.18: Published results on the dataset using AWI and subsequent FWI (Warner et al., 2015)

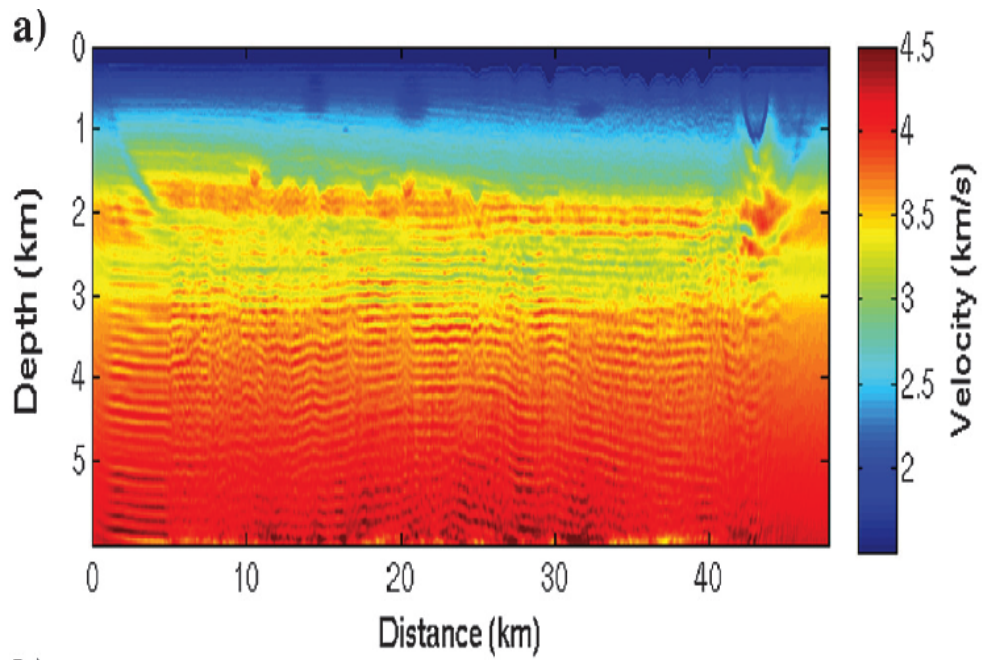


Figure 6.19: Published results on the dataset using scattering angle filters and subsequent FWI ([Alkhalifah, 2016](#))

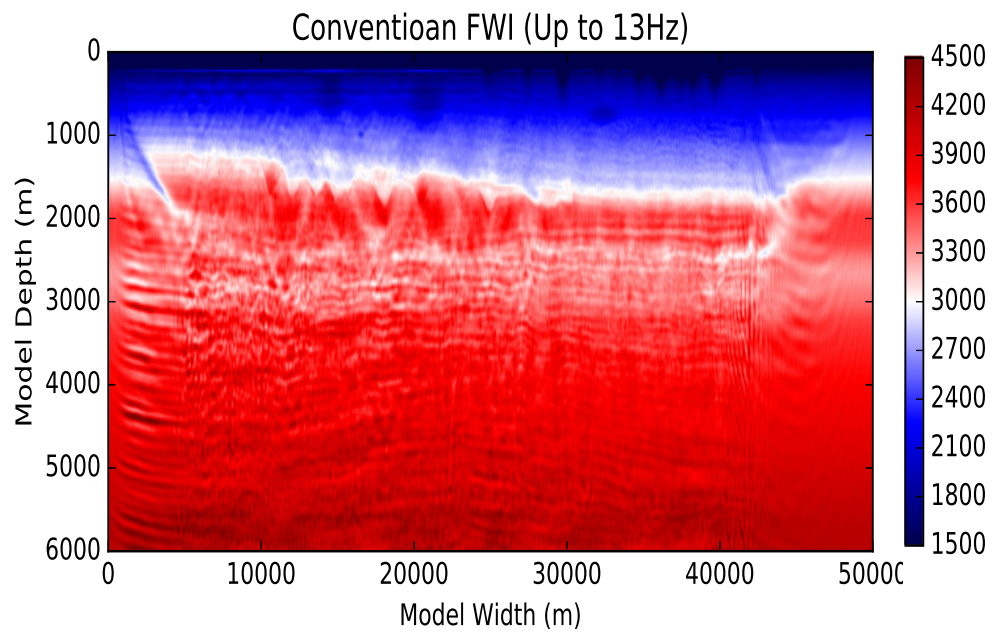


Figure 6.20: Finall results of finite difference explicit wavefield decomposition and subsequent FWI

Chapter 7

Conclusion

7.1 Conclusion

The finite difference method ([Durrán, 2013](#); [Mitchell and Griffiths, 1980](#); [Smith, 1985](#)) is widely used in the geophysics community to model the propagation of waves through material media ([Marfurt, 1984](#); [Moczo et al., 2000](#)). These wavefields have been used in Fourier decomposition schemes ([Liu et al., 2011](#); [Wang, Chauris, Donno and Calandra, 2013](#)) to achieve directional separation of wavefields. These wavefields have been known to suffer from evanescent energies due to energies propagation at right angle to the direction of separation.

The full potential of the finite difference modelling in realising a clean, seamless and evanescent free Fourier decomposition of wavefields has not been realised until now. In this thesis, it has been shown that the standard finite difference method offers clean separation of the wavefields in orthogonal directions, with very negligible noise. The scheme has been called the finite difference explicit wavefield decomposition (FDEWD). It is based on based separating the wavefields into orthogonal directions on the finite difference grid, and storing snapshots of these orthogonal wavefields ([Irabor and Warner, 2016](#)).

In this thesis, the finite difference explicit wavefield decomposition has been derived and formulated. Snapshots of two orthogonal component wavefields are

stored compared against conventional finite difference schemes that store snapshots of just one wavefield (the total wavefield). However, with compact finite difference schemes (Lele (1992), Jo et al. (1996), Chen et al. (2013)), wavefields propagating in more directions are stored. Hence, it is possible to have more separation in along these directions. Expressions for the various gradients possible with this scheme have also been derived.

This FDEWD scheme has been applied to reflection FWI in this thesis. However, it is also possible to apply this scheme in other domains of FWI as outlined in this chapter.

Reflection FWI

In this thesis, two examples of the application of the FDEWD scheme to reflection FWI have been shown. The scheme used is a variant of the MBTT scheme (Chavent and Clément, 1993). In the synthetic example shown, the performance of the method was compared to that of a conventional multi-scale FWI, and was seen to yield better performance. The experience has been carefully to mimic a real case where the starting model is poor and the recorded wavefields are lacking very low frequencies. Earlier in the iteration, the tomographic components of the wavefields were used in improving the start model. The inversion engine was later switched to conventional FWI when a reasonable start model has been achieved.

In the blind data example the FDEWD scheme accurately inverted the shallow parts of the model, as evident in the tight match between the provided well log and the inverted model. The scheme, as implemented in this thesis, was unable to invert the velocity inversion layer. It did, however, improved the deeper parts of the model.

Macro Model

The tomographic update generated by an MBTT-style inversion scheme is closely dependent on the ability of the scheme to cleanly separate wavefields into direc-

tional components. The FDEWD generates low wavenumber macro models as it is able to produce cleaner directional wavefields.

