

THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Marchenko-Lippmann-Schwinger Inversion

Dominic Gerard Cummings



Submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2023

For Nellie,

forever in my heart

1930-2021

Abstract

Seismic wave reflections recorded at the Earth's surface provide a rich source of information about the structure of the subsurface. These reflections occur due to changes in the material properties of the Earth; in the acoustic approximation, these are the density of the Earth and the velocity of seismic waves travelling through it. Therefore, there is a physical relationship between the material properties of the Earth and the reflected seismic waves that we observe at the surface. This relationship is non-linear, due to the highly scattering nature of the Earth, and to our inability to accurately reproduce these scattered waves with the low resolution velocity models that are usually available to us. Typically, we linearize the scattering problem by assuming that the waves are singly-scattered, requiring multiple reflections to be removed from recorded data at great effort and with varying degrees of success. This assumption is called the Born approximation.

The equation that describes the relationship between the Earth's properties and the fully-scattering reflection data is called the Lippmann-Schwinger equation, and this equation is linear if the full scattering wavefield inside the Earth could be known. The development of Marchenko methods makes such wavefields possible to estimate using only the surface reflection data and an estimate of the direct wave from the surface to each point in the Earth. Substituting the results from a Marchenko method into the Lippmann-Schwinger equation results in a linear equation that includes all orders of scattering. The aim of this thesis is to determine whether higher orders of scattering improve the linear inverse problem from data to velocities, by comparing linearized inversion under the Born approximation to the inversion of the linear Lippmann-Schwinger equation.

This thesis begins by deriving the linear Lippmann-Schwinger and Born inverse problems, and reviewing the theoretical basis for Marchenko methods. By deriving the derivative of the full scattering Green's function with respect to the model parameters of the Earth, the gradient direction for a new type of least-squares full waveform inversion called Marchenko-Lippmann-Schwinger full waveform inversion is defined

that uses all orders of scattering.

By recreating the analytical 1D Born inversion of a boxcar perturbation by Beydoun and Tarantola (1988), it is shown that high frequency-sampling density is required to correctly estimate the amplitude of the velocity perturbation. More importantly, even when the scattered wavefield is defined to be singly-scattering and the velocity model perturbation can be found without matrix inversion, Born inversion cannot reproduce the true velocity structure exactly. When the results of analytical inversion are compared to inversions where the inverse matrices have been explicitly calculated, the analytical inversion is found to be superior. All three matrix inversion methods are found to be extremely ill-posed. With regularisation, it is possible to accurately determine the edges of the perturbation, but not the amplitude.

Moving from a boxcar perturbation with a homogeneous starting velocity to a many-layered 1D model and a smooth representation of this model as the starting point, it is found that the inversion solution is highly dependent on the starting model. By optimising an iterative inversion in both the model and data domains, it is found that optimising the velocity model misfit does not guarantee improvement in the resulting data misfit, and vice versa. Comparing unregularised inversion to inversions with Tikhonov damping or smoothing applied to the kernel matrix, it is found that strong Tikhonov damping results in the most accurate velocity models. From the consistent under-performance of Lippmann-Schwinger inversion when using Marchenko-derived Green's functions compared to inversions carried out with true Green's functions, it is concluded that the fallibility of Marchenko methods results in inferior inversion results.

Born and Lippmann-Schwinger inversion are tested on a 2D syncline model. Due to computational limitations, using all sources and receivers in the inversion required limiting the number of frequencies to 5. Without regularisation, the model update is uninterpretable due to the presence of strong oscillations across the model. With strong Tikhonov damping, the model updates obtained are poorly scaled, have low resolution, and low amplitude oscillatory noise remains.

By replacing the inversion of all sources simultaneously with single source inversions, it is possible to reinstate all frequencies within our limited computational

resources. These single source model updates can be stacked similarly to migration images to improve the overall model update. As predicted by the 1D analytical inversion, restoring the full frequency bandwidth eliminates the oscillatory noise from the inverse solution. With or without regularisation, Born and Lippmann-Schwinger inversion results are found to be nearly identical. When Marchenko-derived Green's functions are introduced, the inversion results are worse than either the Born inversion or the Lippmann-Schwinger inversion without Marchenko methods. On this basis, one concludes that the inclusion of higher order scattering does not improve the outcome of solving the linear inverse scattering problem using currently available methods. Nevertheless, some recent developments in the methods used to solve the Marchenko equation hold some promise for improving solutions in future.

Lay summary

"The only true voyage of discovery, the only fountain of Eternal Youth, would be not to visit strange lands but to possess other eyes, to behold the universe through the eyes of another, of a hundred others, to behold the hundred universes that each of them beholds, that each of them is. . . "
Marcel Proust, La Prisonnière (The Captive), 1923
(translated. C. K. Scott-Moncrieff, 1929)

The Earth is, perhaps, the most contradictory place in the known universe, for while we have the capability to observe far-distant stars and planets, to study the Universe at every scale from the smallest particles to structures that span incomprehensible distances, there is still a vast amount we do not know about our own planet. This is despite having never travelled further than the Moon, not even a toe dipped in the ocean of all existence. Much of this is because we are limited, by and large, to the surface of the Earth. Geologists study and infer the Earth's inner processes by collecting samples of materials brought up to the surface by such processes, but since we can only mine or drill to depths of kilometres compared to the thousands of kilometres from the surface of the Earth to its core, it is clear that we cannot directly sample the vast majority of the material that makes up the Earth, and that another complementary method is needed to fully understand our planet. It is, as Marcel Proust said above, *"the only true voyage of discovery. . . would be not to visit strange lands but to possess other eyes, to behold the universe through the eyes of another, of a hundred others"*, and indeed this is what we do as geophysicists when we record geophysical data at the surface of the Earth – we see the Earth through the response of the Earth at its surface to forces acting within it. Even more impressive is that from these data we can infer the physical properties of the Earth's subsurface through their effect on the data we record, *"to behold the hundred universes that each of them beholds"*. One method for doing this is by recording seismic waves.

Seismic waves are similar to sound waves we hear, except that they travel through the Earth instead of the air. They may have natural sources, such as earthquakes or volcanic eruptions, or man-made sources such as dynamite. As they propagate through the Earth, the wave is affected by the material properties of the Earth; it may speed up or slow down, change direction, or become stronger or weaker in its effect on its surroundings. Sometimes, what we are interested in is where the seismic waves originated from; for earthquakes, this location is called the focus, but more familiar is the location above the focus on the surface called the epicentre. This is a relatively simple example of a class of problem called an *inverse problem* that maps the data we record to a property of our *model* of the Earth. So if we record the traveltime of the waves from the source location, the property that transforms times to locations is the velocity of the waves in the Earth. This example requires that we already know the velocity of seismic waves in the Earth, but since seismic waves are impacted by the velocity of waves in the Earth, we can also solve an inverse problem for the velocity of the waves in the Earth called *seismic tomography*, also called *seismic imaging*.

The seismic imaging problem is a complex one because the velocity structure of the Earth is also complex, waves can change directions and *reflect* many times inside the Earth between where they originate and where we record them – such waves are called multiple reflections, or *multiples* for short. We typically do not have enough information for the transformation of multiples in our data to velocities in the Earth to be accurate, so we simplify the problem – we assume that waves only reflect once at most, called the single-scattering or *Born approximation*. This makes the problem solvable, but since it is not true in most cases we either accept that our answer will be less accurate, or we try to process our data so that it looks like singly-scattered data to improve the accuracy of our answer. Ideally we would choose neither option, and instead we would be able to transform the multiples to the correct velocities in the Earth.

In this thesis, I give an overview of the inverse scattering problem that maps recorded seismic waves to velocities in the Earth. By writing the forward scattering problem (the equation that describes the scattered waves observed for a given ve-

locity model) in terms of matrices, the inverse scattering problem can be solved by calculating the inverse matrices. I also describe a method of seismic imaging called *full waveform inversion* that maps recorded seismic waves to velocities in the Earth. Ordinarily, these methods require we assume the Born approximation holds; however, a group of new methods called *Marchenko* methods provide the additional information needed to correctly handle the multiples. I derive a new method of full waveform inversion that uses these Marchenko methods to correctly handle the multiple reflections.

Solving the matrix inverse problem for one dimensional and two dimensional examples, I show that including the multiple reflections does not result in better velocity models. The potential for Marchenko-Lippmann-Schwinger full waveform inversion and future advances in Marchenko methods leaves the door open for a new family of least-squares inverse methods using Marchenko-derived Green's functions and the adjoint-state method employed in methods such as full waveform inversion.

Declaration

I declare that this thesis has been composed solely by myself, and that it contains only my work except where otherwise specified, or where the work is explicitly indicated below to have formed part of a jointly-authored publication. This work has not been submitted for any other degree or professional qualification.

Dominic Cummings

November 2022

Acknowledgements

Christ, what a slog! After what feels like a lifetime in Edinburgh, I can honestly say that none of this would have been possible by myself. In fact, without the help and support of a great many people this thesis would never have seen the light of day. So whilst it seems unlikely that any human being other than myself will ever read this, this thank you is but a small token of my appreciation.

More than anyone else, I would like to thank my supervisor Prof. Andrew Curtis. A man of immense wisdom and insight, it is difficult to enunciate how much I have learned since our first meeting - in fact I have changed so much since I came to Edinburgh that I can hardly remember the person I was before. But more than that, without your kindness and patience I would not have found the strength to pick myself up after each and every fall, and perhaps this more than anything will be your lasting impact on me. It might be cliché, but to paraphrase Newton, "*If I have seen further it is by standing on the shoulders of Andrew Curtis and occasionally kicking him in the eye*".

I would also like to thank my second supervisors, Drs Giovanni Meles and Satyan Singh, whose doors were always open to me, and whose support helped me stay afloat while I struggled to learn how to become a scientist. Giovanni, without your wonderful idea I doubt that the work this thesis became would ever have come to fruition, and without your guidance it might never have become more than an idea. Satyan, you burst into our research group like a ray of Trinidadian sunshine and put us, your students, above all else when we were still struggling to find our way.

An immense thank you goes out to all of my office mates and colleagues/co-conspirators in the Edinburgh Imaging Project: Carlos, Claire, Steph, Xin, Erica, Hugo, Lou & Xuebin; who made coming to work every day worth doing, but especially to Angus, who has put up with me longer than anyone here and never once complained. Not even a little bit.

To my fellow sufferers Hannah, Amelia, Berit, Auntie G, Phil, Rory, Ellen, Rachel, Lyss, Eva, and so many others; thank you, and thank God it's over!

To Sophie Ramette in the PGR office, a massive thank you for everything you've done for me over the years. Words cannot describe how valuable your help has been.

To Rev. Steven, thank you for being there for me, emotionally and financially. Without your support I don't think I would have made it this far.

To Jess, thank you for being my rock and my best friend, for being there through all the misery to pick me up without ever thinking about yourself and pushing me to be better than I am. Barely a day goes by without a kind word and a kind thought, and it feels strange to think there was ever a time it wasn't this way.

To the sponsors of the Edinburgh Interferometry Project and Edinburgh Imaging Projects and their representatives James Rickett, Reda Baina, Paul Williamson, Ed Kragh and so many others; thank you for supporting our research group and helping make the work we do as great as it can be.

To my examiners, Professor Anton Ziolkowski and Dr Lluis Guasch, I would like to say thank you for an insightful critique and discussion of my work, as well as your kind words afterwards. I don't know where I will end up in life, but perhaps my career as a scientist is not as finished as I had believed it to be.

Finally, I would like to thank my family without whom I would be nothing, both literally and figuratively. Mum, Dad, Nellie, Gerard, Damien, and Noel; thank you for being there for me come rain or shine, when all hope seemed lost and when it wasn't. Much love, always.

Dom

Contents

Abstract v

Lay Summary ix

Declaration xiii

Acknowledgements xv

List of Figures xxi

List of Tables xxix

Glossary xxxi

1 Introduction 1
- 1.1 The Seismic Inverse Problem . . . 2
- 1.2 Seismic Imaging-based Inversion . . . 5
- 1.3 Green's function retrieval . . . 8
- 1.4 Thesis plan . . . 10

2 Marchenko-Lippmann-Schwinger Methods 13
- 2.1 The Lippmann-Schwinger Equation . . . 14
  - 2.1.1 The Born Approximation . . . 21
- 2.2 The Inverse Scattering Problem . . . 25
  - 2.2.1 Full-Waveform Inversion . . . 31
- 2.3 Marchenko Methods . . . 37
- 2.4 Marchenko-Lippmann-Schwinger Inversion . . . 41
  - 2.4.1 Marchenko-Lippmann-Schwinger Least-Squares Inversion . . . 42
  - 2.4.2 Adjoint-State Methods of Marchenko-Lippmann-Schwinger Inversion . . . 43

3 Experimental Design & Computational Requirements for Implementation of Marchenko-Lippmann-Schwinger Methods 51
- 3.1 Experimental Design . . . 51
  - 3.1.1 P-wave velocity and density models . . . 51
  - 3.1.2 Wavelet design . . . 57
  - 3.1.3 Wavelet design for Marchenko methods . . . 58
  - 3.1.4 Source and receiver arrays . . . 60

3.2 Seismic modelling and Marchenko routines 62
3.2.1 Choice of programs 62
3.2.2 Workflow for obtaining Marchenko Green's functions 64
3.3 Summary 68

4 Benchmarking Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger-Inversions in One Dimension 69
4.1 Benchmarking of one-dimensional inversion of finite-difference Green's functions against an analytic result 70
4.2 Comparison of Born inversions using analytical Green's function solutions to Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inverse solutions using finite-difference wavefields 81
4.3 Discussion 86

5 Linear Inversion of a Many-Layered Velocity Model in One Dimension 97
5.1 Inversion optimised in the model domain 100
5.2 Inversion optimised in the data domain 109
5.3 Discussion 116

6 Linear inversion of the Lippmann-Schwinger equation in two dimensions 131
6.1 Verifying the Lippmann-Schwinger Green's function reconstruction in two dimensions 134
6.2 Born inversion versus Lippmann-Schwinger inversion in two dimensions 139
6.3 Discussion 149

7 Reconsideration of the Lippmann-Schwinger inverse problem in two dimensions 157
7.1 Born inversion versus Lippmann-Schwinger inversion in two dimensions 161
7.2 Marchenko-Lippmann-Schwinger inversion in two dimensions 168
7.3 Discussion 173
7.4 Summary 178

8 Discussion 181
8.1 A logical order of work 181
8.2 The dominance of single-scattering 183
8.3 Relation to Diekmann et al. (2021) 184
8.4 The Marchenko method 186
8.5 Future research 188

9 Conclusion 193

Bibliography 199

A Addendum to Chapter 4 207
A.1 Source function deconvolution 207
A.1.1 Theory of deconvolution 208
A.1.2 Results 209

A.2 Inversion in the time domain 211

B Refractions, Head Waves & Diving Waves 223

C Analytic Green's Function Solutions to the Helmholtz Equation in Two- & Three- Dimensions 227

List of Figures

2.1 The perturbation theory model description, which says that the true medium (left) can be described by a known estimate of the true medium (centre) plus a small perturbation to that estimate (right). . 16
2.2 The Lippmann-Schwinger description of the Green's function between a source (red and orange explosion) and a receiver (blue triangle) for the medium shown in Figure 2.1. 20
2.3 The Lippmann-Schwinger description of the scattered wavefield as the sum of scattering orders, both for the formulation in equation 2.17 (a to d) and the formulation in equation 2.19 (e to h). Only the first (b & f), second (c & g), and third (d & h) orders of scattering are shown. 23
2.4 The virtual source formulation of the scattered wavefield. The real source excites a wavefield (a), which radiates out and where there are contrasts in the medium (b) partially backscatters the incoming wavefield as if a source was located at the location of the medium contrast (c) radiating a secondary wavefield back towards the surface. 35

3.1 True Earth P-wave velocity model. Every 7th source (red star) and receiver (cyan triangle) of the surface arrays are shown. 52
3.2 Smoothed P-wave starting velocity model. 53
3.3 Comparison of 1st derivative of a Gaussian wavelet and the Ricker wavelet at peak frequency = 15Hz. Top: Wavelets in the time domain; and bottom: their (frequency) amplitude spectra. 58
3.4 Comparison of the 1st derivative of a Gaussian wavelet and the flat spectrum wavelets. Top: The Gaussian 1st derivative and normalized amplitude flat spectrum wavelets in the time domain. Bottom: the normalized amplitude spectra of the wavelets compared against the amplitude spectrum of the autoconvolution of the flat spectrum wavelet. 60
3.5 A pictorial representation of how data is transformed during the workflow of producing Marchenko Green's functions and subsequently wavefields. 67

4.1 The 1D slab model. Adapted from Beydoun and Tarantola (1988). . 71
4.2 The Gabor wavelet with $f_M = 12.5$Hz, $\gamma = 3$, $t_0 = 0.1$s, and $\nu = 0$. . 73
4.3 True and initial models for 1D Born inversion 74
4.4 1D Born inversion of Beydoun and Tarantola (1988) slab model. . . . 75
4.5 Analytical Born inversion of the 1D slab model for varying lengths of time series, increasing in powers of two. 77

4.6 Data following 1D Born inversion of Beydoun and Tarantola (1988) slab model. . . . 79
4.7 1D Born inversion of a slab model with a positive velocity perturbation. 80
4.8 Comparison of analytic inversion to least-squares Born inversion (top), Lippmann-Schwinger inversion (middle), and Marchenko-Lippmann-Schwinger inversion (bottom) for the slab model of Beydoun and Tarantola (1988). . . . 83
4.9 Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversion models after 50 iterations of each scheme. . . . 84
4.10 Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversion models after a single iteration of each scheme with Tikhonov damping. . . . 84
4.11 Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversion models after 50 iterations of each scheme with Tikhonov damping optimised at every iteration. . . . 85
4.12 Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversion models after a single iteration of each scheme with smoothing applied. . . . 86
4.13 Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversion models after 50 iterations of each scheme with applied smoothing optimised at every iteration. . . . 87
4.14 Close-up plot of the amplitude spectrum of the 12.5Hz Gabor wavelet in the range 0 - 1.5Hz as a function of time series length in samples. 90
4.15 Comparison of the wavefield in the true model and its estimate using a Marchenko method before the first iteration of inversion. Both wavefields have been amplitude normalised for comparison. Depth sections (c)-(e) are marked by dashed black lines in panels (a) & (b). 95
4.16 Comparison of the 12.5Hz Gabor wavelet used by Beydoun and Tarantola (1988) to its time derivative, in both time and frequency domains. 96

5.1 1D slab model of Beydoun and Tarantola (1988) plotted against the absolute percentage error of the initial model. . . . 99
5.2 1D many-layered model plotted against the absolute percentage error of the initial model. . . . 100
5.3 Velocity models after a single iteration of inversion without regularisation. . . . 101
5.4 Velocity models after 50 iterations of inversion without regularisation. 102
5.5 Model misfit trend over 50 iterations of inversion. . . . 102
5.6 Data misfit trend over 50 iterations of inversion. . . . 103
5.7 Velocity models after a single iteration of inversion regularised with Tikhonov damping. . . . 104
5.8 Velocity models after 50 iterations of inversion regularised with Tikhonov damping. . . . 104
5.9 Model misfit trend over 50 iterations of inversion regularised with Tikhonov damping. . . . 105

5.10 Data misfit trend over 50 iterations of inversion regularised with Tikhonov damping. . . . 106
5.11 Velocity models after a single iteration of inversion with smoothing regularisation. . . . 107
5.12 Velocity models after 50 iterations of inversion with smoothing regularisation. . . . 107
5.14 Data misfit trend over 50 iterations of inversion with smoothing inversion. . . . 108
5.13 Model misfit trend over 50 iterations of inversion with smoothing regularisation. . . . 108
5.15 Velocity models after a single iteration of inversion optimised with the least-squares data misfit. . . . 110
5.16 Velocity models after 50 iterations of inversion optimised with the least-squares data misfit. . . . 110
5.17 Data misfit trend over 50 iterations of inversion optimised with the least-squares data misfit. . . . 111
5.18 Model misfit trend over 50 iterations of inversion optimised with the least-squares data misfit. . . . 112
5.19 Velocity models after a single iteration of inversion, regularised with Tikhonov damping and optimised with the least-squares data misfit. . . 113
5.20 Velocity models after 50 iterations of inversion, regularised with Tikhonov damping and optimised with the least-squares data misfit. . . . 114
5.21 Data misfit trend over 50 iterations of inversion regularised with Tikhonov damping and optimised by the least-squares data misfit. . . . 114
5.22 Model misfit trend over 50 iterations of inversion regularised with Tikhonov damping and optimised by the least-squares data misfit. . . 115
5.23 Velocity models after a single iteration of inversion with smoothing regularisation and optimised with the least-squares data misfit. . . . 116
5.24 Velocity models after 50 iterations of inversion with smoothing regularisation and optimised with the least-squares data misfit. . . . 117
5.25 Data misfit trend over 50 iterations of inversion with smoothing regularisation and optimised by the least-squares data misfit. . . . 118
5.26 Model misfit trend over 50 iterations of inversion with smoothing regularisation and optimised by the least-squares data misfit. . . . 118
5.27 Comparison of data modelled in the true model to data modelled after 50 iterations of inversion (top) without regularisation, (middle) with Tikhonov damping, and (bottom) with smoothing. . . . 119
5.28 Comparison of the amplitude spectrum of the data from the true model to an amplitude spectrum derived from the wavenumber spectrum of the velocity model obtained after 1 iteration of Marchenko-Lippmann-Schwinger inversion with smoothing. . . . 121
5.29 Comparison of the wavefield in the true model and its estimate using a Marchenko method before the first iteration of inversion. Both wavefields have been amplitude normalised for comparison. Depth sections (c)-(e) are marked by dashed black lines in panels (a) & (b). 124

5.30 Variation of model and data misfits for inversions optimised in the model and data domains. 127

6.1 True model for the 2D inverse scattering problem. The true model for the 1D many-layered model in chapter 5 is taken from the profile shown by the dashed line. 133
6.2 Initial model for the 2D inverse scattering problem. The initial model for the 1D many-layered model in chapter 5 is taken from the profile shown by the dashed line. 133
6.3 The true model for testing the accuracy of Lippmann-Schwinger data reconstruction in 2D, formed by extending the magenta depth profile in figure 6.1 in the x-direction. The single source and receiver location are indicated by the star and inverted triangle respectively. 135
6.4 The background model for testing the accuracy of Lippmann-Schwinger data reconstruction in 2D. The single source and receiver location are indicated by the star and inverted triangle respectively. 135
6.5 Comparison of the true model data modelled with finite-differences to the Lippmann-Schwinger reconstruction of the same data. Arrivals within the black and red dashed lines are artefacts from the Lippmann-Schwinger process. 136
6.6 Comparison of the true model data modelled with finite-differences to the Lippmann-Schwinger reconstruction of the same data when the bottom layer velocity in the background model is exactly that in the true model. The black and red windows are the same as those in figure 6.5. 137
6.7 The kernel of the Lippmann-Schwinger modelling equation plotted in the time domain as a common receiver gather, taking only the contributions from model pixels on the left, bottom, and right edges of the model. Trace number increases from 0 in the top left corner of the model to 660 in the top right corner. The black and red dashed lines correspond to the same lines in figures 6.5 and 6.6. 138
6.8 Schematic of the spatial tapering applied to the model. The original model area, enclosed by the dashed line, remains fixed. The model area is then extended homogeneously using the edge values of the model to the solid white line, with the taper running from 0 at the solid line to 1 at the dashed line. The taper length can be increased by increasing the amount the model is extended up to any taper length required (the effect on the solid boundary is indicated by the double-headed arrows). 139
6.9 Comparison of the true model data modelled with finite-differences to the Lippmann-Schwinger reconstruction of the same data for various lengths of taper applied at the model edges. 140
6.10 Comparison of Born inversion velocity models to the true and initial models. 142
6.11 Comparison of Lippmann-Schwinger inversion velocity models to the true and initial models. 143

6.12 Comparison of slowness-squared model updates from Born and Lippmann-Schwinger inversions with the true model perturbation. Panels (b) - (f) are shown on a common scale, panel (a) is shown at true amplitude. The dashed lines in panel (c) indicate the location of interfaces in the true velocity model, whilst purple arrows in panels (e) & (f) match the orange arrows in figure 6.11. 145
6.13 Comparison of velocity model residuals after Born and Lippmann-Schwinger inversions with the initial velocity model residual. 147
6.14 Comparison of data residual gathers before and after inversion by the Born and Lippmann-Schwinger methods. Panel (d) is shown at true scale, all other panels are shown on a common scale. 148
6.15 Model and data misfits for each inversion method. 149
6.16 Raypaths shot through the true velocity model from source position 21 (x=1300m), demonstrating the lack of illumination in deeper layers from wide offset sources. 152
6.17 A demonstration of the Gibbs phenomenon in the Fourier representation of a square wave (reproduced from Gubbins (2004)). 155

7.1 25Hz Ricker wavelet in both time and frequency domains. 160
7.2 A pictorial demonstration of the source subsampling scheme used. Coloured explosions represent sources used for inversion and grey explosions sources omitted from the inversion. 161
7.3 Comparison of Born inversion velocity models to the true and initial models. 163
7.4 Comparison of Lippmann-Schwinger inversion velocity models to the true and initial models. 164
7.5 Comparison of slowness-squared model updates from Born and Lippmann-Schwinger inversions with the true model perturbation. Panels (b) - (f) are shown on a common scale, panel (a) is shown at true amplitude. 166
7.6 Comparison of data residual gathers before and after inversion by Born and Lippmann-Schwinger methods. 167
7.7 Flat spectrum wavelet for use in a Marchenko method in both time and frequency domains. 169
7.8 Comparison of Marchenko-derived Green's function to the true Earth Green's function between the same two points. 171
7.9 Comparison of Marchenko-Lippmann-Schwinger inversion velocity models to the true and initial models. 172
7.10 Comparison of slowness-squared model updates from Marchenko-Lippmann-Schwinger inversions with the true model perturbation. Panels (b), (c), and (e) are shown on a common scale, panels (a) and (d) are shown at true amplitude. 174
7.11 Comparison of velocity model and data misfits before and after inversion. 175
7.12 Comparison of the impact of applying different bounds to a Brent search for the optimal scaling for model updates from Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversions. 177

7.13 Comparison of slowness-squared model updates from Marchenko-Lippmann-Schwinger inversions with the true model perturbation. Panels (b), (c), and (e) are shown on a common scale, panels (a) and (d) are shown at true amplitude. 179

A.1 Comparison of data traces modelled using 1001 time samples when (top) flat spectrum Green's functions are used, (second top) flat spectrum data is convolved with the Gabor wavelet, (second bottom) the convolved wavelet is removed by deconvolution, and (bottom) the convolved wavelet is removed by deconvolution with the Gabor wavelet after zero-phasing. 210

A.2 Comparison of data traces modelled using 64001 time samples, shown truncated after 1001 samples. Subfigures are arranged as per figure A.1. 211

A.3 1D Born inversion of a boxcar perturbation in the time domain using the analytic method of Beydoun and Tarantola (1988) with flat spectrum Green's functions. 213

A.4 1D Born inversion of a boxcar perturbation in the time domain using the analytic method of Beydoun and Tarantola (1988) with data convolved with the Gabor wavelet. 214

A.5 1D Born inversion of a boxcar perturbation in the time domain by replacing the integration limits of the analytic method of Beydoun and Tarantola (1988) with a sum over all time samples. 214

A.6 1D Born inversion of a boxcar perturbation in the time domain using the analytic method of Beydoun and Tarantola (1988) and data modelled with 64001 time samples. 216

A.7 Comparison of the inversion amplitude with the upper integration limit in equation A.6 evaluated at various depths above, within, and below the velocity perturbation using flat spectrum Green's functions. . . . 218

A.8 Comparison of the inversion amplitude with the upper integration limit in equation A.6 evaluated at various depths above, within, and below the velocity perturbation using data convolved with the Gabor wavelet. 220

A.9 1D Born inversion of a boxcar perturbation in the time domain using only the 0Hz component of the data with the analytic method of Beydoun and Tarantola (1988). 221

B.1 A schematic of refractions in a layered medium. For incident angles below the critical angle $\theta_{critical}$ (black arrows), the propagation direction of the wave changes at the interfaces between velocity layers. For incidence angles equal to or greater than the critical angle (red arrows), the wave is critically refracted at the velocity interface and continues to propagate perpendicular to the interface in the second layer. Per Snell's law, the critically refracted wavefront acts as a source of waves propagating upwards at the critical angle (dashed red arrows). The recorded refraction is called a head wave (upwards red arrow). 224

B.2 A schematic of diving wave formation. In a smooth model where velocity increases linearly with depth, the wavefield is continuously refracted since the velocity changes continuously. In a finitely discretized model, it thus refracts after each unit of length propagated (a). In the limit of infinitesimal unit size, the wavefront is refracted along a smoothly curving trajectory until it reaches the criticial incidence angle (b), at which point it propagates horizontally emitting Huygen's wavelets at the critical angle upwards. Since the medium smoothly increases with depth, the upward trajectory of the wavefront curves smoothly towards the vertical (c), this is the diving wave. We can image using diving waves by backpropagating the wavefield from the receiver through the same model (d). 225

List of Tables

3.1 Wavelet settings for the flat spectrum wavelet in figure 3.4 59

Glossary

ADCIG	Angle-domain common image gather
AWI	Adaptive waveform inversion
CIG	Common image gather
FWI	Full-waveform inversion
GF	Green's function
LHS	Left-hand side
LSI	Lippmann-Schwinger inversion
LSM	Least-squares migration
MLS	Marchenko-Lippmann-Schwinger
MLS-FWI	Marchenko-Lippmann-Schwinger full-waveform inversion
MLSI	Marchenko-Lippmann-Schwinger inversion
MVA	Migration velocity analysis
NMO	Normal moveout analysis
RHS	Right-hand side
RSF	Regularly sampled format
SEG	Society of Exploration Geophysicists
SRME	Surface Related Multiple Elimination
SU	Seismic Unix

Chapter 1

Introduction

Seismic wavefields, both from naturally occurring sources such as earthquakes or artificial sources such as dynamite, illuminate the Earth at all times. These waves are a rich source of information about the composition of the subsurface and the processes working within the Earth, especially deep below the surface where human beings can never travel and examine the Earth directly. As the wavefield propagates outwards from the source, its characteristics are changed by the properties of the materials through which it travels. These changes are encoded within recordings of wavefields we make at the surface of the Earth called seismograms. Mathematically speaking, we can describe this relationship between the Earth and the seismic wavefield by saying that the data we obtain when we record the seismic wavefield is a function of the material properties of the Earth called the model. The model does not need to describe every property of the Earth, or even the Earth in its entirety; rather, the model describes the properties of the medium in the domain being studied that can alter the data observed. For example, if we wanted to know the fastest traveltime of a wave between two nearby locations, the model parameters would be the slownesses (the reciprocal of velocities) at different locations between the two locations. Even though the medium has other properties, e.g. electrical conductivity, these properties are irrelevant to the problem at hand; similarly, even if our two locations are on or inside the Earth, we do not need to consider a wave travelling around the entire Earth to know that it is not the solution we are looking for.

The relationship between the data and the model can be written as an equation of the form

$$\boldsymbol{d} = F(\boldsymbol{m}) \tag{1.1}$$

where $\boldsymbol{d}$ is the vector of the recorded seismic wavefield, $\boldsymbol{m}$ is the model vector encapsulating the properties of the Earth, and F is the function mapping between the two. The function F and the model parameters $\boldsymbol{m}$ change depending on the physics of the problem; for example, if we treat the seismic wavefield as an elastic wave, then the model parameters would be the density of the Earth and the elements of a stiffness matrix, and the function F would encode the elastic wave equation. If the phenomenon we are interested in can be adequately described by the acoustic approximation, then the model parameters could be reduced to the density of the Earth and the velocity of compressional P-waves through the Earth, while the function F would encode the acoustic wave equation. As written, equation 1.1 says that if we know the medium properties of the Earth, then we can predict the wavefield that we would record at the surface. This is called the forward problem.