An integrated FWI scheme

Conventionally, FWI proceeds by initially muting all arrivals the transmission and turning waves. This approach is to force the inversion to makes of the turning waves and transmission in building the low wavenumbers of the model early in the inversion process ([Warner et al., 2013](#)). This method involves defining an inner geometric filter to mute the initial unwanted energies. This is a a semi-automatic process that becomes demanding as the orientation of the sea-bed changes rapid laterally.

An alternative approach is to employ the FDEWD without any Fourier decomposition. In this approach, the horizontal components of the wavefields, composed largely of transmission and turning waves, are used in the initial stages of the inversion. This approach waves the extra steps of having to define a inner geometric filter used to mute the energies.

7.2 Recommendation and Future Work

Effective Combination of Gradient Components

The thesis has exploited the FDEWD in an MBTT scheme. A scheme worth investigating is the performance of the FDEWD in a sub-space scheme ([Kennett and Williamson, 1988](#); [Oldenburg et al., 1993](#); [Rawlinson and Sambridge, 2003](#)).

FDEWD and scattering angle filters

Wavefield decomposition by scattering angle filtering ([Alkhalifah, 2016](#)) yield tomographic updates capable to building start model for conventional FWI. A study of how the FDEWD scheme can be combined with methods like the scattering

angle filtering scheme can provide additional benefits of the application of the method.

3D acoustic, Elastic and Anisotropic decomposition

This method so far has been tested with 2D dataset. It is recommended to further extend the method to 3D acoustic case, elastic case and anisotropic case. It is hoped that some other application of this scheme might be realised by such extension.

reverse time migration

Wavefield decomposition techniques have been applied previously in muting low wavenumber noise associated with reverse time migration ([Liu et al., 2011](#)). The noise is due to the tomographic component of the RTM. FDEWD scheme should be able to eliminate the low wavenumber noise of the RTM update.

Appendix A

Wavefields

A.1 Evolution of wavefields through the Marmousi model

A.1 Evolution of wavefields through the Marmousi model

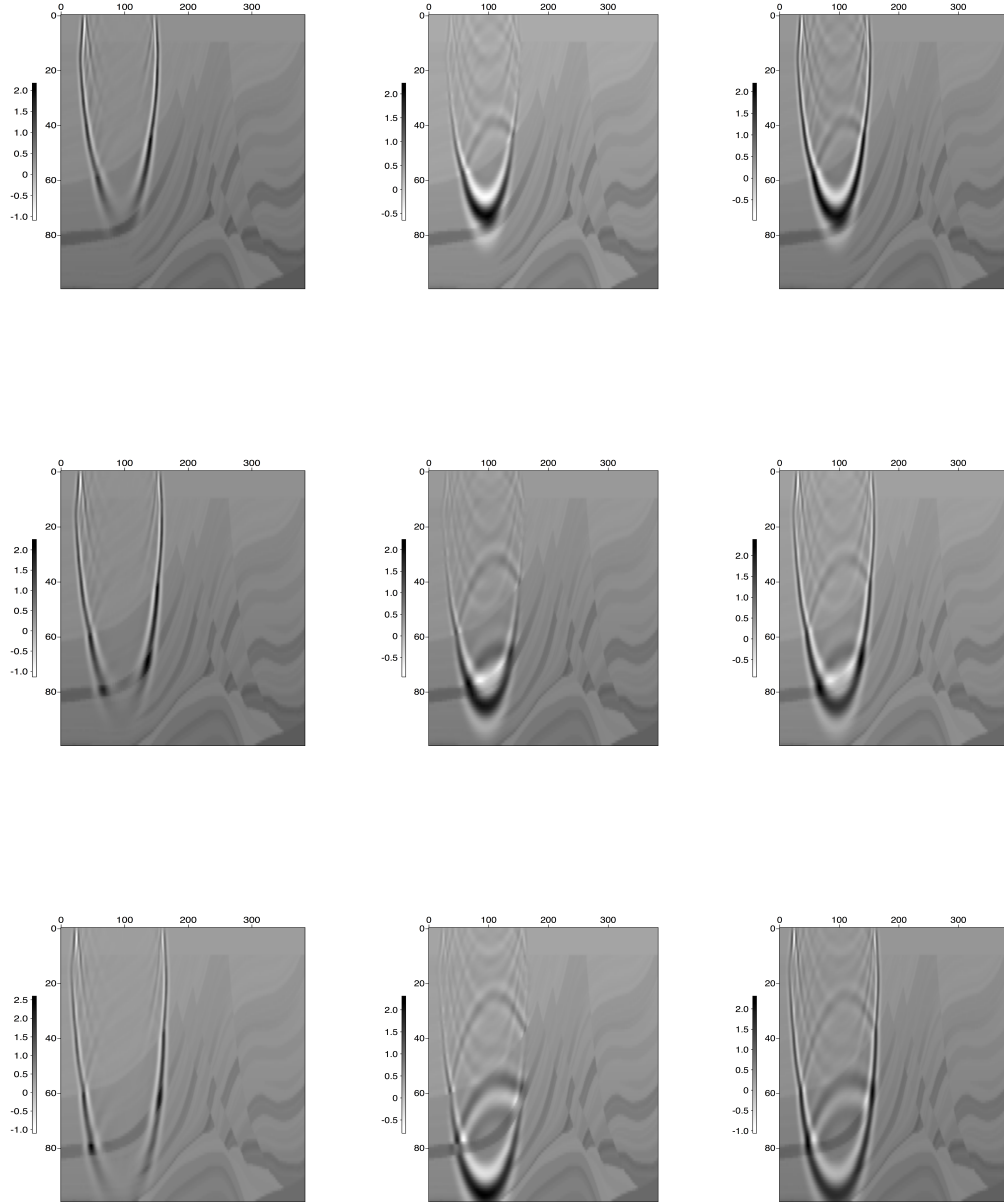


Figure A.1: Evolution of a wavefield in time through a section of the Marmousi model. Shot position is on the left side of the model