We are generally not interested in solving the forward problem since we can readily record seismic wavefields on the Earth's surface, but generally do not know the medium properties of the Earth. The function that maps the recorded data to the unknown model is the inverse F^{-1} of the function F, and the problem is called the inverse problem.

$$\boldsymbol{m} = F^{-1}(\boldsymbol{d}) \tag{1.2}$$

The focus of this thesis is the solution of the inverse problem (equation 1.2) for the acoustic medium parameters of the Earth.

1.1 The Seismic Inverse Problem

There are many different inverse problems within the field of seismology, and even for the same conceptual problem i.e. how to find x given knowledge of y, there may be several functions $F^{-1}(d)$ that map between the two spaces; and for each of these functions there may be several algorithms for determining the elements of the matrix

that describe that function, or several approximations to that function, each of which may have unique algorithms designed to determine the elements of the approximate matrix. The effectiveness of different functions describing the same conceptual inverse problem is not equal, and a great deal of effort goes into finding appropriate functions that can be easily solved and give an accurate solution when compared to a known result. One of the most commonly observed properties of seismograms is that the recorded wavefield appears to be highly scattered, and since scattering is known to occur where changes in medium properties occur, the inverse problem for the Earth's medium properties can be posed using scattering theory. Such problems are called inverse scattering problems.

Within the subset of inverse scattering problems, there are still many scattering equations that can be used. Early inverse scattering approaches were based on quantum scattering theory and used the time-independent Schrödinger equation as the forward modelling equation; by taking the potential term in the Schrödinger equation equal to zero, the Schrödinger equation reduces to the Helmholtz equation. Adapting the method of Jost and Kohn (1952), Moses (1956) calculated the scattering potential from reflection coefficients for the Schrödinger equation in one- and three-dimensions, forming the basis for a series solution to the Helmholtz equation with variable scattering potential provided the scattering potential varies smoothly (Prosser, 1969, 1976).

Cohen and Bleistein (1979) developed a data-driven method for inverting for discontinuous 2D scattering potentials based on the constant-density acoustic wave equation starting from a homogeneous velocity model. Beydoun and Tarantola (1988) developed a 1D inversion method based on exact representations of the wavefield and scattering potential for a slab inclusion in a medium with homogeneous background velocity. A similar analysis was performed by Rajan and Frisk (1989) for a number of more complex 1D velocity structures. The commonality between all of these examples is that they do not include the full scattering in the wavefield in the inversion, and instead require that the scattering potential be sufficiently weak so that the wavefield is singly-scattering. This is the so-called Born approximation (Born, 1926), the integral representation of which can be written using Green's functions, the response

observed due to an impulsive source i.e. a Dirac delta function. Due to the sifting property of the delta function, that the convolution of a function with the delta function returns the prior function evaluated at the position of the delta function, the Green's function solution can give the observed response due to any source function by convolution of the Green's function with said source function; for this reason, the Green's function solution is sometimes called the fundamental solution. The Green's function representation of the Born approximation is written

$$G_{Born}\left(g,s;\omega\right)=\omega^{2}\int V(x)G_{0}\left(g,x;\omega\right)G_{0}\left(x,s;\omega\right)dx \tag{1.3}$$

Here, G_{Born} is the singly-scattering Green's function between the source at s and receiver at g with frequency ω i.e. the singly-scattered component of wavefield recorded at the surface, G_0 is the Green's function in the background medium, and the scattering potential denotes the difference between the background medium and the true medium and is given by $V(x)=\frac{1}{c^{2}(x)}-\frac{1}{c_{0}^{2}(x)}$ where $c(x)$ is the true velocity structure of the Earth and $c_0(x)$ is the background medium velocity structure. A method of inversion applied directly to the integral representation of the Born equation without any exact representations for the wavefield or scattering potential was found by Freter (1992) for constant velocity background media. The Born integral equation is a linearization of the canonical equation of acoustic scattering, the Lippmann-Schwinger integral equation:

$$G_{S}\left(g,s;\omega\right)=\omega^{2}\int V(x)G_{0}\left(g,x;\omega\right)G\left(x,s;\omega\right)dx \tag{1.4}$$

The left-hand side of the Lippmann-Schwinger equation, G_S is the scattering Green's function containing all orders of scattering, not only the first order – the additional orders of scattering are introduced by the Green's function G in the true medium on the right hand side of the equation. Unfortunately, without receivers placed inside the subsurface of the Earth (even if borehole seismic sensors are available, the limited receiver coverage does not span the entire model area), we cannot measure the Green's functions in the true medium G. This makes the Lippmann-Schwinger equation nonlinear with respect to the model parameters (since both V and G are dependent on

the unknown model parameters). Thus in the absence of suitable methods to obtain the Green's functions G, the linearization from the Lippmann-Schwinger equation to the Born integral equation is necessary, and simply replaces G by G_0, which is accessible since it can be modelled computationally in the background medium.

1.2 Seismic Imaging-based Inversion

If we accept that the true relationship between the seismic wavefield and the medium parameters of the Earth is non-linear, and that we cannot access the Green's functions in the true Earth G that make the Lippmann-Schwinger equation linear, then we have three options for solving the inverse scattering problem.

The first option is to force the Lippmann-Schwinger equation to be linear by assuming scattering is weak enough that the single scattering approximation holds. The Lippmann-Schwinger equation (equation 1.4) then becomes the Born integral equation (equation 1.3) and the methods referred to in the previous section become valid. We know that in general this approximation is invalid, because the Earth is a highly scattering medium; consequently, to use such methods on real seismic wavefields, we must process the data to remove higher order scattering events called "multiples" whilst preserving the single scattering events. Unlike random noise sources, multiples present coherently in the data and cannot be removed by summing or "stacking" over data time-series traces. For exploration-scale seismic experiments, a number of methods have been devised for removing different types of multiple scattering: ocean-bottom multiple scattering due to the free surface can be eliminated by methods such as Surface Related Multiple Elimination (SRME) (Verschuur et al., 1992; Berkhout and Verschuur, 1997) that predict the arrival time of multiples and adaptively subtract them from the data; whilst intra-layer multiple scattering can be removed by filtering in alternate domains such as the f-k domain (Embree et al., 1963; Fail and Grau, 1963) or τ-p domain (Hampson, 1986) if the moveout velocity contrast is sufficiently great, or by predictive deconvolution in the τ-p domain (Alam and Austin, 1981; Treitel et al., 1982). Unfortunately, such processing is never perfect and artefacts from the processing will remain in the data, and unless the single- and multiple-scattering

events are well spaced in the relevant domains, a compromise between removing useful data alongside the multiples or preserving the data and allowing multiples to remain must be made. An in-depth discussion of seismic processing methods is given by Yilmaz (2001).

The second and, perhaps, most obvious choice is to attempt to solve a related problem using a different modelling equation. We have already discussed methods based on quantum scattering using the Schrödinger equation, however other quantum scattering equations exist – including the Lippmann-Schwinger equation (Lippmann and Schwinger, 1950). The method of Moses (1956) is equivalent to the Gelfand-Levitan equation (Gelfand and Levitan, 1951), whilst Snieder (1990a,b) uses the Marchenko equation (Marchenko, 1955; Agranovich and Marchenko, 1964) to invert the 1D plasma equation with a Born series solution.

The third option is to transform the non-linear problem into a linear problem that can be iterated towards the solution of the non-linear problem; this is often called seismic tomography. Schuster (2017) divides seismic inversion or tomography (from the Greek *tomos*, meaning "slice") into four categories: traveltime tomography, phase inversion, full-waveform inversion, and migration velocity analysis. Traveltime tomography uses an initial estimate of the slowness of the Earth to calculate the traveltime of a ray between two points, from which one can infer the average velocity (or its reciprocal, slowness) along the path of the ray. By using many rays travelling on different raypaths, one can infer the differences in slowness between the initial slowness model of the Earth and the true Earth. This procedure of forward modelling rays and inverting for slowness can be iterated many times to converge to an optimal solution. The first instance of traveltime tomography being used in geophysics is generally accepted to be by Aki and Lee (1976), who used P-wave arrival times from local earthquakes to invert for the 3D crustal structure beneath California by tracing linear ray paths and applying a damped least-squares inversion. P-wave traveltime tomography was extended to the global scale by Dziewonski et al. (1977) to image the velocity structure of the mantle. Bishop et al. (1985) were the first to use the traveltimes of seismic reflections to determine the velocities and depths of a laterally homogeneous velocity structure. For a detailed overview, I refer the reader to reviews

of the field by Rawlinson and Sambridge (2003) or Rawlinson et al. (2010) for example.

Phase inversion utilizes the fact that in the frequency domain, the seismic wavefield can be separated into amplitude and phase components. Similar to traveltime tomography, the difference between the phase of recorded and modelled wavefield traces can be used instead of the difference in traveltimes to update the current velocity model to be closer to the velocity structure of the Earth. Examples include Woodward (1992), Sun and Schuster (1993), and Shin and Min (2006).

Full-waveform inversion (FWI) is an optimisation problem that seeks to find the (velocity) model that minimises the difference between recorded and modelled waveforms, it does this by casting the non-linear inverse problem as an iterative linear problem typically solved by the method of least-squares. Although the name "full-waveform inversion" may be considered a misnomer, since it is not necessary or useful to always use the full-waveform in the inversion (Fichtner, 2010), by using the full-waveform in place of traveltimes, FWI is capable of using both phase and amplitude information in the velocity inversion to produce high-resolution models (Alkhalifah, 2016). The earliest works widely considered to be full-waveform inversions were by Lailly (1983) and Tarantola (1984) who demonstrated that the Fréchet derivative in the least-squares formulation of the inverse problem could be obtained using an adjoint-state method, that is to say, a suitable iterative scheme to update the model could be obtained by a migration operation. In fact, the model update in the first iteration IS the migration image[1]. Efficient FWI in the frequency domain was formulated by Pratt et al. (1998) and Pratt (1999), and has developed into a broad research field with numerous variations such as Adaptive Waveform Inversion (AWI) (Warner and Guasch, 2016; Guasch et al., 2019) and multiple applied contexts, having being applied in medical imaging to the imaging of breast tissue (Pratt et al., 2007; Agudo et al., 2018) and brain tissue (Guasch et al., 2020). For an extensive historical and technical overview of FWI, I refer the reader to Virieux and Operto (2009).

[1]Strictly speaking, this is only true when considering reverse-time migration using a smooth velocity model that does not cause reflections. Whilst such a migration operation is commonplace in modern seismic data processing, it is not necessarily the only combination of migration algorithm and velocity model that will be used.

Migration velocity analysis (MVA) is different from the previous inversion methods in that it does not seek to find the model that best fits the data but rather the model that optimises the image in the image domain, typically assessed using common image gathers (CIGs). Much like normal moveout (NMO) analysis in which the correct NMO velocity is found when the primary reflection response has been flattened in time across space, so too in migration velocity analysis the aim is to flatten the primary reflection response across the CIGs. By determining the relationship between the image perturbation (i.e. the moveout of the image across the gather) to the velocity model perturbation, one can update the velocity model from the CIG. This process can be iterated by imaging with the new velocity model until a satisfactory result has been achieved. Originally formulated for offset domain common image gathers (e.g. Liu and Bleistein, 1995; Liu, 1997), MVA has also been formulated for angle domain common image gathers (Biondi and Sava, 1999; Mosher et al., 2001; Liu et al., 2001). Classically, MVA uses ray-based methods such as Kirchoff migration, making it computationally efficient compared to wave equation based inversion methods such as FWI, although wave equation methods of MVA do exist (Sava and Biondi, 2004; Symes, 2008).

1.3 Green's function retrieval

So far, we have assumed that the Green's functions G containing all orders of scattering cannot be known inside the Earth without a receiver or a source also inside the Earth, and because of this we justify ourselves of the appropriateness of the Born approximation in seismic inversion – because there is no other choice. However, recent methodological advances mean that this is no longer true: it is indeed possible to estimate the Green's functions G from the surface of the Earth to a location inside its subsurface.

Seismic interferometry allows for the homogeneous Green's function (combining both the causal and anti-causal Green's functions) to be retrieved between two receivers (inter-receiver interferometry) as if there were a "virtual" (i.e. imagined) source at the location of one of the receivers (e.g. Lobkis and Weaver, 2001), between

two sources (inter-source interferometry) as if there were a virtual receiver at the location of one of the sources (Curtis et al., 2009), or between a real source position and a real receiver position (source-receiver interferometry) as if there were a virtual receiver located at the position of the real source and a virtual source located at the position of the real receiver (Curtis and Halliday, 2010). As powerful as these methods are, there are two requirements for seismic interferometry that restrict its application: the first is that inter-source/receiver interferometry requires a closed boundary of receivers/sources surrounding the sources/receivers between which the Green's function is to be retrieved, in the case of source-receiver interferometry enclosing boundaries of both sources and receivers are required. Snieder (2004) showed using stationary phase analysis that only sources/receivers on the boundary approximately in-line with receiver/source pair contribute constructively in the Green's function retrieval, but even with this easing of the imposed acquisition geometry, two-sided access to be medium being studied is still required. The second requirement is that to retrieve the Green's function to/from a virtual receiver/source, a real source/receiver must already be located there. So seismic interferometry cannot be used to recreate data from a virtual source or receiver at an inaccessible location in the subsurface unless a seismic source (either active or passive) already fired there and was recorded.

Around the same time, Rose (2001, 2002) developed the so-called "autofocusing" algorithm for using the 1D single-sided reflection response due to an unknown scattering potential to focus the wavefield at a location inside the medium, that is, to collapse the wavefield to a delta function in space and time at that location. Furthermore, it was shown that the autofocusing scheme gave the same result as solving the Marchenko equation (Marchenko, 1955). Broggini et al. (2012) later demonstrated that this focusing solution to the Marchenko equation could be used to retrieve the 1D Green's function from a virtual source located at the subsurface focusing location. This result was extended to three-dimensions by Wapenaar et al. (2013) for acoustic waves, and to elastic waves by da Costa Filho et al. (2014) and Wapenaar (2014). The method of Wapenaar et al. (2013), and later applications and variations of the method, are generally known as "Marchenko methods". Extensions of the Marchenko method exist to include free surface multiples (Singh et al., 2017), for plane wave

data (Meles et al., 2018, 2020), for internal multiple prediction and removal (Meles et al., 2015; da Costa Filho et al., 2017), and the reconstruction of reflection data containing only primary reflections (Meles et al., 2016). Marchenko methods have also been successfully applied to field data (Ravasi et al., 2016; Staring et al., 2018; Staring and Wapenaar, 2020) and under controlled laboratory conditions (da Costa Filho et al., 2018). Crucially, Marchenko methods inherit their single-sided acquisition geometry from Rose's autofocusing scheme and have no requirement for a real source or receiver at the virtual receiver/source location, allowing their application in a range of scenarios.

1.4 Thesis plan

We can summarise the preceding information by the following statements:

1. Our aim is to solve the inverse problem of mapping the recorded wavefield at the surface of the Earth to the medium properties of the Earth. Specifically, since scattering occurs wherever changes in the Earth's medium properties occur, we seek to solve the inverse scattering problem.
2. The canonical equation of acoustic scattering, the Lippmann-Schwinger equation, is non-linear with respect to the Earth's medium properties, since the unknown Green's functions from the surface into the Earth are a function of the unknown medium properties.
3. To solve the Lippmann-Schwinger inverse problem, linearization is applied to form the Born integral equation, which assumes wavefields are singly-scattering. The Born equation can then be solved by analytical methods if the wavefield is processed to remove higher orders of scattering. Alternatively, we can pose the non-linear problem as an iterative linear problem. Common methods for this include traveltime tomography and full-waveform inversion.
4. Methods have recently been developed to retrieve the unknown Green's functions from the surface into the Earth. Of particular interest are so-called "Marchenko methods".

The aim of this thesis is to combine the Lippmann-Schwinger equation with

Green's functions estimated by a Marchenko method. This linearizes the Lippmann-Schwinger equation with respect to the unknown medium properties without the requirement for the single scattering assumption. In so doing I aim to answer the following two questions:

- *Can the linearized Lippmann-Schwinger equation be inverted for the medium properties?*
- *Does the inclusion of all orders of scattering within the inverse problem give an improved inversion result when compared to single scattering inversion?*

I now outline the structure of the remainder of this thesis. In Chapter 2, I present a theoretical overview of the problem as it exists currently, followed by the theory that is the focus of this thesis: Marchenko-Lippmann-Schwinger methods, that go beyond the existing framework. I begin by deriving the Lippmann-Schwinger equation and linearize to form the Born integral equation. I then present the theory required to understand the inverse problem, both from the perspective of direct matrix inversion and by using full-waveform inversion. I conclude the existing theory with a discussion of Marchenko methods. In the second part of this chapter, I show how the Green's functions from a Marchenko method can be incorporated into the existing inversion framework in two ways. The first of these is by direct matrix inversion of the linearized Lippmann-Schwinger equation, the second is using a new form of full-waveform inversion that requires a new definition for the cost function derivative that correctly handles the Marchenko Green's functions.

In Chapter 3, I discuss the computational implementation of Marchenko-Lippmann-Schwinger methods. I describe the model used to test these methods, as well as the necessary acquisition geometry. I also explain my choice of software packages from a workflow integration perspective.

In Chapter 4, I repeat the 1D analytical Born inversion of a boxcar perturbation by Beydoun and Tarantola (1988), and compare the results to Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversions by calculating the inverse of the kernel matrix. In chapter 5, I present the results for the inversion of a many-layered model in one dimension.

In chapter 6, I show the results of Born and Lippmann-Schwinger inversions for a

2D syncline model, using all available sources and receivers but limited frequencies. In chapter 7, I rewrite the inverse problem for a single source and all available frequencies. The single source inversions can then be stacked as might a migration image to produce a single model update using all sources.

In Chapter 8, I present a discussion of the full body of this work, including the limitations of the Marchenko method as it currently exists and the impacts of using Marchenko estimates of Green's functions in inverse problems. I end this thesis in Chapter 9 with a short summary of my findings.

Chapter 2

Marchenko-Lippmann-Schwinger Methods

In this chapter I describe the theory that underlies the current state-of-the-art methodologies for solving the inverse seismic scattering problem, before building upon and extending this theory to form the basis for the methodologies used in the following chapters, which I call Marchenko-Lippmann-Schwinger or MLS methods. I begin by deriving the fundamental equation of forward- and back- scattering, the Lippmann-Schwinger equation, and show how this reduces to the commonly-used single-scattering approximation, also known as the Born approximation. I then introduce the problem that I aim to solve, the least-squares inversion of the Lippmann-Schwinger equation, and explain the adjoint-state method for solving this problem using the Fréchet derivative of the Born approximation, before briefly explaining the so-called Marchenko methods. Finally, I will bring together the previous three methods to determine the Fréchet derivative of the Lippmann-Schwinger equation, which is the key advancement motivating a new series of least-squares inversion methods that do not require the single-scattering approximation. I call these methods Marchenko-Lippmann-Schwinger methods.

2.1 The Lippmann-Schwinger Equation

Geophysical problems can be broadly separated into two types: forward, and inverse problems. Before we can write the inverse problem, we must first define the forward problem; that is, for a given model, what is the data we expect to observe? For a seismological experiment, we could express that as follows: for a given medium, what is the wavefield we predict we would observe? In this instance, the model is the set of fundamental physical properties that describe the elastic or acoustic behaviour of the Earth. The fundamental equation describing the forward scattering problem is called the Lippmann-Schwinger equation. The approach I will use to derive it follows that of Bleistein et al. (2001) and Schuster (2009).

We begin by writing the wave equation. For the simplest case of an acoustic medium (one in which there is no shear deformation) with constant density, the wave equation is

$$\nabla^2 G\left(\boldsymbol{x}, \boldsymbol{x}'; t\right) - \frac{1}{c^2\left(\boldsymbol{x}\right)} \frac{\partial^2 G}{\partial t^2} = -\delta\left(\boldsymbol{x} - \boldsymbol{x}'\right) \delta\left(t - t'\right) \tag{2.1}$$

where $c(\boldsymbol{x})$ is the speed of a propagating wave at a location $\boldsymbol{x} = (x, y, z)$ in the medium, and $G\left(\boldsymbol{x}, \boldsymbol{x}'; t\right)$ is the Green's function between a source located at $\boldsymbol{x}'$ and the location $\boldsymbol{x}$ inside the medium. The Green's function is the response due to a temporally impulsive, spatially monopole source function (i.e. the right hand side (RHS) of equation 2.1). The observed wavefield $u\left(\boldsymbol{x}, t\right)$ is related to the Green's functions through convolution with the source signature:

$$u\left(\boldsymbol{x}, t\right) = S\left(t\right) * G\left(\boldsymbol{x}, \boldsymbol{x}'; t\right) \equiv \int_{-\infty}^{\infty} S\left(\tau\right) G\left(\boldsymbol{x}, \boldsymbol{x}', t - \tau\right) d\tau \tag{2.2}$$

In the (temporal) frequency domain, equation 2.1 becomes the Helmholtz equation:

$$\nabla^2 \tilde{G}\left(\boldsymbol{x}, \boldsymbol{x}'; \omega\right) + \frac{\omega^2}{c^2\left(\boldsymbol{x}\right)} \tilde{G}\left(\boldsymbol{x}, \boldsymbol{x}'; \omega\right) = -\delta\left(\boldsymbol{x} - \boldsymbol{x}'\right) \tag{2.3}$$

where $\omega = 2\pi f$ is the angular frequency. Since convolution in the time domain is equivalent to multiplication in the frequency domain, the observed wavefield in the

frequency domain is thus

$$\tilde{u}\left(\boldsymbol{x},\omega\right)=\tilde{S}\left(\omega\right)\tilde{G}\left(\boldsymbol{x},\boldsymbol{x}';\omega\right) \tag{2.4}$$

For a three-dimensional constant velocity medium, equation 2.1 has two propagating Green's function solutions, G_{in} and G_{out}, describing an incoming and an outgoing spherical plane wave respectively:

$$G_{in}=\frac{e^{-ikr}}{4\pi r},\quad G_{out}=\frac{e^{ikr}}{4\pi r}$$

where $k=\omega/c$ is the wavenumber, $r=\sqrt{(x^2+y^2+z^2)}$ is the radial distance from the origin, and $i=\sqrt{-1}$ is the imaginary constant. For a physical source excited at time $t=0$, energy causally radiates outwards from the source location at positive times; all energy incoming from infinity must be acausal. Hence solution G_{in} must be excluded.

To ensure the uniqueness of the (causal) solution, we require that the Sommerfeld outgoing radiation condition is applied at infinity and held uniformly in all directions (Sommerfeld, 1912; Schot, 1992):

$$\lim_{r\to\infty} r\left(\frac{\partial\tilde{G}}{\partial r}-ik\tilde{G}\right)=0 \tag{2.5}$$

The non-trivial solution to equation 2.5 requires that the parentheses tend towards zero in the limit that the radial distance from the origin approaches infinity. This means that the outward directional derivative of the Green's function is proportional to the amplitude of the Green's function as the wavefront approaches infinity. For a physical source, the outgoing energy per unit area on the spherical wavefront decreases proportionally to $1/r^2$, and since energy is proportional to the amplitude squared, both the outgoing derivative and the amplitude tend to zero as the wavefront approaches infinity. For an incoming wavefield due to, say, an infinite array of sources located at infinity, as one approaches infinity the wavefront amplitude increases, since one is approaching the sources. However, the outward-directed derivative of the wavefront becomes increasingly negative, since the wavefront is moving towards the origin. Thus

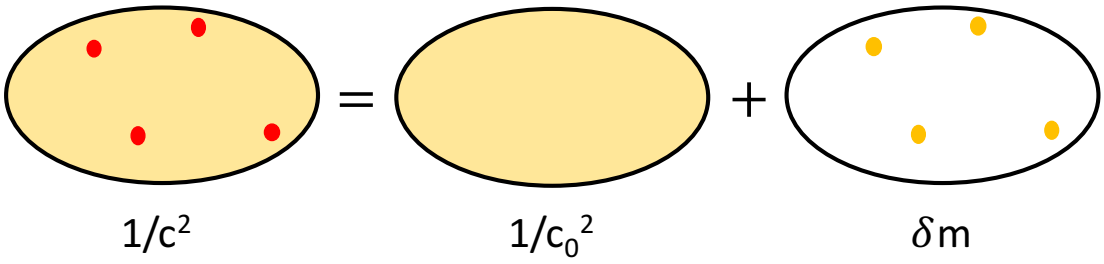


Figure 2.1: The perturbation theory model description, which says that the true medium (left) can be described by a known estimate of the true medium (centre) plus a small perturbation to that estimate (right).

the Sommerfeld radiation condition is only valid for a wavefield incoming from infinity **if** the amplitude of the sources is zero i.e. **there is no incoming wavefield**.

A useful property of the frequency domain Green's function is that it exhibits source-receiver reciprocity; that is, the Green's function is unchanged if the source and receiver spatial coordinates are interchanged. We can write this as

$$\tilde{G}\left(\boldsymbol{x}, \boldsymbol{x}'; \omega\right) = \tilde{G}\left(\boldsymbol{x}', \boldsymbol{x}; \omega\right) \tag{2.6}$$

Herein I will **only use the frequency domain formulation**, and will thus drop the " ˜ " notation for functions of frequency; where a function needs to be expressed in a time domain formulation, the time dependency is explicitly stated.

From equations 2.1 and 2.2, it is clear that the Green's function at a given time or frequency depends solely on the reciprocal of the wavespeed-squared, $1/c^2(\boldsymbol{x})$, also called the slowness-squared. That is to say, the model parameter m of the forward problem is the slowness-squared in the medium, and not the wavespeed in the medium. We now apply perturbation theory and suppose that, in fact, the exact medium with slowness-squared $1/c^2(\boldsymbol{x})$ can be described as consisting of some reference background medium with a slowness-squared $1/c_0^2(\boldsymbol{x})$, plus a small perturbation $\delta m(\boldsymbol{x})$:

$$\frac{1}{c^2\left(\boldsymbol{x}\right)} = \frac{1}{c_0^2\left(\boldsymbol{x}\right)} + \delta m\left(\boldsymbol{x}\right) \tag{2.7}$$

An example of this configuration is shown in Figure 2.1.

We can also write a wave equation for the background medium with its own Green's function , so that

$$\nabla^2 G_0(\boldsymbol{x}, \boldsymbol{x}'; \omega) + \frac{\omega^2}{c_0^2(\boldsymbol{x})} G_0(\boldsymbol{x}, \boldsymbol{x}'; \omega) = -\delta(\boldsymbol{x} - \boldsymbol{x}') \tag{2.8}$$

Within the framework of our perturbation theory, we can decompose the wavefield $G(\boldsymbol{x}, \boldsymbol{x}'; \omega)$ into a wavefield that propagates solely in the background medium $G_0(\boldsymbol{x}, \boldsymbol{x}'; \omega)$, plus a wavefield due to the perturbations to the background model which we will label $G_S(\boldsymbol{x}, \boldsymbol{x}'; \omega)$, where the "S" stands for scattering. Thus we define the scattering Green's function

$$G_S(\boldsymbol{x}, \boldsymbol{x}'; \omega) \equiv G(\boldsymbol{x}, \boldsymbol{x}'; \omega) - G_0(\boldsymbol{x}, \boldsymbol{x}'; \omega) \tag{2.9}$$

Our goal is to formulate the scattering Green's function in terms of the model parameter, the slowness-squared of the medium $1/c^2(x)$, so that we may ultimately find an expression for the model parameter in terms of the scattered Green's function.

We start by substituting equation 2.7 into equation 2.3 so that

$$\left[\nabla^2 + \frac{\omega^2}{c_0^2(\boldsymbol{x})}\right] G(\boldsymbol{x}, \boldsymbol{x}'; \omega) = -\delta(\boldsymbol{x} - \boldsymbol{x}') - \omega^2 \delta m(\boldsymbol{x}) G(\boldsymbol{x}, \boldsymbol{x}'; \omega) \tag{2.10}$$

Using the product rule of differentiation, one finds

$$\nabla^2 G(\boldsymbol{x}, \boldsymbol{x}'; \omega) G_0(\boldsymbol{x}'', \boldsymbol{x}; \omega) = \nabla \cdot [G(\boldsymbol{x}, \boldsymbol{x}') \nabla G_0(\boldsymbol{x}'', \boldsymbol{x})] - \nabla G(\boldsymbol{x}, \boldsymbol{x}') \cdot \nabla G_0(\boldsymbol{x}'', \boldsymbol{x}) \tag{2.11a}$$

$$\nabla^2 G_0(\boldsymbol{x}'', \boldsymbol{x}; \omega) G(\boldsymbol{x}, \boldsymbol{x}'; \omega) = \nabla \cdot [G_0(\boldsymbol{x}'', \boldsymbol{x}) \nabla G(\boldsymbol{x}, \boldsymbol{x}')] - \nabla G_0(\boldsymbol{x}'', \boldsymbol{x}) \cdot \nabla G(\boldsymbol{x}, \boldsymbol{x}') \tag{2.11b}$$

where the index x'' is the observation point of a wave originating at x' and scattering about the point x. Multiplying equation 2.10 by $G_0(\boldsymbol{x}'', \boldsymbol{x})$ and using equations 2.11a

& 2.11b gives

$$\nabla \cdot \left[G\left(\boldsymbol{x}, \boldsymbol{x}'\right) \nabla G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right) - G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right) \nabla G\left(\boldsymbol{x}, \boldsymbol{x}'\right) \right] + \left[\nabla^2 + \frac{\omega^2}{c_0^2\left(\boldsymbol{x}\right)} \right] G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right) G\left(\boldsymbol{x}, \boldsymbol{x}'\right)$$
$$= -\delta\left(\boldsymbol{x} - \boldsymbol{x}'\right) G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right) - \omega^2 \delta m\left(\boldsymbol{x}\right) G\left(\boldsymbol{x}, \boldsymbol{x}'\right) G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right) \tag{2.12}$$

Substituting equation 2.8 into equation 2.12, and using source-receiver reciprocity (equation 2.6) gives

$$\begin{aligned} \nabla \cdot \left[G\left(\boldsymbol{x}, \boldsymbol{x}'\right) \nabla G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right) - G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right) \nabla G\left(\boldsymbol{x}, \boldsymbol{x}'\right) \right] - \delta\left(\boldsymbol{x} - \boldsymbol{x}''\right) G\left(\boldsymbol{x}, \boldsymbol{x}'\right) \\ = -\delta\left(\boldsymbol{x} - \boldsymbol{x}'\right) G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right) - \omega^2 \delta m\left(\boldsymbol{x}\right) G\left(\boldsymbol{x}, \boldsymbol{x}'\right) G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right) \end{aligned} \tag{2.13}$$

The final step is to integrate equation 2.13 over the volume enclosed by a sphere at infinity. To do this, we first use Green's theorem,

$$\begin{aligned} \int_V \nabla \cdot \left[G\left(\boldsymbol{x}, \boldsymbol{x}'\right) \nabla G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right) - G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right) \nabla G\left(\boldsymbol{x}, \boldsymbol{x}'\right) \right] d\boldsymbol{x} \\ = \int_{\partial V} \left[G\left(\boldsymbol{x}, \boldsymbol{x}'\right) \left(\hat{\boldsymbol{n}} \cdot \nabla G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right)\right) - G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right) \left(\hat{\boldsymbol{n}} \cdot \nabla G\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right) \right] dS \end{aligned} \tag{2.14}$$

where V is the volume enclosed by the surface ∂V, $\hat{\boldsymbol{n}}$ is the outward-directed unit vector normal to the surface ∂V; to convert the integral over the first term on the LHS of equation 2.13 from a volume integral to a surface integral:

$$\int_V \left[\delta\left(\boldsymbol{x} - \boldsymbol{x}''\right) G\left(\boldsymbol{x}, \boldsymbol{x}'\right) - \delta\left(\boldsymbol{x} - \boldsymbol{x}'\right) G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right) - \omega^2 \delta m\left(\boldsymbol{x}\right) G\left(\boldsymbol{x}, \boldsymbol{x}'\right) G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right) \right] d\boldsymbol{x}$$
$$= \int_{\partial V} \left[G\left(\boldsymbol{x}, \boldsymbol{x}'\right) \left(\hat{\boldsymbol{n}} \cdot \nabla G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right)\right) - G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right) \left(\hat{\boldsymbol{n}} \cdot \nabla G\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right) \right] dS \tag{2.15}$$

The surface integral on the RHS of equation 2.15 is evaluated by applying the Sommerfeld radiation condition (equation 2.5). The requirement that the spatial derivative of the Green's function decay to zero in the limit that $r \to \infty$ means that the dot products, and thus the entire RHS of equation 2.15, are equal to zero.