A.1 Evolution of wavefields through the Marmousi model

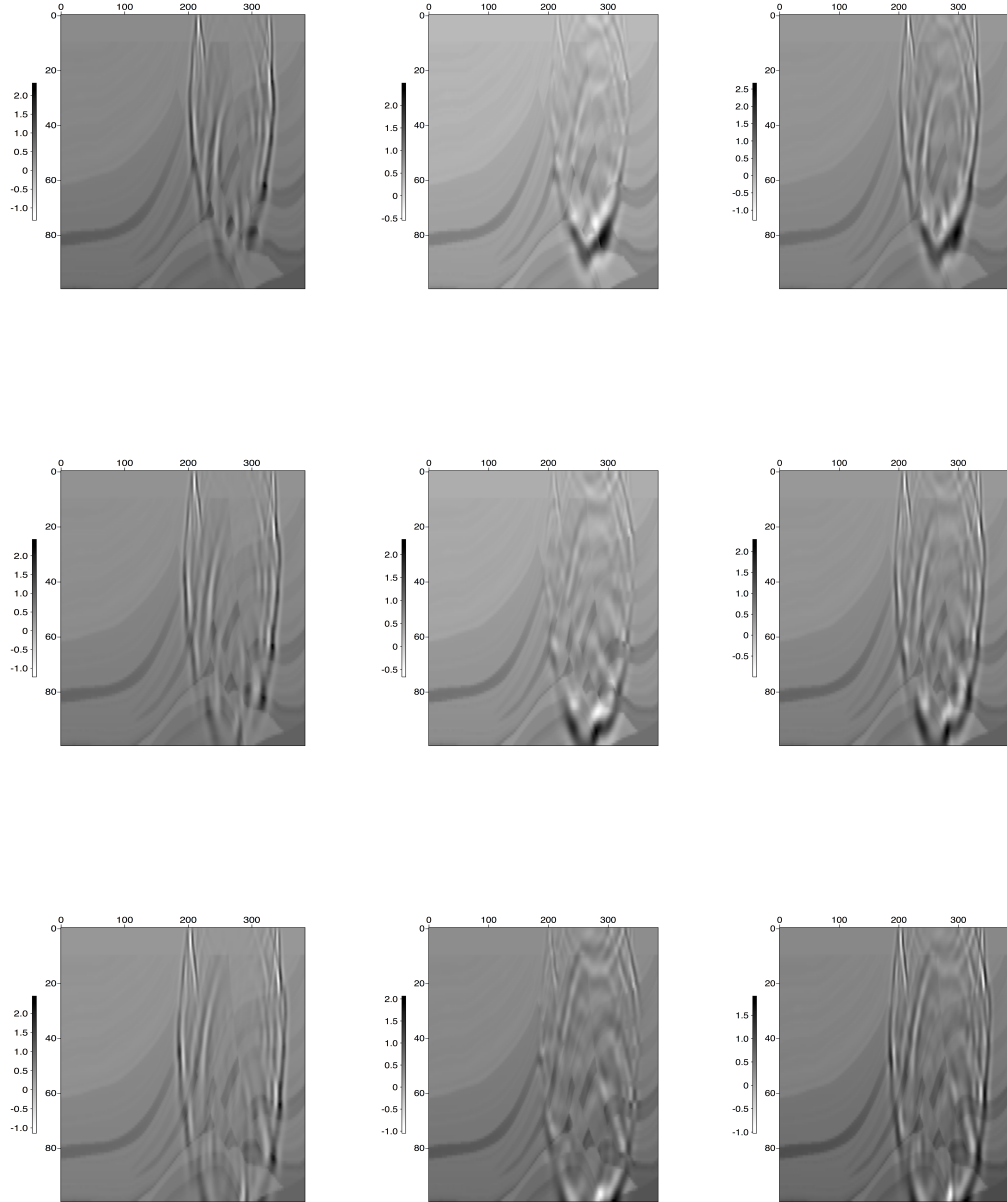
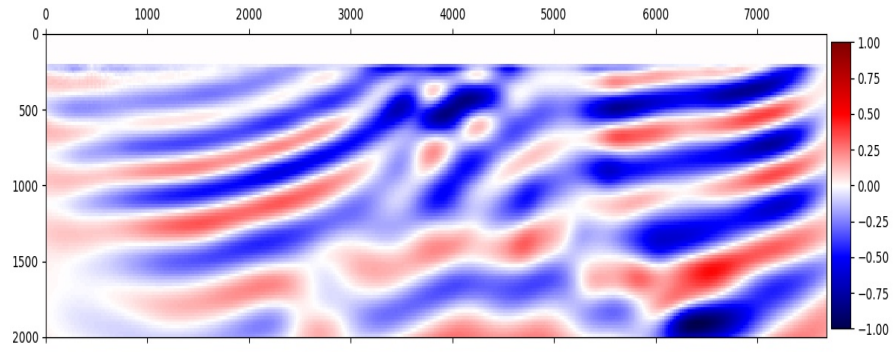


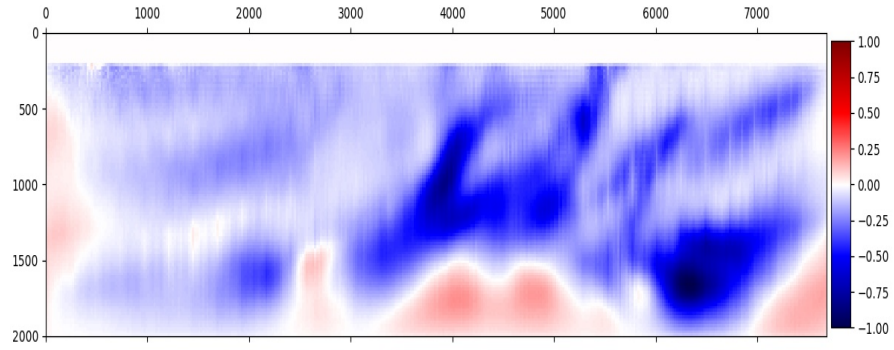
Figure A.2: Evolution of a wavefield in time through a section of the Marmousi model. Shot position is on the right side of the model