Evaluating the volume integral gives us our final result:

$$G(\boldsymbol{x}'', \boldsymbol{x}') = G_0(\boldsymbol{x}'', \boldsymbol{x}') + \omega^2 \int_V \delta m(\boldsymbol{x}) G_0(\boldsymbol{x}'', \boldsymbol{x}) G(\boldsymbol{x}, \boldsymbol{x}') \, d\boldsymbol{x} \tag{2.16}$$

This is the Lippmann-Schwinger equation. It says that the full Green's function $G(\boldsymbol{x}'', \boldsymbol{x}')$ between a source located at $\boldsymbol{x}'$ and a receiver at $\boldsymbol{x}''$ in the true medium is equal to the background Green's function $G_0(\boldsymbol{x}'', \boldsymbol{x}')$ between the same source and receiver pair plus a sum of Green's functions over all points in the medium.

From equation 2.9, it is evident the integral on the RHS of equation 2.16 is equal to the scattered Green's function $G_S(\boldsymbol{x}'', \boldsymbol{x}')$. Thus

$$G_S(\boldsymbol{x}'', \boldsymbol{x}') \equiv \omega^2 \int_V \delta m(\boldsymbol{x}) G_0(\boldsymbol{x}'', \boldsymbol{x}) G(\boldsymbol{x}, \boldsymbol{x}') \, d\boldsymbol{x} \tag{2.17}$$

This is the canonical form of the Lippmann-Schwinger equation. From it we see that the scattered wavefield is constructed from the product of a full Green's function from the source to every point in the medium, and a background Green's function from every point in the medium to the receiver(s), weighted by the model perturbation at each point in the medium. It is convenient to imagine there exists a virtual source (virtual in the sense that it does not exist in the actual physical medium e.g. the Earth's subsurface) at each and every point in the medium. These sources take as their source functions the full Green's functions from the true source at $\boldsymbol{x}'$ to each point in the medium $\boldsymbol{x}$, the amplitude or gain applied to each source is determined by the magnitude of the model perturbation at that point $\delta m(\boldsymbol{x})$. It is for this reason that the model perturbation is also known as the *scattering potential*, it controls how strong the scattering at any location is – if the model perturbation is zero at any point, then the strength of the scattering interaction about that point is also zero; that is there is no scattering at that location. These virtual sources then transmit a background Green's function directly to the receiver. This is shown in Figure 2.2.

If we were to naïvely guess the form of the Lippmann-Schwinger equation, we might assume it to be a combination of Green's functions G in the true medium, after all, the Lippmann-Schwinger equation is supposed to describe all orders of scattering

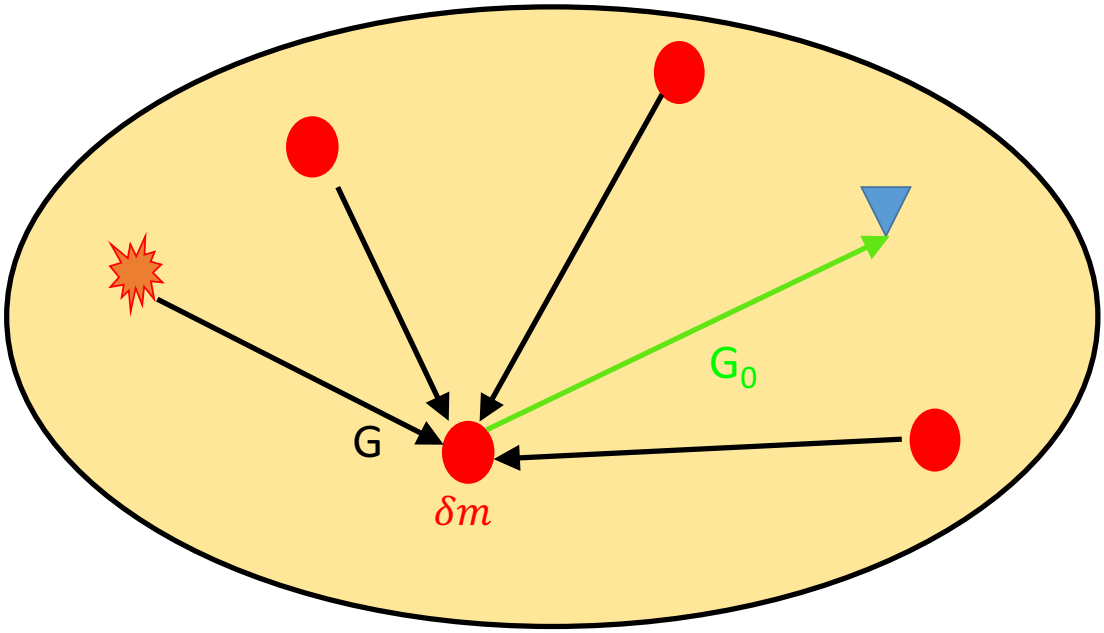


Figure 2.2: The Lippmann-Schwinger description of the Green's function between a source (red and orange explosion) and a receiver (blue triangle) for the medium shown in Figure 2.1.

in the true medium. The virtual source model helps us intuitively understand why this naïve description cannot be true. Substituting equations 2.8 & 2.9 into equation 2.10 gives

$$\left[\nabla^2 + \frac{\omega^2}{c_0^2(\boldsymbol{x})}\right] G_S(\boldsymbol{x}, \boldsymbol{x}'; \omega) = -\omega^2 \delta m(\boldsymbol{x})\, G(\boldsymbol{x}, \boldsymbol{x}'; \omega) \tag{2.18}$$

which is a wave equation of the Helmholtz-type for the background medium with slowness-squared $1/c_0^2(\boldsymbol{x})$ that propagates the wavefield G_S when a source function $-\omega^2 \delta m(\boldsymbol{x})\, G(\boldsymbol{x}, \boldsymbol{x}'; \omega)$ is excited in the background medium. The observed scattering Green's function $G_S(\boldsymbol{x}, \boldsymbol{x}'; \omega)$ is thus the combination of the Green's function G we would observe inside the true medium, and a background model Green's function G_0 that propagates this information to our receivers without further scattering. The fact that the scattering wavefield G_S propagates in the background medium means that the full scattering wavefield G cannot be a function solely of the true medium and its Green's functions.

Finally, it is worth noting that the physics of the Lippmann-Schwinger equation is independent of the choice of source and receiver coordinates in the wave equations 2.3

& 2.8 – you can choose to derive the Lippmann-Schwinger equation by considering a background medium Green's function $G_0(\boldsymbol{x}, \boldsymbol{x}'; \omega)$ from real source location $\boldsymbol{x}'$ to the point $\boldsymbol{x}$, and a full Green's function $G(\boldsymbol{x}'', \boldsymbol{x}; \omega)$ in the true medium from the point $\boldsymbol{x}$ to the real receiver $\boldsymbol{x}''$. In this case, the Lippmann-Schwinger equation reads

$$G_S(\boldsymbol{x}'', \boldsymbol{x}') \equiv \omega^2 \int_V \delta m(\boldsymbol{x})\, G(\boldsymbol{x}'', \boldsymbol{x})\, G_0(\boldsymbol{x}, \boldsymbol{x}')\, d\boldsymbol{x} \tag{2.19}$$

2.1.1 The Born Approximation

In the previous section, we showed how the fundamental equation of forward scattering, the Lippmann-Schwinger equation, is derived from the acoustic wave equation for the case of a constant density medium. We will now explore the meaning of the Lippmann-Schwinger equation and show how the Born approximation emerges from it.

We begin by returning to equation 2.16. Observe that the equation is implicit, that is, that the function we are trying to solve for (in this case the true medium Green's function G) is itself a function of the true medium Green's function, albeit evaluated between different spatial coordinates. This means two things: the first is that without the means to evaluate the Green's function G, either by measurement, mathematical modelling, or some other means; that the Lippmann-Schwinger equation **cannot** be used to solve for the Green's function on the LHS of the equation. The second, is that we can re-write the equation by substituting the equation into itself. For example,

$$\begin{aligned}
G(\boldsymbol{g}, \boldsymbol{s}) = G_0(\boldsymbol{g}, \boldsymbol{s}) \quad & + \omega^2 \int_V \delta m(\boldsymbol{x})\, G_0(\boldsymbol{g}, \boldsymbol{x})\, G(\boldsymbol{x}, \boldsymbol{s})\, d\boldsymbol{x} \\
= G_0(\boldsymbol{g}, \boldsymbol{s}) \quad & + \omega^2 \int_V \delta m(\boldsymbol{x})\, G_0(\boldsymbol{g}, \boldsymbol{x}) \{G_0(\boldsymbol{x}, \boldsymbol{s}) \\
& + \omega^2 \int_V \delta m(\boldsymbol{x}')\, G_0(\boldsymbol{x}, \boldsymbol{x}')\, G(\boldsymbol{x}', \boldsymbol{s})\, d\boldsymbol{x}'\} d\boldsymbol{x} \qquad (2.20a) \\
= G_0(\boldsymbol{g}, \boldsymbol{s}) \quad & + \omega^2 \int_V \delta m(\boldsymbol{x})\, G_0(\boldsymbol{g}, \boldsymbol{x}) \{G_0(\boldsymbol{x}, \boldsymbol{s}) \\
& + \omega^2 \int_V \delta m(\boldsymbol{x}')\, G_0(\boldsymbol{x}, \boldsymbol{x}') [G_0(\boldsymbol{x}', \boldsymbol{s}) \\
& + \omega^2 \int_V \delta m(\boldsymbol{x}'')\, G_0(\boldsymbol{x}', \boldsymbol{x}'')\, G(\boldsymbol{x}'', \boldsymbol{s})\, d\boldsymbol{x}''] d\boldsymbol{x}'\} d\boldsymbol{x} \qquad (2.20b)
\end{aligned}$$

etc.

Equation 2.20a says that the scattered Green's function can be described as the sum of a single scattering term $\delta m\left(\boldsymbol{x}\right) G_0\left(\boldsymbol{g},\boldsymbol{x}\right) G_0\left(\boldsymbol{x},\boldsymbol{s}\right)$ plus a scattering integral. From equation 2.20b we see that this is equivalent to the sum of the single scattering term and a scattering integral plus a second-order term
$\delta m\left(\boldsymbol{x}\right) G_0\left(\boldsymbol{g},\boldsymbol{x}\right) \delta m\left(\boldsymbol{x}\right) G_0\left(\boldsymbol{x},\boldsymbol{x}'\right) G_0\left(\boldsymbol{x}',\boldsymbol{s}\right)$. This process can be iterated *ad infinitum*, and thus we conclude that the scattering Green's function can be constructed from the sum of every order of scattering. In principle this is an infinite sum, however in practice there may only be a finite number of scattering events detectable due to a combination of attenuation effects and the sensitivity of the instruments used in an experiment. Figure 2.3 shows the deconstructed scattered wavefield in a layered medium for the first three orders of scattering.

At this point we must address what we mean physically by the first, second etc. orders of scattering. This is especially important in reflection seismology where it is commonplace to describe the data in terms of singly-scattered waves or "primary reflections", and multiply-scattered waves or "multiple reflections", often reduced simply to "multiples". In the Earth's subsurface, reflections and other scattering events occur due to contrasts in the material properties (e.g. P-wave speed, S-wave speed, density etc.) which is assumed to correlate with changes in the geological structure of the subsurface. The scattering described by the Lippmann-Schwinger equation is due to a contrast between the true medium and the reference medium. The choice of reference medium, whilst generally informed by the data, can in principle be completely arbitrary; as a consequence, if we were able to compare the wavefield propagating in the true medium to the scattered wavefield calculated at a given frequency or snapshot in time, the two wavefields may disagree despite reproducing the observed data exactly. It is only when the reference medium is informed by accurate geological or geophysical prior information that the scattering potential of the Lippmann-Schwinger equation may correlate to changes in the underlying geological structure of the true medium, which is where physical scattering of waves occurs. If the reference medium is chosen without the use of prior information, or if the prior information is inaccurate, the scattering potential will be non-zero at locations other

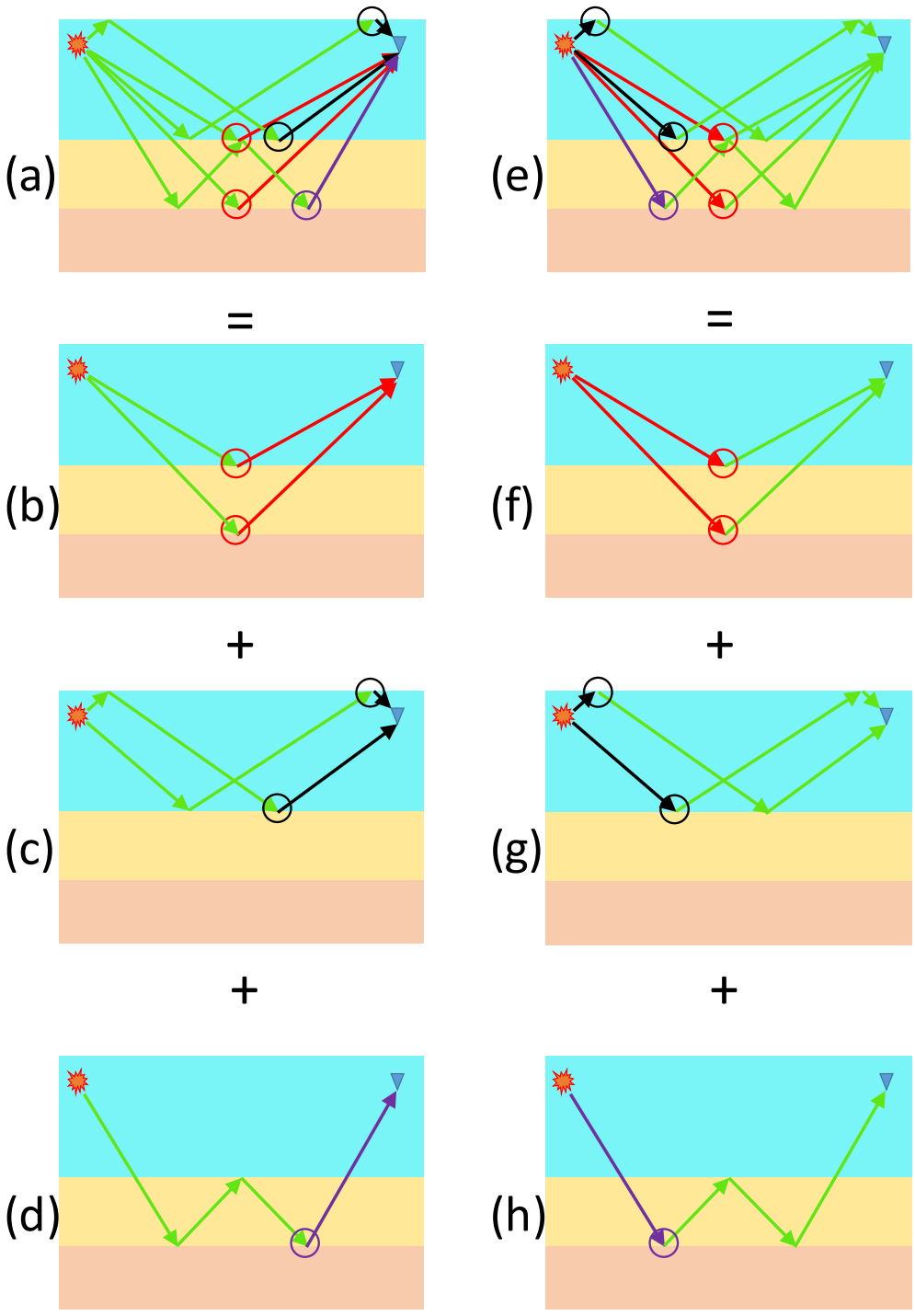


Figure 2.3: The Lippmann-Schwinger description of the scattered wavefield as the sum of scattering orders, both for the formulation in equation 2.17 (a to d) and the formulation in equation 2.19 (e to h). Only the first (b & f), second (c & g), and third (d & h) orders of scattering are shown.

than where discontinuities occur in the true medium. These locations in turn become scatterers in the Lippmann-Schwinger equation, and the scattered wavefield described by the Lippmann-Schwinger equation no longer resembles the scattered wavefield in the true medium. This non-uniqueness is important to recognise when we come to formulate the inverse problem.