Appendix B

Intermediate Wavenumbers

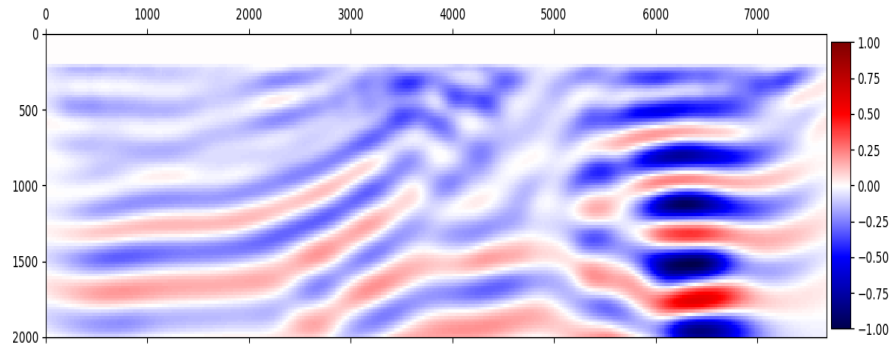


(a) Conventional FWI gradient

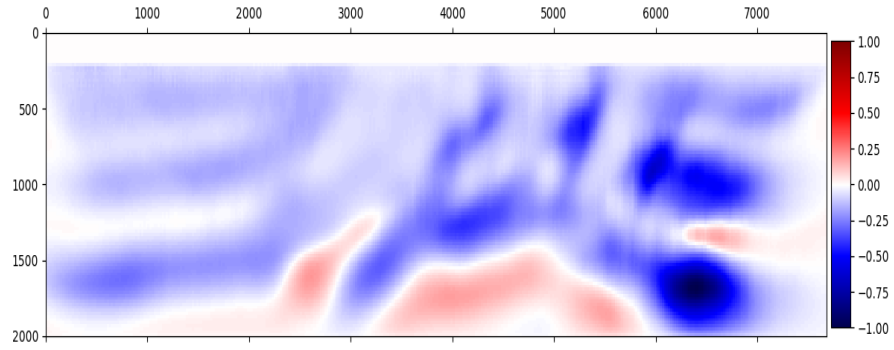


(b) FDEWD tomographic gradient

Figure B.1: Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 3Hz

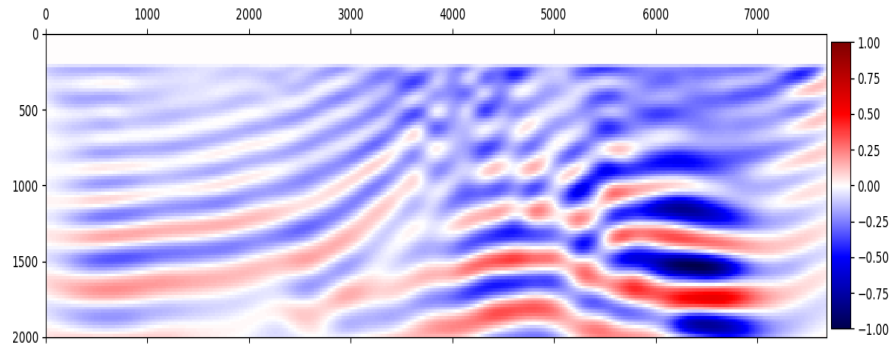


(a) Conventional FWI gradient

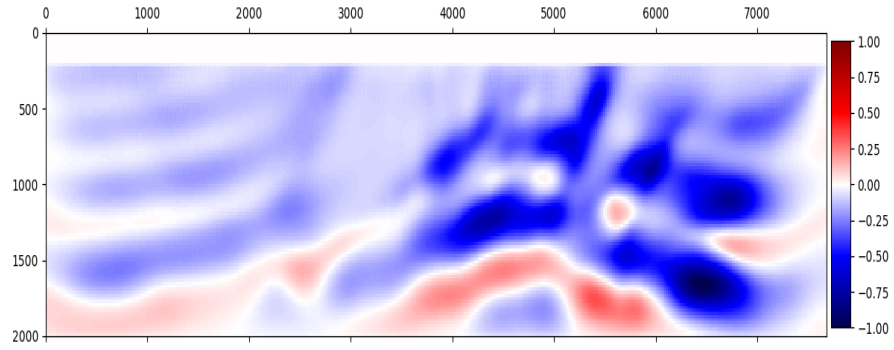


(b) FDEWD tomographic gradient

Figure B.2: Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 4Hz

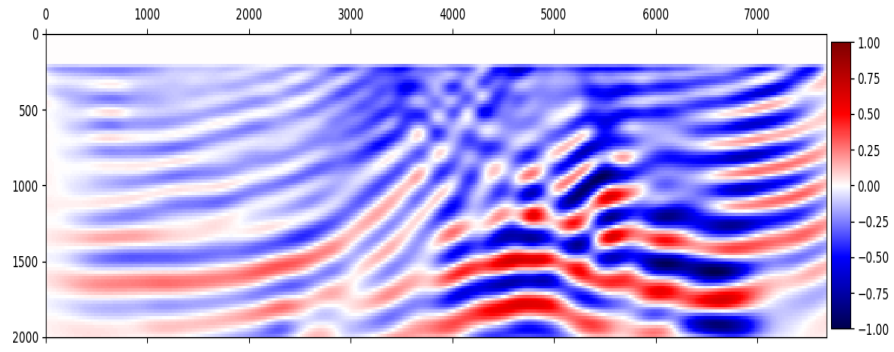


(a) Conventional FWI gradient

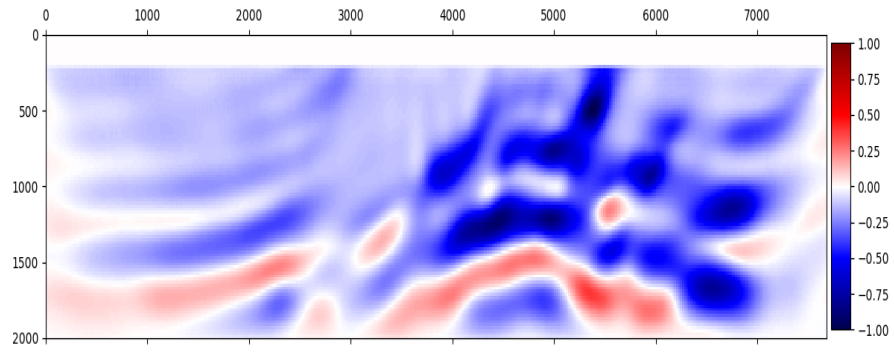


(b) FDEWD tomographic gradient

Figure B.3: Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 5Hz

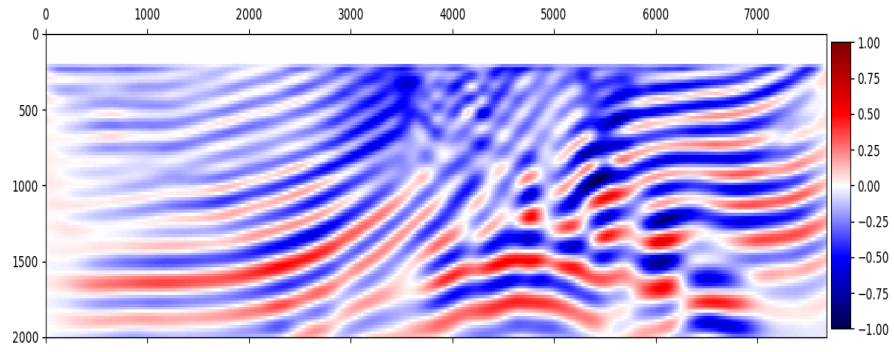


(a) Conventional FWI gradient

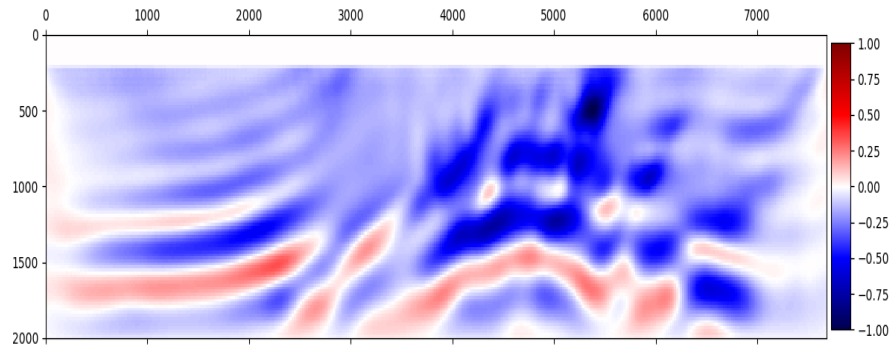


(b) FDEWD tomographic gradient

Figure B.4: Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 6Hz

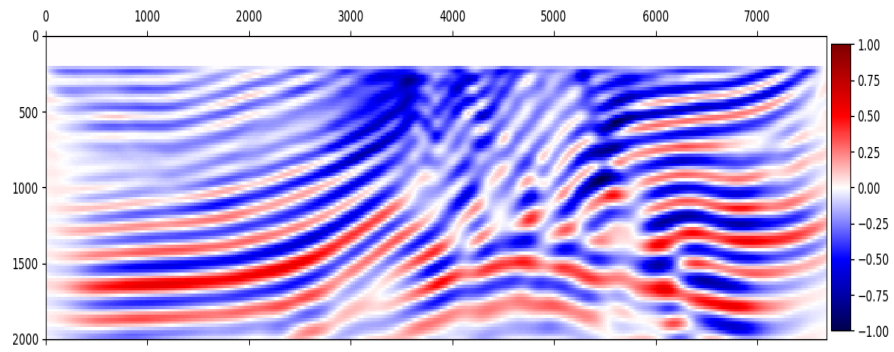


(a) Conventional FWI gradient

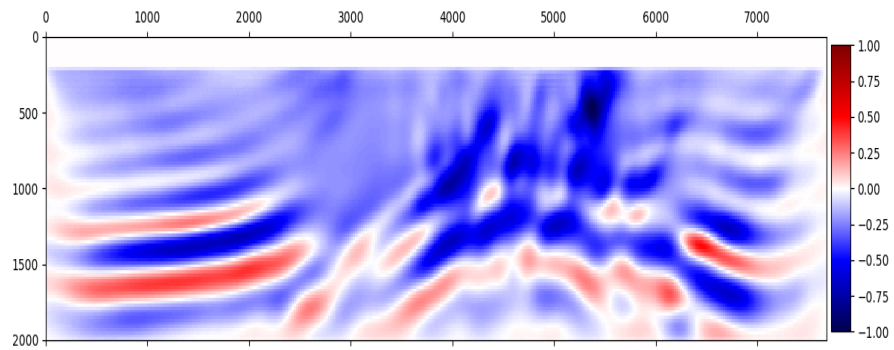


(b) FDEWD tomographic gradient

Figure B.5: Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 7Hz



(a) Conventional FWI gradient



(b) FDEWD tomographic gradient

Figure B.6: Comparison between the conventional FWI gradient and the tomographic gradient from FDEWD using only the vertical wavefields. Both gradients are generated at 8Hz

References

- Aki, K. and Richards, P. G. (2002), *Quantitative seismology*, Vol. 1. [11](#), [12](#), [29](#)
- Al-Yaqoobi, A. M. A. (2013), Full-waveform inversion to 3D seismic land data, PhD thesis, Imperial College London. [1](#)
- Alkhalifah, T. (2016), ‘Full-model wavenumber inversion: An emphasis on the appropriate wavenumber continuation’, *Geophysics* **81**(3), R89–R98. [21](#), [39](#), [40](#), [124](#), [171](#), [173](#), [176](#)
- Baeten, G., de Maag, J. W., Plessix, R.-E., Klaassen, R., Qureshi, T., Kleemeyer, M., Kroode, F. t. and Rujie, Z. (2013), ‘The use of low frequencies in a full-waveform inversion and impedance inversion land seismic case study’, *Geophysical Prospecting* **61**(4), 701–711. [77](#)
- Bal, G. (2012), ‘Introduction to inverse problems’, *Lecture Notes-Department of Applied Physics and Applied Mathematics, Columbia University, New York* . [11](#)
- Bartholomew-Biggs, M. (2006), *Nonlinear optimization with financial applications*, Springer Science & Business Media. [17](#)
- Bass, L. (2007), *Software architecture in practice*, Pearson Education India. [88](#)
- Berenger, J.-P. (1994), ‘A perfectly matched layer for the absorption of electromagnetic waves’, *Journal of computational physics* **114**(2), 185–200. [81](#)
- Bernholdt, D. E., Allan, B. A., Armstrong, R., Bertrand, F., Chiu, K., Dahlgren, T. L., Damevski, K., Elwasif, W. R., Epperly, T. G., Govindaraju, M. et al.

REFERENCES

- (2006), ‘A component architecture for high-performance scientific computing’, *The International Journal of High Performance Computing Applications* **20**(2), 163–202. [88](#)
- Bertsekas, D. P. (2014), *Constrained optimization and Lagrange multiplier methods*, Academic press. [17](#)