Data-driven methods of velocity analysis can be used to find the reference model such as picking velocity spectra semblances for NMO stacking velocities (e.g. Taner and Koehler, 1969; Yilmaz, 2001), migration-velocity analysis for migration velocities (e.g. Sava and Biondi, 2004), and ray-based traveltime tomography (e.g. Rawlinson and Sambridge, 2003). Unfortunately, such models are often a smoothed representation of the true medium, and the contrasts within the model are unable to accurately reproduce the scattering we observe in data – it may only be possible to reconstruct the strongest scattering events i.e. the singly-scattered waves. In this case the Lippmann-Schwinger equation is a non-linear function of the model perturbation, since we cannot evaluate the true medium Green's function.

To account for this, it is commonplace to linearize the Lippmann-Schwinger equation by truncating the infinite scattering series after the first order of scattering; this is commonly known as the (first) Born approximation or the Born modelling equation when referring to a forward problem:

$$G_{Born}\left(\boldsymbol{x}'', \boldsymbol{x}'\right) \equiv \omega^2 \int_V \delta m\left(\boldsymbol{x}\right) G_0\left(\boldsymbol{x}'', \boldsymbol{x}\right) G_0\left(\boldsymbol{x}, \boldsymbol{x}'\right) d\boldsymbol{x} \tag{2.21}$$

The main consequence of this linearization is that the forward problem cannot describe the multiply-reflected events we observe in seismic data recorded in the Earth. As such, if we formulate an inverse problem from the Born modelling equation we must first remove as well as possible any events in the observed data that are multiply-scattered, since such events cannot be correctly handled by the algorithm we apply to it.

2.2 The Inverse Scattering Problem

Having defined the forward problem, we can now write the corresponding inverse problem; that is, given that I have measured some data, what is the corresponding model that predicts the data I observed? If this definition seems circumlocutory, it is because while the solution of the forward problem has a clear meaning, the same cannot be said of the inverse problem, if indeed a solution even exists. I now briefly explain the necessary theory of discrete inverse problems to understand my later results; for a more complete discussion see, for example, Menke (1984), Parker (1994) or Scales et al. (1994).

Let us begin by writing the data $\boldsymbol{d}$ as an ordered vector of length N. Then we may write the forward problem as

$$\boldsymbol{d} = A(\boldsymbol{m}) + \boldsymbol{e} \tag{2.22}$$

where $\boldsymbol{m}$ is an ordered vector of length M of the model parameters describing the medium, $\boldsymbol{e}$ is the error associated with the data, and A is a function that maps the model parameters to values of data. The function A is sometimes called the data kernel in analogy to the continuous forward problem. The most general solution to this problem is

$$\hat{\boldsymbol{m}} = A^{-1}(\boldsymbol{d}) \tag{2.23}$$

where $\hat{\boldsymbol{m}}$ is the estimated model vector and A^{-1} is an inverse mapping of the function $A(\boldsymbol{m})$, if it exists. If the function A is linear in the model, then we may write equation 2.22 as

$$\boldsymbol{d} = A\boldsymbol{m} + \boldsymbol{e} \tag{2.24}$$

where A is an NxM matrix operator, which I will refer to as the kernel operator. As written in the previous section, the Lippmann-Schwinger equation is a non-linear forward problem of the form of equation 2.22, whilst the Born forward modelling equation is a linear problem of the form of equation 2.24. Ideally we would solve exactly the non-linear inverse problem to the Lippmann-Schwinger equation, however given that we cannot in general solve that forward problem due to the lack of access

to the full Green's functions G required, I will limit the discussion to the linear inverse problem for which we know the forward problem (the Born modelling equation) can be solved. Those familiar with inverse theory may skip forward to equation 2.34, wherein I frame the Born forward problem (equation 2.21) as a linear problem of the form of equation 2.24.

The most general solution to the linear inverse problem may be written

$$\hat{\boldsymbol{m}} = A^{-g}\boldsymbol{d} \tag{2.25}$$

where A^{-g} is the generalised inverse of the matrix A, which unlike the regular matrix inverse A^{-1} is guaranteed to exist (the regular matrix inverse only exists for square non-singular matrices). Substituting equation 2.24 into 2.25 we can express the linear inverse problem as

$$\hat{\boldsymbol{m}} = \boldsymbol{m} + \left(A^{-g}A - I\right)\boldsymbol{m} + A^{-g}\boldsymbol{e} \tag{2.26}$$

where

$$R_m = A^{-g}A \tag{2.27}$$

is the model resolution matrix and $A^{-g}\boldsymbol{e}$ is the data error propagated into the model domain (Snieder and Trampert, 1999). If $R_m = 1$ then the model is said to be perfectly resolved, meaning that each model parameter in $\hat{\boldsymbol{m}}$ may be determined independently of any other. If the model is not perfectly resolved, then the model parameters cannot be solved for independently, and the estimated model is said to be blurred. Since all the results I present in this thesis are from synthetic experiments, we can expect that the data error due to random noise etc. to be zero. However, even in this ideal case there will still be a propagated error due to the linearization of the problem from the Lippmann-Schwinger equation to the Born modelling equation – unless the data are exactly de-multipled before the inversion process we will observe artefacts in the estimated model because the physics of multiple scattering is incompatible with the linearized physics of the Born modelling equation.

The problems 2.23 and 2.25 are said to be "well-posed" if the following conditions are satisfied (Hadamard, 1902):

a. A solution exists,

b. The solution is unique,

c. The solution depends continuously on the data.

Any problem that does not satisfy all three conditions is said to be "ill-posed", of which the inverse scattering problem is one such example. Dorren et al. (1994) showed that even when the exact inverse problem can be defined, as is the case for the Marchenko equation, that the inversion is unstable i.e. that the solution can change rapidly for small changes in the data, and thus violates condition (c). Even for the linearized inverse problem, the solution may be non-unique.

Accepting that the problem is ill-posed, the most common approach to solving the inverse problem 2.25 is the least-squares approach. The least-squares approach seeks to find the model that minimises the error function

$$E(\boldsymbol{m}) = \|\boldsymbol{d} - A\boldsymbol{m}\|^2 \tag{2.28}$$

where the norm is the Euclidean or $l2$-norm defined as

$$\|\boldsymbol{y}\| \equiv \sqrt{\sum_{i=1}^{n} y_i^2} \tag{2.29}$$

for a vector $\boldsymbol{y}$ of length n. The model that **minimises the norm** is

$$\hat{\boldsymbol{m}}_{LS} = \left(A^T A\right)^{-1} A^T \boldsymbol{d} \tag{2.30}$$

If there are exactly as many data as model parameters, then there exists a unique solution that minimises the error function provided that the inverse matrix $\left(A^T A\right)^{-1}$ exists; such problems are described as exactly-determined. Often the inverse matrix does not exist due to noise in the data.

The least squares solution is ideal for over-determined problems, where there are more data than model parameters $(N > M)$, and so no model exists that can exactly fit all the data. The opposite of an over-determined problem is an under-determined problem, where the number of model parameters exceeds the number of data $(M > N)$, and so there may exist many models that minimise the error function.

It is therefore impossible to uniquely determine every model parameter. We can obtain a unique solution to this problem by asserting the solution satisfy additional conditions, which constitutes a priori information about the model. Following Occam's razor, if we seek the solution with the minimum size that also minimises the error function, that is

$$\hat{\boldsymbol{m}}_{MN} = \min\|\hat{\boldsymbol{m}}\|^2 \text{ subject to } E(\boldsymbol{m}) = 0$$

the so-called "minimum-norm" solution is found to be

$$\hat{\boldsymbol{m}}_{MN} = A^T \left(AA^T\right)^{-1} \boldsymbol{d} \tag{2.31}$$

assuming that the inverse matrix exists.

It is further possible for a problem to be simultaneously over-determined and under-determined, this is called a mixed-determined problem. This is the typical state of the seismic inverse scattering problem. Consider the seismic tomography problem: if we break down the medium into equally sized pixels with independent values of p-wave slowness assigned to each pixel, those pixels closest to the sources and receivers will be sampled by many ray-paths, whilst those furthest from the sources and receivers will be sampled less, or even not at all. Thus the model parameters in the pixels closest to the sources and receivers may be over-determined, whilst the model parameters in pixels far away may be under-determined. The problem is mixed-determined due to the limitations in geophysical experimental design; with greater access to the subsurface we could eliminate this effect. Even with access to the subsurface, a typical seismic tomography experiment would be formally over-determined since we generally have many more data than model parameters. We desire a solution that both constrains the data error close to the source and receiver arrays whilst also preventing model spikes far from the arrays where the model is under sampled. This can be achieved by regularisation of the model (Tikhonov and Arsenin, 1977).

To regularise the problem, we modify the error function to include the a priori information of the solution. By including a priori information, it is possible that one or more of the family of problems that include the a priori information is better conditioned than the original problem, whilst hopefully having solutions similar enough to

the unregularised problem. A popular choice of regularisation is Tikhonov regularisation (also known as ridge regression), which states that the model length (or norm) is small. Adding the model length to the least-squares error function, the Tikhonov regularised error function is

$$E(\boldsymbol{m}) = \|\boldsymbol{d} - A(\boldsymbol{m})\|^2 + \mu^2 \|\boldsymbol{m}\|^2 \tag{2.32}$$

where μ^2 is the scalar regularisation parameter. Defining the regularisation parameter in this way ensures the regularisation term is always positive, preventing the error function from becoming negative. The model that minimises this error function is

$$\hat{\boldsymbol{m}}_{reg} = \left(A^T A + \mu^2 I\right)^{-1} A^T \boldsymbol{d} \tag{2.33}$$

Tikhonov regularisation differs from the minimum-norm solution because the minimum-norm solution aims to minimise the model length of the subset of models that minimise the data error, whilst the regularisation solution solves for the model that minimises the sum of the data error and model length simultaneously. We can change the kind of a priori information used to regularise the problem by modifying the regularisation term in equation 2.32; for example, if we expect that the model is smoothly varying in a given spatial direction, then we might replace the model length by the length of a spatial derivative of the model in the smooth direction.

The discussion thus far is valid for any inverse problem, subject to linearization where appropriate. Let us now return to the inverse Born scattering problem for a single frequency, whose forward problem we can write as a vector equation of the form of equation 2.24 so that

$$\boldsymbol{G}_S = \omega^2 [G_0^{sx} G_0^{xr}] \boldsymbol{V} \tag{2.34}$$

where $\boldsymbol{G}_S$ is the vector of scattered Green's functions between sources and receivers, and $\boldsymbol{V}$ is the vector of scattering potentials. $[G_0^{sx} G_0^{xr}]$ is the matrix whose components are the products of the background model Green's functions from a source s to a given location in the model x, and the background model Green's functions from the same

location x in the model to a receiver r. For a constant-density acoustic model with m model coordinates, $\boldsymbol{V}$ is always a vector of size 1xm of slowness-squared model perturbations at each of the m model coordinates. The size of $\boldsymbol{G}_S$ and $[G_0^{sx}G_0^{xr}]$ is determined both by the size of the model and the acquisition geometry. For a single source-receiver pair, $\boldsymbol{G}_S$ is a 1x1 vector and the matrix $[G_0^{sx}G_0^{xr}]$ reduces to a mx1 vector; for p sources and q receivers there are a maximum of pq source-receiver pairs, thus $\boldsymbol{G}_S$ is a 1xpq vector and $[G_0^{sx}G_0^{xr}]$ is a mxpq matrix (assuming all channels are available) where the rows of $\boldsymbol{G}_S$ and $[G_0^{sx}G_0^{xr}]$ are arranged in the same source-receiver pair order. This form equation neatly mirrors the form of equation 2.24, where $\boldsymbol{G}_S$ corresponds to the data vector $\boldsymbol{d}$, $\boldsymbol{V}$ corresponds to the model vector $\boldsymbol{m}$, and $[G_0^{sx}G_0^{xr}]$ corresponds to the kernel operator A. This can be summarised as follows:

$$\begin{array}{lll} \boldsymbol{G}_S & = \omega^2 & [G_0^{sx}G_0^{xr}] & \boldsymbol{V} \end{array}$$

$$\begin{bmatrix} G_{S,1} \\ G_{S,2} \\ \vdots \\ G_{S,pq} \end{bmatrix} = \omega^2 \begin{bmatrix} (G_0G_0)_{1,1} & (G_0G_0)_{2,1} & \dots & (G_0G_0)_{m,1} \\ (G_0G_0)_{1,2} & (G_0G_0)_{2,2} & \dots & (G_0G_0)_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ (G_0G_0)_{1,pq} & (G_0G_0)_{2,pq} & \dots & (G_0G_0)_{m,pq} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_m \end{bmatrix} \tag{2.35}$$

$$\begin{aligned} \boldsymbol{G}_S &\longleftrightarrow \boldsymbol{d} \\ \omega^2[G_0^{sx}G_0^{xr}] &\longleftrightarrow A \\ \boldsymbol{V} &\longleftrightarrow \boldsymbol{m} \end{aligned}$$

In the literature (e.g. Araújo et al., 1994; Weglein et al., 1997; Kouri and Vijay, 2003; Lesage et al., 2013), the forward problem (equation 2.34) is more commonly written as

$$\boldsymbol{G}_S = \omega^2 \boldsymbol{G}_0^T V \boldsymbol{G}_0 \tag{2.36}$$

where the vectors $\boldsymbol{G}_0$ have size 1 x m for the single source-receiver pair case, and V is a m x m diagonal matrix whose off-diagonal elements are equal to 0. In these cases, the authors seek to use the iterative nature of the Lippmann-Schwinger equation to

pose a different, but nonetheless equivalent, inverse problem to the direct inversion I seek to describe. Due to matrix multiplication being a non-commutative operation (i.e.$BC \neq CB$ in general), direct inversion of equation 2.36 requires that both a left-sided inverse matrix and a right-sided inverse matrix exist; that is, for a matrix B, the left-sided inverse matrix B^{-1}_{left} and right-sided inverse matrix B^{-1}_{right} may be defined as

$$I \equiv B^{-1}_{left} B$$
$$I \equiv B B^{-1}_{right}$$

Applying left- and right-sided inverse operators as defined above to equation 2.36 requires use of the outer product on the left-hand side of the equation so that the sizes of the resultants are consistent on both sides of the equation. It therefore makes sense to formulate the inverse problem from equation 2.34 and not equation 2.36. The direct inverse problem can then be stated as

$$\omega^2 \boldsymbol{V} = [G_0^{sx} G_0^{xr}]^{-g} \boldsymbol{G}_S \tag{2.37}$$

where "-g" indicates that the generalized inverse has been applied, since the kernel operator $(G_0 G_0)$ is only square if the number of source-receiver pairs equals the number of model parameters. The least-squares solution to equation 2.34 is

$$\omega^2 \boldsymbol{V}_{LS} = \{[G_0^{sx} G_0^{xr}]^T [G_0^{sx} G_0^{xr}]\}^{-1} [G_0^{sx} G_0^{xr}]^T \boldsymbol{G}_S \tag{2.38}$$

and the regularised least-squares solution is

$$\omega^2 \boldsymbol{V}_{reg} = \{[G_0^{sx} G_0^{xr}]^T [G_0^{sx} G_0^{xr}] + \mu^2 I\}^{-1} [G_0^{sx} G_0^{xr}]^T \boldsymbol{G}_S \tag{2.39}$$

2.2.1 Full-Waveform Inversion

The inverse problem as I have already described it has an elegant and compact solution. Unfortunately it is not always appropriate to approach the problem in this manner. One consideration is the computational complexity. The matrix $A^T A$ has

size M x M. To solve the inverse of this matrix by Gauss-Jordan elimination would require $\mathcal{O}(M^3)$ calculations, thus for large models the cost of calculating the matrix inverse becomes intractable. And whilst it is true that computational power continues to increase, so too does the size of the models, so direct inversion remains a viable option only for specific problems where the matrix to be inverted has a special form that significantly reduces the computational cost. Another consideration is the ill-posedness of the problem. As the severity of the ill-posedness increases, so too does the difficulty in obtaining a meaningful solution, and problems with instability of the algorithm become unmanageable or the level of regularisation required to control it becomes so excessive as to provide a physically implausible solution. We thus require alternatives to direct inversion to solve inverse problems.