- Billette, F. and Lambaré, G. (1998), ‘Velocity macro-model estimation from seismic reflection data by stereotomography’, *Geophysical Journal International* **135**(2), 671–690. [5](#)
- Bishop, T., Bube, K., Cutler, R., Langan, R., Love, P., Resnick, J., Shuey, R., Spindler, D. and Wyld, H. (1985), ‘Tomographic determination of velocity and depth in laterally varying media’, *Geophysics* **50**(6), 903–923. [5](#)
- Bleistein, N., Cohen, J. K., John Jr, W. et al. (2013), *Mathematics of multidimensional seismic imaging, migration, and inversion*, Vol. 13, Springer Science & Business Media. [29](#)
- Boschetti, F., Dentith, M. C. and List, R. D. (1996), ‘Inversion of seismic refraction data using genetic algorithms’, *Geophysics* **61**(6), 1715–1727. [6](#)
- Boyd, S. and Vandenberghe, L. (2004), *Convex optimization*, Cambridge university press. [20](#)
- Bracewell, R. N. and Bracewell, R. N. (1986), *The Fourier transform and its applications*, Vol. 31999, McGraw-Hill New York. [54](#)
- Brainerd, W. S. et al. (2009), *Guide to Fortran 2003 programming*, Springer. [87](#)
- Brossier, R., Operto, S. and Virieux, J. (2009), ‘Seismic imaging of complex on-shore structures by 2d elastic frequency-domain full-waveform inversion’, *Geophysics* **74**(6), WCC105–WCC118. [1](#)
- Bunks, C., Saleck, F. M., Zaleski, S. and Chavent, G. (1995), ‘Multiscale seismic waveform inversion’, *Geophysics* **60**(5), 1457–1473. [6](#), [7](#), [121](#)

REFERENCES

- Callahan, J., Cox, D., Hoffman, K., O'shea, D., Pollatsek, H. and Senechal, L. (1995), *Calculus in context: The five college calculus project*, WH Freeman. 15
- Chandra, R. (2001), *Parallel programming in OpenMP*, Morgan kaufmann. 88
- Chapman, B., Jost, G. and Van Der Pas, R. (2008), *Using OpenMP: portable shared memory parallel programming*, Vol. 10, MIT press. 88
- Chapman, C. (2004), *Fundamentals of seismic wave propagation*, Cambridge university press. 29
- Chapman, C. t. and Pratt, R. (1992), 'Traveltime tomography in anisotropic media: theory', *Geophysical Journal International* **109**(1), 1–19. 5
- Chapman, S. J. (2008), *Fortran for scientist and engineers: 1995-2003*, McGraw-Hill. 87
- Chavent, G. and Clément, F. (1993), Waveform inversion through MBTT formulation, PhD thesis, INRIA. 52, 121, 175
- Chavent, G., Clément, F., Gómez, S. et al. (1994), Automatic determination of velocities via migration-based traveltime waveform inversion: A synthetic data example, in 'SEG Technical Program Expanded Abstracts', Vol. 328, pp. 1179–1182. 8, 52
- Chen, J.-B. (2007), 'High-order time discretizations in seismic modeling', *Geophysics* **72**(5), SM115–SM122. 32
- Chen, J.-B. (2011), 'A stability formula for lax-wendroff methods with fourth-order in time and general-order in space for the scalar wave equation', *Geophysics* **76**(2), T37–T42. 32
- Chen, Z., Cheng, D., Feng, W. and Wu, T. (2013), 'An optimal 9-point finite difference scheme for the helmholtz equation with pml', *Int. J. Numer. Anal. Model* **10**, 389–410. 175
- Chi, B., Dong, L. and Liu, Y. (2015), 'Correlation-based reflection full-waveform inversion', *Geophysics* **80**(4), R189–R202. 8, 52

REFERENCES

- Chivers, I. D. and Sleightholme, J. (2006), *Introduction to Programming with FORTRAN*, Springer. [87](#)
- Claerbout, J. F. (1985), *Imaging the earth's interior*, Vol. 6. [7](#), [77](#)
- Cohen, J. K. and Bleistein, N. (1977), A velocity inversion procedure for acoustic waves, Technical report. [4](#)
- Collino, F. and Tsogka, C. (2001), ‘Application of the perfectly matched absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media’, *Geophysics* **66**(1), 294–307. [81](#)
- Courant, R. and Hilbert, D. (1965), *Methods of mathematical physics*, Vol. 1, CUP Archive. [16](#)
- Crnkovic, I. (2001), ‘Component-based software engineering new challenges in software development’, *Software Focus* **2**(4), 127–133. [88](#)
- da Silva, N. V., Ratcliffe, A., Vinje, V. and Conroy, G. (2016), ‘A new parameter set for anisotropic multiparameter full-waveform inversion and application to a north sea data set’, *Geophysics* **81**(4), U25–U38. [1](#)
- Dablain, M. (1986), ‘The application of high-order differencing to the scalar wave equation’, *Geophysics* **51**(1), 54–66. [32](#)
- Dagum, L. and Menon, R. (1998), ‘Openmp: an industry standard api for shared-memory programming’, *IEEE computational science and engineering* **5**(1), 46–55. [88](#)
- Datta, D. and Sen, M. K. (2016), ‘Estimating a starting model for full-waveform inversion using a global optimization method’, *Geophysics* . [6](#)
- Drábek, P. and Milota, J. (2013), *Methods of nonlinear analysis: applications to differential equations*, Springer Science & Business Media. [20](#), [23](#)
- Durran, D. R. (2013), *Numerical methods for wave equations in geophysical fluid dynamics*, Vol. 32, Springer Science & Business Media. [174](#)

REFERENCES

- Etgen, J. T. and OBrien, M. J. (2007), ‘Computational methods for large-scale 3d acoustic finite-difference modeling: A tutorial’, *Geophysics* **72**(5), SM223–SM230. [32](#), [33](#)
- Fichtner, A., Kennett, B. L., Igel, H. and Bunge, H.-P. (2009), ‘Full seismic waveform tomography for upper-mantle structure in the australasian region using adjoint methods’, *Geophysical Journal International* **179**(3), 1703–1725. [1](#)
- Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y. and Villaseñor, A. (2013), ‘Multiscale full waveform inversion’, *Geophysical Journal International* p. ggt118. [1](#)
- Fletcher, R. (2013), *Practical methods of optimization*, John Wiley & Sons. [20](#), [23](#)
- Floudas, C. A. and Pardalos, P. M. (2001), *Encyclopedia of optimization*, Springer Science & Business Media. [17](#)
- Fokkema, J. T. and Berg, P. M. (1989), ‘Seismic inversion by a rms born approximation in the space-time domain¹’, *Geophysical prospecting* **37**(1), 53–72. [4](#)
- Garlan, D. and Shaw, M. (1993), ‘An introduction to software architecture’, *Advances in software engineering and knowledge engineering* **1**(3.4). [88](#)
- Gauthier, O., Virieux, J. and Tarantola, A. (1986), ‘Two-dimensional nonlinear inversion of seismic waveforms: Numerical results’, *Geophysics* **51**(7), 1387–1403. [5](#), [7](#), [41](#)
- Gelfand, I. M., Silverman, R. A. et al. (2000), *Calculus of variations*, Courier Corporation. [24](#)
- Gill, P. E., Murray, W. and Wright, M. H. (1981), ‘Practical optimization’. [17](#)
- Gisolf, D. and Vershuur, E. (2010), *The principles of quantitative acoustical imaging*. [29](#), [30](#)

REFERENCES

- Gockenbach, M. S. (2010), *Partial differential equations: analytical and numerical methods*, Siam. [30](#)
- Heinbockel, J. (2006), *Introduction to the variational calculus*, Trafford Publishing. [16](#), [24](#)
- Heineman, G. T. and Councill, W. T. (2001), ‘Component-based software engineering’, *Putting the pieces together, addison-westley* p. 5. [88](#)
- Hijab, O. and Hijab, O. (2007), *Introduction to calculus and classical analysis*, Springer. [15](#)
- Hu, L.-Z. and McMechan, G. A. (1987), ‘Wave-field transformations of vertical seismic profiles’, *Geophysics* **52**(3), 307–321. [56](#), [85](#)
- Hudson, J. and Heritage, J. (1981), ‘The use of the born approximation in seismic scattering problems’, *Geophysical Journal International* **66**(1), 221–240. [6](#)
- Ikelle, L. T. and Amundsen, L. (2005), *Introduction to petroleum seismology*, Society of Exploration Geophysicists Tulsa, Oklahoma. [11](#)
- Irabor, K. and Warner, M. (2016), Reflection fwi, in ‘SEG Technical Program Expanded Abstracts 2016’, Society of Exploration Geophysicists, pp. 1136–1140. [174](#)
- Isakov, V. (2006), *Inverse problems for partial differential equations*, Vol. 127, Springer Science & Business Media. [11](#)
- Jannane, M., Beydoun, W., Crase, E., Cao, D., Koren, Z., Landa, E., Mendes, M., Pica, A., Noble, M., Roeth, G. et al. (1989), ‘Wavelengths of earth structures that can be resolved from seismic reflection data’, *Geophysics* **54**(7), 906–910. [77](#)
- Jin, S. and Madariaga, R. (1993), ‘Background velocity inversion with a genetic algorithm’, *Geophysical research letters* **20**(2), 93–96. [6](#)
- Jo, C.-H., Shin, C. and Suh, J. H. (1996), ‘An optimal 9-point, finite-difference, frequency-space, 2-d scalar wave extrapolator’, *Geophysics* **61**(2), 529–537. [175](#)

REFERENCES

- Jost, J. and Li-Jost, X. (1998), *Calculus of variations*, Vol. 64, Cambridge University Press. [24](#)
- Jovanović, B. S. and Süli, E. (2013), *Analysis of finite difference schemes: for linear partial differential equations with generalized solutions*, Vol. 46, Springer Science & Business Media. [30](#)
- Kemball, A., Crutcher, R. and Hasan, R. (2008), ‘A component-based framework for radio-astronomical imaging software systems’, *Software: Practice and Experience* **38**(5), 493–507. [88](#)
- Kennett, B. and Williamson, P. (1988), Subspace methods for large-scale nonlinear inversion, in ‘Mathematical geophysics’, Springer, pp. 139–154. [83](#), [176](#)
- Kirsch, A. (2011), *An introduction to the mathematical theory of inverse problems*, Vol. 120, Springer Science & Business Media. [11](#)
- Komatitsch, D. and Tromp, J. (2003), ‘A perfectly matched layer absorbing boundary condition for the second-order seismic wave equation’, *Geophysical Journal International* **154**(1), 146–153. [81](#)
- Kreyszig, E. (2005), *Advanced engineering mathematics*, John Wiley & Sons. [24](#)
- Kreyszig, E. (2007), *Advanced engineering mathematics*, John Wiley & Sons. [16](#)
- Krosche, M. and Hautefeuille, M. (2007), ‘Simulation and solution of stochastic systems with a component-based software design’, *PAMM* **7**(1), 2140001–2140002. [88](#)
- Lailly, P. (1983), The seismic inverse problem as a sequence of before stack migrations, in ‘Conference on Inverse Scattering, Theory and Application’, Society for Industrial and Applied Mathematics, pp. 206–220. [5](#)
- Lange, K. (2004), Elementary optimization, in ‘Optimization’, Springer. [17](#), [18](#)
- Lele, S. K. (1992), ‘Compact finite difference schemes with spectral-like resolution’, *Journal of computational physics* **103**(1), 16–42. [175](#)