One such approach is full-waveform inversion (FWI), which aims to solve the full non-linear inverse problem whilst also addressing some of the problems of direct inversion. Full-waveform inversion transforms the non-linear inverse problem (equation 2.23) into an iterative linear optimization problem (Lailly, 1983; Tarantola, 1984) which can more easily be solved. Another benefit of this transformation is that the inversion of the square of the kernel operator is replaced by a migration-style operation, which can be solved efficiently by forward modelling. For a broader overview, I refer the reader to Virieux and Operto (2009).

We begin the full-waveform inversion approach by defining the data residual vector $\Delta\boldsymbol{d}$ as the difference between the observed data vector $\boldsymbol{d}_{obs}$ and the modelled data vector $\boldsymbol{d}_{cal}(\boldsymbol{m})$, which is a function of the model:

$$\Delta\boldsymbol{d} \equiv \boldsymbol{d}_{obs} - \boldsymbol{d}_{cal}(\boldsymbol{m}) \tag{2.40}$$

where the model $\boldsymbol{m}$ is the slowness-squared of the medium i.e. $\boldsymbol{m} = 1/c^2(\boldsymbol{x})$. We define the error function, known as the cost function, as half of the L2-norm of the data residual, so that

$$\begin{aligned} C(\boldsymbol{m}) &= \frac{1}{2}\|\Delta\boldsymbol{d}(\boldsymbol{m})\|^2 \\ &= \frac{1}{2}\Delta\boldsymbol{d}^\dagger\Delta\boldsymbol{d} \end{aligned} \tag{2.41}$$

where † represents the adjoint or conjugate transpose of the vector. The cost function implicitly assumes summing over all sources and receivers. The choice of norm is arbitrary; however, it is standard to choose the L2-norm.

If we impose a perturbation solution onto the model so that

$$\boldsymbol{m} = \boldsymbol{m}_0 + \Delta\boldsymbol{m} \tag{2.42}$$

where $\boldsymbol{m}_0 = 1/c_0^2(\boldsymbol{x})$ is the slowness-squared of a known reference medium and $\Delta\boldsymbol{m}$ is an unknown perturbation to the reference medium, our objective is then to find the model $\boldsymbol{m}$ in the vicinity of the reference model that minimises the cost function. Substituting our perturbation solution into the cost function, we can expand the cost function in a Taylor series about the known reference model $\boldsymbol{m}_0$. This gives us that

$$C\left(\boldsymbol{m}\right) = C\left(\boldsymbol{m}_0 + \Delta\boldsymbol{m}\right) \tag{2.43a}$$

$$= C\left(\boldsymbol{m} = \boldsymbol{m}_0\right) + \sum_{j=1}^{M} \frac{\partial C\left(\boldsymbol{m}_0\right)}{\partial m_j}\Delta m_j + \sum_{j=1}^{M}\sum_{k=1}^{M} \frac{\partial^2 C\left(\boldsymbol{m}_0\right)}{\partial m_j \partial m_k}\Delta m_j \Delta m_k + \mathcal{O}(\Delta\boldsymbol{m})^3 \tag{2.43b}$$

Taking the derivative of the cost function with respect to a single model parameter m_l gives

$$\frac{\partial C\left(\boldsymbol{m}\right)}{\partial m_l} = \frac{\partial C\left(\boldsymbol{m}_0\right)}{\partial m_l} + \sum_{j=1}^{M} \frac{\partial^2 C\left(\boldsymbol{m}_0\right)}{\partial m_j \partial m_l}\Delta m_j \tag{2.44}$$

Since the model we are looking for minimises the cost function in the local vicinity of the reference model, the model perturbation that updates the reference model to that model is given by

$$\Delta\boldsymbol{m} = -\left[\sum_{j=1}^{M} \frac{\partial^2 C\left(\boldsymbol{m}_0\right)}{\partial \boldsymbol{m}^2}\right]^{-1} \frac{\partial C\left(\boldsymbol{m}_0\right)}{\partial \boldsymbol{m}} \tag{2.45}$$

where $\partial C/\partial\boldsymbol{m}$ is called the Fréchet derivative. The second derivative matrix $\partial^2 C/\partial\boldsymbol{m}^2$, called the Hessian matrix, is defined as

$$\left[\frac{\partial^2 C}{\partial \boldsymbol{m}^2}\right]_{jl} \equiv \frac{\partial^2 C}{\partial m_j \partial m_l}$$

The class of methods that seeks to calculate the solution using the full Hessian matrix is known as Newton methods. Because the Hessian matrix is expensive to calculate, and even more expensive to invert (since it has dimensionality M x M, to invert the matrix by Gauss-Jordan elimination requires $\mathcal{O}(M^3)$ calculations), if the second derivative of the wavefield with respect to the model parameters is small then the Hessian maybe be approximated as

$$\left[\frac{\partial^2 C}{\partial \boldsymbol{m}^2}\right]_{jl}^{GN} \approx \frac{\partial C}{\partial \boldsymbol{m}}\frac{\partial C^\dagger}{\partial \boldsymbol{m}}$$

Such methods are called Gauss-Newton methods. I assume that the Hessian has the form of a scalar multiplied by the identity matrix. This gives us the steepest-descent algorithm for updating the model,

$$\Delta \boldsymbol{m} = -\alpha I_M \frac{\partial C\left(\boldsymbol{m}_0\right)}{\partial \boldsymbol{m}} \tag{2.46}$$

where the scalar α is called the step-length. Thus the inverse problem has been reduced to solving the derivative of the cost function evaluated at the reference model.

To calculate the derivative of the cost function, we simply apply the product rule to equation 2.41 so that

$$\begin{aligned}
\frac{\partial C\left(\boldsymbol{m}\right)}{\partial \boldsymbol{m}} &= \frac{1}{2}\left[\frac{\partial\left(\boldsymbol{d}_{obs}-\boldsymbol{d}_{cal}(\boldsymbol{m})\right)^\dagger}{\partial \boldsymbol{m}}\Delta\boldsymbol{d} + \Delta\boldsymbol{d}^\dagger\frac{\partial\left(\boldsymbol{d}_{obs}-\boldsymbol{d}_{cal}(\boldsymbol{m})\right)}{\partial \boldsymbol{m}}\right] \\
&= -\frac{1}{2}\left[\frac{\partial\left(\boldsymbol{d}_{cal}(\boldsymbol{m})\right)^\dagger}{\partial \boldsymbol{m}}\Delta\boldsymbol{d} + \Delta\boldsymbol{d}^\dagger\frac{\partial\left(\boldsymbol{d}_{cal}(\boldsymbol{m})\right)}{\partial \boldsymbol{m}}\right] \\
&= -\Re\left[\frac{\partial\left(\boldsymbol{d}_{cal}(\boldsymbol{m})\right)^\dagger}{\partial \boldsymbol{m}}\Delta\boldsymbol{d}\right]
\end{aligned} \tag{2.47}$$

where $\Re$ denotes the real part of the complex-valued argument. Pratt et al (1998) explained the meaning of the Fréchet derivative in terms of virtual sources in the medium by writing the wave equation in the form of a linear forward problem

$$B\boldsymbol{u} = \boldsymbol{f} \Longleftrightarrow \boldsymbol{u} = B^{-1}\boldsymbol{f} \tag{2.48}$$

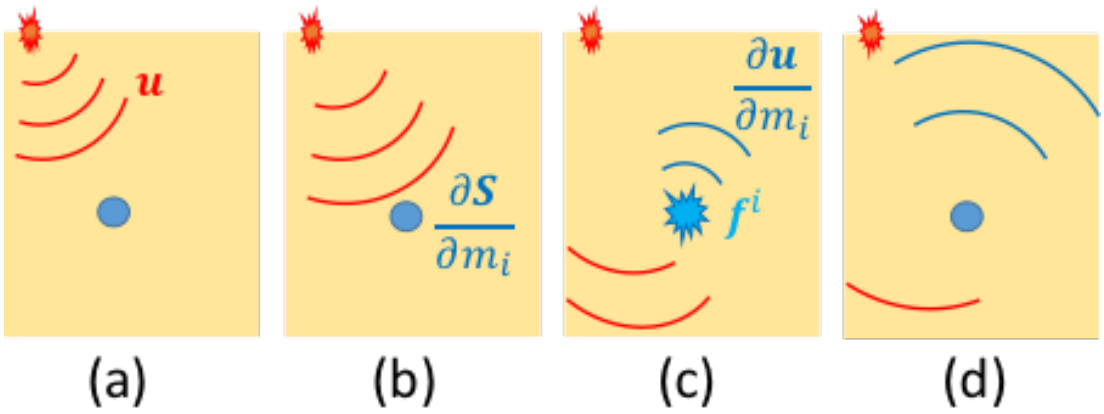


Figure 2.4: The virtual source formulation of the scattered wavefield. The real source excites a wavefield (a), which radiates out and where there are contrasts in the medium (b) partially backscatters the incoming wavefield as if a source was located at the location of the medium contrast (c) radiating a secondary wavefield back towards the surface.

where $\boldsymbol{u}$ is the wavefield excited by a source function $\boldsymbol{f}$, and the matrix B is the impedance matrix which acts as the wave equation operator does in the continuous wave equation. It is clear that the impedance matrix is a representation of the model parameters. By replacing the source function $\boldsymbol{f}$ with a spatio-temporal delta function, the columns of the matrix B^{-1} must also be discrete representations of the Green's functions. Taking the derivative of both sides and applying the product rule, one finds that

$$\frac{\partial B}{\partial \boldsymbol{m}}\boldsymbol{u} = -B\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{m}} \tag{2.49}$$

since the source function is independent of the model. Then

$$\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{m}} = -B^{-1}\frac{\partial B}{\partial \boldsymbol{m}}\boldsymbol{u} \tag{2.50}$$

We see that the Fréchet derivative of the wavefield constitutes a secondary wavefield with a virtual source function $-(\partial B/\partial \boldsymbol{m})\boldsymbol{u}$ and propagated by the matrix B^{-1}, this is shown in Figure 2.4. Since the discrete data is nothing more than the wavefield restricted to the receiver array coordinates i.e. $\boldsymbol{d}_{cal}(\boldsymbol{m}) = \mathcal{R}\boldsymbol{u}(\boldsymbol{m})$ where $\mathcal{R}$ is the restriction matrix. Comparing this equation to equation 2.47 proves this description is also true for the Fréchet derivative of the data. It is for this reason that the derivative is often interpreted as the scattering wavefield due to the model perturbation.

Alternatively, we could have come to this conclusion by inspection of the Born forward modelling equation 2.21 and the definition of the Fréchet derivative $DF(a)$,

$$\lim_{h\to 0}\frac{\|F(a+h)-F(a)+DF(a)h\|}{\|h\|}=0 \tag{2.51}$$

that the Fréchet derivative of the data is given by

$$\frac{\partial\left(\boldsymbol{d}_{cal}(\boldsymbol{m})\right)}{\partial m_j}=\begin{cases}\omega^2 G_0(g,x_i)G_0(x_j,s) & ,i=j\\ 0 & \text{otherwise}\end{cases} \tag{2.52}$$

This is the approach I employ when deriving the Marchenko-Lippmann-Schwinger FWI. By substituting equation 2.52 into equation 2.47, we can then say that

$$\frac{\partial C(\boldsymbol{m})}{\partial \boldsymbol{m}}=-\Re\Big[\omega^2\left(G_0(g,x_i)G_0(x_j,s)\right)^{\dagger}\Delta\boldsymbol{d}\Big] \tag{2.53}$$

Like the direct inversion method in section 2.2, we can also apply regularisation to FWI by modification of the cost function. Adding a Tikhonov regularisation term to equation 2.41, the cost function reads

$$C\left(\boldsymbol{m}\right)=\frac{1}{2}\|\Delta\boldsymbol{d}\left(\boldsymbol{m}\right)\|^2+\mu^2\|\boldsymbol{m}\|^2 \tag{2.54}$$

Consequently, the gradient of the cost function becomes

$$\frac{\partial C(\boldsymbol{m})}{\partial \boldsymbol{m}}=-\Re\Big[\omega^2\left(G_0(g,x_i)G_0(x_j,s)\right)^{\dagger}\Delta\boldsymbol{d}\Big]+\mu^2\boldsymbol{m} \tag{2.55}$$

This type of regularisation has previously been used in FWI by Abubakar et al. (2009), although other regularisation schemes such as Total Variation are also applicable to FWI (Vogel and Oman, 1996; van den Berg and Abubakar, 2001; Vogel, 2002). The total variation functional is defined as (Rudin et al., 1992)

$$J_{TV}=\int_{\Omega}|\nabla u(x)|dx$$

where $|\cdot|$ is the L1-norm. For least-squares problems one may replace the L1-norm

with the L2-norm (Vogel and Oman, 1996). Total variation regularisation is applied by replacing the Tikhonov regularisation term in equation 2.54 by the total variation functional (weighted by some positive scalar).

To conclude this section, I now combine the previously written equations into a single full-waveform inversion algorithm:

1. Record data $\boldsymbol{d}$ in the true Earth as well as its source function
2. Do forward modelling in the reference model $\boldsymbol{m}_0$ to record data $\boldsymbol{d}_{cal}(\boldsymbol{m}_0)$

   While cost function has not converged:
   (a) Calculate residual $\Delta\boldsymbol{d} \equiv \boldsymbol{d}_{obs} - \boldsymbol{d}_{cal}(\boldsymbol{m})$
   (b) Back-propagate adjoints from receiver locations
   (c) Cross-correlate forward- and back-propagated wavefields to get $\frac{\partial C(\boldsymbol{m})}{\partial(\boldsymbol{m})}$
   (d) Linesearch to obtain optimal step length α
   (e) Update reference model according to $\boldsymbol{m}_k = \boldsymbol{m}_{k-1} - \alpha\frac{\partial C(\boldsymbol{m})}{\partial(\boldsymbol{m})}$
   (f) Do forward modelling in the reference model $\boldsymbol{m}_k$ to record data $\boldsymbol{d}_{cal}(\boldsymbol{m}_k)$
3. Accept reference model as solution

2.3 Marchenko Methods

So far, I have discussed how a forward scattering problem can be posed starting from the scalar Helmholtz equation, resulting in the so-called Lippmann-Schwinger equation; and how one might go about solving the subsequent inverse scattering problem. The scalar Helmholtz equation is, however, not the only scalar wave equation in physics; in fact, many different wave equations exist. Perhaps the most well-known of the last century has been the Schrödinger equation, which is central to the study of quantum mechanics and the quantum forward- and inverse scattering problems, but also in geophysics since by Fourier transformation of the Schrödinger equation one can form a Helmholtz-type equation. Starting with the Schrödinger equation, one can write a one-dimensional inverse scattering equation known as the Marchenko equation, and both exact and inexact solutions have been studied extensively in the literature. Of particular interest is the work of Dorren et al. (1994), who showed that the exact solution to the Marchenko equation is unstable.