REFERENCES

- Liberti, L. and Maculan, N. (2006), *Global optimization: from theory to implementation*, Vol. 84, Springer Science & Business Media. [18](#)
- Lindsey, J. (1960), ‘Elimination of seismic ghost reflections by means of a linear filter’, *Geophysics* **25**(1), 130–140. [7](#), [41](#)
- Lines, L. R., Slawinski, R. and Bording, P. R. (n.d.), A recipe for stability analysis of finite-difference wave equations computations, Technical report, CREWES Research Report. [34](#)
- Liu, F., Zhang, G., Morton, S. and Leveille, J. (2011), ‘An effective imaging condition for reverse-time migration using wavefield decomposition: Geophysics, 76’, *S29–S39*. [59](#), [85](#), [174](#), [177](#)
- Liu, Q.-H. and Tao, J. (1997), ‘The perfectly matched layer for acoustic waves in absorptive media’, *The Journal of the Acoustical Society of America* **102**(4), 2072–2082. [81](#)
- Liu, Y. and Sen, M. (2009), 2d acoustic wave equation modeling with a new high-accuracy time-space domain finite-difference stencil, *in* ‘71st EAGE Conference and Exhibition incorporating SPE EUROPEC 2009’. [32](#)
- Long, G., Zhao, Y. and Zou, J. (2013), ‘A temporal fourth-order scheme for the first-order acoustic wave equations’, *Geophysical Journal International* **194**(3), 1473–1485. [32](#)
- Luo, S. and Sava, P. (2011), ‘A deconvolution-based objective function for wave-equation inversion: 81st annual international meeting, seg, expanded abstracts, 2788–2792’. [28](#)
- Ma, Y. and Hale, D. (2013), ‘Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion’, *Geophysics* **78**(6), R223–R233. [8](#), [52](#), [53](#)
- Marfurt, K. J. (1984), ‘Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations’, *Geophysics* **49**(5), 533–549. [174](#)
- Markus, A. (2012), *Modern Fortran in practice*, Cambridge University Press. [87](#)

REFERENCES

- Metcalf, M., Reid, J. and Cohen, M. (2011), *Modern Fortran Explained*, Oxford University Press. [87](#)
- Michiels, W., Aarts, E. and Korst, J. (2007), *Theoretical aspects of local search*, Springer Science & Business Media. [18](#)
- Mitchell, A. R. and Griffiths, D. F. (1980), *The finite difference method in partial differential equations*, John Wiley. [174](#)
- Moczo, P., Kristek, J. and Halada, L. (2000), ‘3d fourth-order staggered-grid finite-difference schemes: Stability and grid dispersion’, *Bulletin of the Seismological Society of America* **90**(3), 587–603. [174](#)
- Moczo, P., Robertsson, J. O. and Eisner, L. (2007), ‘The finite-difference time-domain method for modeling of seismic wave propagation’, *Advances in Geophysics* **48**, 421–516. [32](#)
- Mora, P. (1987), ‘Nonlinear two-dimensional elastic inversion of multioffset seismic data’, *Geophysics* **52**(9), 1211–1228. [8](#)
- Mora, P. (1989), Inversion= migration+ tomography, in ‘Parallel Computing 1988’, Springer, pp. 78–101. [8](#), [52](#)
- Morton, K. and Mayers, D. (1998), ‘Numerical solution of partial differential equations’, *Journal of Fluid Mechanics* **363**, 349–349. [30](#)
- Nocedal, J. and Wright, S. (2006), *Numerical optimization*, Springer Science & Business Media. [5](#), [17](#), [18](#), [23](#), [26](#)
- Oldenburg, D., McGillivray, P. and Ellis, R. (1993), ‘Generalized subspace methods for large-scale inverse problems’, *Geophysical Journal International* **114**(1), 12–20. [176](#)
- Oldenburg, D. W. and Li, Y. (1994), ‘Subspace linear inverse method’, *Inverse problems* **10**(4), 915. [83](#)
- Pintér, J. D. (2006), *Global optimization: scientific and engineering case studies*, Vol. 85, Springer Science & Business Media. [18](#)

REFERENCES

- Plessix, R.-E. (2006), ‘A review of the adjoint-state method for computing the gradient of a functional with geophysical applications’, *Geophysical Journal International* **167**(2), 495–503. [37](#)
- Plessix, R.-E., Baeten, G., de Maag, J. W., Klaassen, M., Rujie, Z. and Zhifei, T. (2010), Application of acoustic full waveform inversion to a low-frequency large-offset land data set, *in* ‘SEG Technical Program Expanded Abstracts 2010’, Society of Exploration Geophysicists, pp. 930–934. [77](#)
- Pratt, R. G., Shin, C. and Hick, G. (1998), ‘Gauss–newton and full newton methods in frequency–space seismic waveform inversion’, *Geophysical Journal International* **133**(2), 341–362. [6](#)
- Pratt, R., Song, Z.-M., Williamson, P. and Warner, M. (1996), ‘Two-dimensional velocity models from wide-angle seismic data by wavefield inversion’, *Geophysical Journal International* **124**(2), 323–340. [5](#)
- Quintino, T. (2008), A component environment for high-performance scientific computing: design and implementation, PhD thesis, Ph. D. Thesis, Katholieke Universiteit Leuven, von Karman Institute for Fluid Dynamics, Leuven. [88](#)
- Rawlinson, N. and Sambridge, M. (2003), ‘Seismic traveltime tomography of the crust and lithosphere’, *Advances in Geophysics* **46**, 81–199. [84](#), [176](#)
- Riley, K. F., Hobson, M. P. and Bence, S. J. (2006), *Mathematical methods for physics and engineering: a comprehensive guide*, Cambridge university press. [24](#)
- Rosenbrock, H. (1960), ‘An automatic method for finding the greatest or least value of a function’, *The Computer Journal* **3**(3), 175–184. [27](#)
- Schäffler, S. (2012), *Global optimization: a stochastic approach*, Springer Science & Business Media. [17](#), [18](#)
- Sen, M. K. and Stoffa, P. L. (1991), ‘Nonlinear one-dimensional seismic waveform inversion using simulated annealing’, *Geophysics* **56**(10), 1624–1638. [6](#)

REFERENCES

- Shearer, P. M. (2009), *Introduction to seismology*, Cambridge University Press. [29](#)
- Sirgue, L., Barkved, O., Van Gestel, J., Askim, O. and Kommedal, J. (2009), 3d waveform inversion on valhall wide-azimuth obc, in ‘71st EAGE Conference and Exhibition incorporating SPE EUROPEC 2009’. [6](#)
- Smith, G. D. (1985), *Numerical solution of partial differential equations: finite difference methods*, Oxford university press. [30](#), [174](#)
- Snyman, J. (2005), *Practical mathematical optimization: an introduction to basic optimization theory and classical and new gradient-based algorithms*, Vol. 97, Springer Science & Business Media. [17](#)
- Sorenson, H. W. (1980), *Parameter estimation: principles and problems*, Vol. 9, M. Dekker. [12](#)
- Stolt, R. H. and Weglein, A. B. (2012), *Seismic imaging and inversion: Volume 1: Application of linear inverse theory*, Vol. 1, Cambridge University Press. [29](#)
- Symes, W. and Carazzone, J. J. (1991), ‘Velocity inversion by differential semblance optimization’, *Geophysics* **56**(5), 654–663. [5](#)
- Szyperski, C., Bosch, J. and Weck, W. (1999), ‘Component-oriented programming’, *Lecture Notes in Computer Science* **1743**, 184–184. [88](#)
- Tang, Y., Lee, S., Baumstein, A., Hinkley, D. et al. (2013), Tomographically enhanced full waveform inversion, in ‘2013 SEG Annual Meeting’, Society of Exploration Geophysicists. [8](#), [52](#)
- Tarantola, A. (1984), ‘Linearized inversion of seismic reflection data’, *Geophysical prospecting* **32**(6), 998–1015. [5](#), [6](#)
- Tarantola, A. (2005), *Inverse problem theory and methods for model parameter estimation*, siam. [12](#)

REFERENCES

- van Leeuwen, T. and Herrmann, F. J. (2013), ‘Mitigating local minima in full-waveform inversion by expanding the search space’, *Geophysical Journal International* **195**(1), 661–667. [28](#)
- Van Leeuwen, T. and Mulder, W. (2010), ‘A correlation-based misfit criterion for wave-equation travelttime tomography’, *Geophysical Journal International* **182**(3), 1383–1394. [28](#)
- Versteeg, R. (1994), ‘The marmousi experience: Velocity model determination on a synthetic complex data set’, *The Leading Edge* **13**(9), 927–936. [103](#)
- Virieux, J. and Operto, S. (2009), ‘An overview of full-waveform inversion in exploration geophysics’, *Geophysics* **74**(6), WCC1–WCC26. [5](#), [30](#)
- Wang, F., Chauris, H., Donno, D. and Calandra, H. (2013), ‘Taking advantage of wavefield decomposition in full waveform inversion: 75th annual international conference and exhibition, eage’, *Extended Abstracts, paper Tu0708* . [53](#), [59](#), [174](#)
- Wang, F., Donno, D., Chauris, H., Calandra, H. and Audebert, F. (2015), Full waveform inversion based on wave field decomposition, in ‘2015 SEG Summer Research Workshop’. [59](#)
- Wang, S., Chen, F., Zhang, H., Shen, Y. et al. (2013), Reflection-based full waveform inversion (RFWI) in the frequency domain, in ‘SEG Technical Program Expanded Abstracts’, pp. 877–881. [8](#), [52](#)
- Wang, Y. and Rao, Y. (2009), ‘Reflection seismic waveform tomography’, *Journal of Geophysical Research: Solid Earth* **114**(B3). [8](#), [52](#)
- Warner, M. and Guasch, L. (2014), Adaptive waveform inversion-fwi without cycle skipping-theory, in ‘76th EAGE Conference and Exhibition 2014’. [171](#)
- Warner, M., Guasch, L. and Yao, G. (2015), Adaptive waveform inversion–fwi without cycle skipping, in ‘2015 Workshop: Depth Model Building: Full-waveform Inversion, Beijing, China, 18-19 June 2015’, Society of Exploration Geophysicists, pp. 11–14. [21](#), [171](#), [172](#)

REFERENCES

- Warner, M., Morgan, J., Umpleby, A., Stekl, I. and Guasch, L. (2012), Which physics for full-wavefield seismic inversion?, *in* ‘74th EAGE Conference and Exhibition incorporating EUROPEC 2012’. [30](#)
- Warner, M., Ratcliffe, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N., Vinje, V., Štekl, I., Guasch, L., Win, C. et al. (2013), ‘Anisotropic 3d full-waveform inversion’, *Geophysics* **78**(2), R59–R80. [12](#), [1](#), [2](#), [6](#), [7](#), [176](#)
- Warner, M., Stekl, I. and Umpleby, A. (2008), Efficient and effective 3d wavefield tomography, *in* ‘70th EAGE Conference and Exhibition incorporating SPE EUROPEC 2008’. [1](#)
- Weglein, A. B., Araújo, F. V., Carvalho, P. M., Stolt, R. H., Matson, K. H., Coates, R. T., Corrigan, D., Foster, D. J., Shaw, S. A. and Zhang, H. (2003), ‘Inverse scattering series and seismic exploration’, *Inverse problems* **19**(6), R27. [6](#)
- Woodward, M. J., Nichols, D., Zdraveva, O., Whitfield, P. and Johns, T. (2008), ‘A decade of tomography’, *Geophysics* **73**(5), VE5–VE11. [5](#)
- Xu, S., Chen, F., Lambaré, G., Zhang, Y., Wang, D. et al. (2012), Inversion on reflected seismic wave, *in* ‘2012 SEG Annual Meeting’, Society of Exploration Geophysicists. [8](#), [52](#), [53](#), [85](#), [121](#)
- Xu, S., Wang, D., Chen, F., Zhang, Y. and Lambare, G. (2012), Full waveform inversion for reflected seismic data, *in* ‘74th EAGE Conference and Exhibition incorporating EUROPEC 2012’. [40](#)
- Yang, J., Liu, Y. and Dong, L. (2014), A multi-parameter full waveform inversion strategy in acoustic media, *in* ‘76th EAGE Conference and Exhibition 2014’. [1](#)
- Yao, G., Warner, M. and Silverton, A. (2014), Reflection fwi for both reflectivity and background velocity, *in* ‘76th EAGE Conference and Exhibition 2014’. [8](#), [52](#)
- Yilmaz, Ö. (2001), *Seismic data analysis: Processing, inversion, and interpretation of seismic data*, Society of exploration geophysicists. [54](#)

REFERENCES

- Yoon, K., Marfurt, K. J. and Starr, W. (2004), Challenges in reverse-time migration, *in* ‘SEG Technical Program Expanded Abstracts 2004’, Society of Exploration Geophysicists, pp. 1057–1060. [85](#)
- Youn, O. K. and Zhou, H.-w. (2001), ‘Depth imaging with multiples’, *Geophysics* **66**(1), 246–255. [85](#)
- Zhang, Y. and Sun, J. (2008), Practical issues of reverse time migration true-amplitude gathers, noise removal and harmonic-source encoding: 70th annual international conference and exhibition, eage, *in* ‘Extended Abstracts F’, Vol. 47. [85](#)
- Zhdanov, M. S. (2002), *Geophysical Inverse Theory and Regularization Problems*, Methods in Geochemistry and Geophysics Series, Elsevier Science. [11](#), [15](#)
- Zhdanov, M. S. (2015), *Inverse theory and applications in geophysics*, Vol. 36, Elsevier. [11](#), [15](#)