Recently, the Marchenko integral equation has become of interest again to geophysicists thanks to two developments. The first is due to Rose (2001) and Aktosun and Rose (2002), who showed two things: that single-sided focusing of the wavefield for the plasma wave equation (which reduces to the Helmholtz equation in free space) is equivalent to single-sided focusing of the time-dependent Schrödinger equation, and that the single-sided focusing solution of the time-dependent Schrödinger equation is the Marchenko equation. By "focusing" we mean that the energy in the wavefield collapses to a spatio-temporal delta-function, whilst "single-sided" refers to being able to achieve focusing while observing the wavefield from only one side of the medium e.g. the surface of the Earth; for real signals the delta-function should be interpreted as band-limited in time. This development formed the basis for Rose (2002) to describe a single-sided "autofocusing" algorithm for the focusing wavefield in a medium with unknown wavespeed by iteratively solving the Marchenko equation. The second development is due to Broggini et al. (2012), who showed that using the Marchenko equation, one can solve for the Green's function between a surface source and a virtual receiver inside a 1D medium using only the single-sided surface reflection data and an estimate of the direct wave from the real source to the virtual receiver. This directly led to the work of Wapenaar et al. (2012, 2013, 2014a,b), who extended the method of Broggini et al. (2012) from 1D to 3D, and which forms the basis for a family of methods of Green's function retrieval, redatuming, and imaging collectively known as Marchenko methods.

Marchenko methods as devised by Wapenaar et al. (2014a) iteratively solve the Marchenko equation similar to Rose (2002) autofocusing scheme, for the wavefield that focuses at depth inside an idealized, lossless medium that is identical to the actual medium except that a homogeneous half-space exists above the depth of the real source or receiver, and a reflection-free half-space exists below the depth of the virtual receiver or source; this wavefield is known as a focusing function. These focusing functions, that exist solely in the idealized medium, are related to the Green's functions in the true medium that in principle contain all orders of scattering, but only require the reflection response in the true medium and an estimate of the direct wave into the medium to determine said focusing functions. There are two focusing

functions: one that focuses at the depth below which the medium is reflection-free, f_1, and one that focuses at the source or receiver array depth, f_2. Assuming that wavefield is pressure-normalized, the two-way Green's functions and focusing functions can be decomposed into an upgoing wavefield denoted with a "-" sign, and a downgoing wavefield denoted with a "+", so that

$$G = G^+ + G^- \tag{2.56}$$

$$f_1 = f_1^+ + f_1^- \tag{2.57}$$

$$f_2 = f_2^+ + f_2^- \tag{2.58}$$

To recover the two-way full Green's function G only requires solving for f_2, whilst recovering the decomposed one-way Green's functions G^+ and G^- also requires solving for f_1. In the work presented in this thesis I will only require the full two-way Green's functions. Details of how to solve for the decomposed Green's functions are described in Wapenaar et al. (2014b) for example.

The two-way Green's function is given by the sum over all receiver positions along the surface ∂D_0 that runs concurrent to the Earth's surface, of the convolution of the reflection response $R\left(\boldsymbol{x}_0'', \boldsymbol{x}_0, t - t'\right)$ recorded at x_0'' due to a source at x_0 with the focusing function $f_2\left(\boldsymbol{x}_i', \boldsymbol{x}_0, t'\right)$ that focuses at x_0 in space and time $t = 0$, evaluated at location x_i' and time t', plus the time-reversed focusing function that focuses at x_0'' and is evaluated at x_i':

$$G\left(\boldsymbol{x}_i', \boldsymbol{x}_0'', t\right) = -\int_{\partial D_0} d\boldsymbol{x}_0 \int_{-\infty}^{t} R\left(\boldsymbol{x}_0'', \boldsymbol{x}_0, t - t'\right) f_2\left(\boldsymbol{x}_i', \boldsymbol{x}_0, t'\right) dt' + f_2\left(\boldsymbol{x}_i', \boldsymbol{x}_0'', -t\right) \tag{2.59}$$

This equation is called a Marchenko equation, so named for the solution to the 1D inverse scattering problem discovered by Marchenko (1955). The upper limit of integration is set at $t' = t$ because the reflection response is causal. Since the Green's function is also causal, the integral must be equal to zero before the arrival of the Green's function's direct wave at time $t = t_D$. Thus

$$0 = -\int_{\partial D_0} d\boldsymbol{x}_0 \int_{-\infty}^{t} R\left(\boldsymbol{x}_0'', \boldsymbol{x}_0, t-t'\right) f_2\left(\boldsymbol{x}_i', \boldsymbol{x}_0, t'\right) dt' + f_2\left(\boldsymbol{x}_i', \boldsymbol{x}_0'', -t\right) \text{ for } t < t_D \quad (2.60)$$

Similar to the 1D case (Lamb, 1980), we assume that the solution to the focusing function f_2 can be decomposed into a direct wave plus a scattering wave coda M for which we will solve the Marchenko equation. It can be shown (Wapenaar et al., 2014b) that the direct arrival of the focusing function f_2 is equal to the direct arrival of the inverse of the transmission response , so that

$$f_2\left(\boldsymbol{x}_i', \boldsymbol{x}_0'', t\right) = \mathcal{T}_d^{-1}\left(\boldsymbol{x}_i', \boldsymbol{x}_0'', t\right) + M\left(\boldsymbol{x}_i', \boldsymbol{x}_0'', t\right) \quad (2.61)$$

where the inverse of the transmission response $\mathcal{T}$, the component of the wavefield that is **not** back-scattered (i.e. reflected), is defined in the time domain as the upgoing part of the focusing function $f_2\left(\boldsymbol{x}_i', \boldsymbol{x}_0'', t\right)$:

$$f_2^{-}\left(\boldsymbol{x}_i', \boldsymbol{x}_0'', t\right) = \mathcal{T}^{-1}\left(\boldsymbol{x}_i', \boldsymbol{x}_0'', t\right)$$

For a wavefield that focuses at time equal to zero, the direct wave of the focusing function arrives at -t_D, and since the coda follows the direct wave it also stands that

$$M\left(\boldsymbol{x}_i', \boldsymbol{x}_0'', t\right) = 0 \text{ for } t < -t_D \quad (2.62)$$

The iterative Marchenko algorithm for M is obtained by substituting equation 2.61 into equation 2.60 and applying equation 2.62 so that

$$M_k\left(\boldsymbol{x}_i', \boldsymbol{x}_0'', -t\right) = M_0\left(\boldsymbol{x}_i', \boldsymbol{x}_0'', t\right) - \int_{\partial D_0} d\boldsymbol{x}_0 \int_{-t_D^{\epsilon}}^{t} R\left(\boldsymbol{x}_0'', \boldsymbol{x}_0, t-t'\right) M_{k-1}\left(\boldsymbol{x}_i', \boldsymbol{x}_0, t'\right) dt' \quad (2.63)$$

where

$$M_0\left(\boldsymbol{x}_i', \boldsymbol{x}_0'', t\right) = -\int_{\partial D_0} d\boldsymbol{x}_0 \int_{-\infty}^{t_D^{\epsilon}} R\left(\boldsymbol{x}_0'', \boldsymbol{x}_0, t-t'\right) \mathcal{T}_d^{-1}\left(\boldsymbol{x}_i', \boldsymbol{x}_0, t'\right) dt' \quad (2.64)$$

and $t_D^\epsilon = t_D + \epsilon$, where ϵ is a small positive constant that ensures the direct arrival is entirely within the integral for M_0. When the direct arrival of the inverse transmission response is estimated to be the time-reversed direct arrival of the Green's function, the iterative scheme described by equations 2.63 and 2.64 can be evaluated with only the reflection data and a smooth macro-model of the true medium.

The algorithm I describe above has several important assumptions not already mentioned. The first is that the reflection response does not contain free-surface multiples. A Marchenko algorithm that can account for free-surface multiples was devised by Singh et al. (2017), however all the results I present do not contain free-surface reflections. The second is that the full treatment by Wapenaar et al. (2014a) requires source-receiver reciprocity at several steps, which implicitly assumes that the source and receiver arrays are co-located. Such an experimental design is impractical for off-shore streamer surveys, an issue that was resolved by Ravasi (2017). I have imposed Wapenaar's experimental design on all my experiments. Finally, because it is necessary to decompose the wavefield into up-going and down-going components, it is assumed that the focusing functions are free of evanescent waves which cannot be described as either up- or down-going, and consequently the focusing functions should be interpreted as collapsing to a delta-function band-limited in both time and space. This means that the aperture of the source-receiver array is assumed to be infinite! Consequently, when implementing Marchenko methods it may be necessary to taper the geometric edges of the array to prevent artefacts due to finite aperture (e.g. Lomas and Curtis, 2019).

2.4 Marchenko-Lippmann-Schwinger Inversion

In section 2.2 we discussed a number of methods for solving the inverse scattering problem, all of which required us at some point to assume that the data are linear in the model, that is, that the single-scattering or Born approximation holds. Amongst the primary reasons for this assumption is that we cannot access the Green's functions in the true medium $G(\cdot, x)$ between our surface-based source & receiver arrays, and a location inside the medium x; ordinarily, one would argue this is because we do

not know the true medium, and if we did know the true medium, we would have no need to solve the inverse problem! Thus the interdependency of linearized data and inadequate models that replicate the linearized data prevails in geophysics.

Previously this assumption was justified because no method existed to access the Green's functions necessary to solve the Lippmann-Schwinger equation. However, as shown in section 2.3, the emergence of Marchenko methods now allows us to determine these same Green's functions in a data-driven manner that does not require access to the true medium. I make the argument that given the ability to recover the necessary Green's functions, the proper procedure should be to invert the full Lippmann-Schwinger equation and not the Born modelling equation. What is more, inspection of the Lippmann-Schwinger equation (equation 2.16) with the understanding that now $G(\cdot, x)$ is a known variable reveals that the Lippmann-Schwinger equation is now linear with respect to the model perturbation! Thus it should be, and is, possible to design methods based on the linearized Born modelling equation to invert the Lippmann-Schwinger equation. The remaining part of this chapter is dedicated to explaining two such methods for solving this inverse problem. Given that a Marchenko method will be required to solve the Lippmann-Schwinger equation, they are all appropriately named Marchenko-Lippmann-Schwinger methods.

2.4.1 Marchenko-Lippmann-Schwinger Least-Squares Inversion

Since the Lippmann-Schwinger equation is linearized with knowledge of the Green's functions $G(x, s)$, the simplest approach is a one-step inversion by solving the inverse of the kernel operator of the Lippmann-Schwinger equation. The forward problem can be written in the form of equation 2.24 so that

$$G_S = \omega^2 \left(G_0 G\right) V \tag{2.65}$$

It then follows directly from section 2.2 that the direct inverse problem is written

$$\omega^2 \hat{V} = \left(G_0 G\right)^{-g} G_S \tag{2.66}$$

Similar to equation 2.38, the model that minimises the error function of the Lippmann-Schwinger least-squares problem is

$$\omega^2 \hat{V}_{LS} = \{(G_0 G)^T (G_0 G)\}^{-1} (G_0 G)^T G_S \tag{2.67}$$

whilst the regularised least-squares solution is given by

$$\omega^2 \hat{V}_{reg} = \{(G_0 G)^T (G_0 G) + \mu^2 I\}^{-1} (G_0 G)^T G_S \tag{2.68}$$

where μ^2 is the regularisation parameter, which is again positive by definition to maintain the positivity of the error function.

If the inverse Born problem is under-determined, then the comparable Marchenko-Lippmann-Schwinger inverse problem is also under-determined, since replacing G_0 with G does not change the number of data-points (which is determined by the number of source-receiver pairs and the number of different frequencies[1]) and thus does not increase the number of unique equations, but rather changes the solution to those equations by including higher orders of scattering. The question is then whether the additional information from the higher orders of scattering is sufficient to constrain the ill-posedness of the inverse scattering problem, and if so, which solution most accurately reproduces the true medium?

2.4.2 Adjoint-State Methods of Marchenko-Lippmann-Schwinger Inversion

The direct inversion methods just described are suited to small models with a limited-number of source-receiver pairs, since just as with the Born inversion described in section 2.2, the cost of calculating the inverse of a square matrix of order M scales as $\mathcal{O}(M^3)$. For larger problems where it is impractical to calculate the inverse matrix, or problems where the instability of the inversion cannot be adequately constrained by regularisation, we still need alternative methods such as adjoint-state methods. If we are to apply adjoint-state methods to the inversion of the full Lippmann-Schwinger

[1] If the problem is written in the time domain, each individual time sample rather than each individual frequency constitutes a data point.

equation, we will need to be able to calculate the Fréchet derivatives of the Lippmann-Schwinger equation. I will now demonstrate how to calculate the first Fréchet derivative of the Lippmann-Schwinger equation, which has a compact form suitable for use in imaging and inversion. Then I will derive a new scheme that uses this derivative, Marchenko-Lippmann-Schwinger Full-Waveform Inversion, which can be implemented as is its Born approximation counterpart.

2.4.2.1 Derivatives of the Lippmann-Schwinger Equation

We begin by writing again the Lippmann-Schwinger equation:

$$G_S(\boldsymbol{r},\boldsymbol{s}) = \omega^2 \int_V \delta m(\boldsymbol{x})\, G_0(\boldsymbol{r},\boldsymbol{x})\, G(\boldsymbol{x},\boldsymbol{s})\, d\boldsymbol{x} \tag{2.69}$$

where $\boldsymbol{r}$ is the vector of receiver locations and $\boldsymbol{s}$ is the vector of source locations. I reserve all vectors $\boldsymbol{x}$, $\boldsymbol{x}'$ etc. to denote coordinates inside the model. Our goal is to determine the derivative of the Green's function between source and receiver $G(\boldsymbol{r},\boldsymbol{s})$ with respect to the model parameters m. Inspecting the Lippmann-Schwinger equation, we note the variables dependent on the model are $G(\boldsymbol{r},\boldsymbol{s})$, $G(\boldsymbol{x},\boldsymbol{s})$, and $\delta m(\boldsymbol{x})$. The background model Green's functions G_0 are independent of the model $\boldsymbol{m}$. Indeed, the background model $\boldsymbol{m}_0$ is not required to represent the true medium in any meaningful way, although typically it will be determined by one or more data-driven methods. The Lippmann-Schwinger equation describes the relationship between the Green's functions in the media $\boldsymbol{m}$ and $\boldsymbol{m}_0$, when the difference between the models is small compared to the magnitude of the model values, and then we impose the interpretation that any such model $\boldsymbol{m}_0$ adequately describes the low wavenumber components of the true medium $\boldsymbol{m}$. As a consequence, the derivative of the Green's functions G is solely dependent on the scattering term under the integral sign. This does not mean that only the high wavenumber components of the model can be recovered. Indeed, the Born equation is a linearisation of the Lippmann-Schwinger equation with a more restrictive range of applicability, and the Born equation framework is routinely used to resolve the low wavenumber components of velocity models. So clearly the Lippmann-Schwinger equation must also be able to resolve the low

wavenumbers. Rather, the appeal of the Lippmann-Schwinger equation is that it encapsulates the higher wavenumber information contained in short- and mid-offset highly-scattering reflection data that is not well described by low wavenumber models.

Taking the derivative with respect to the model parameters $\boldsymbol{m}$ gives us that

$$\frac{\partial G(\boldsymbol{r},\boldsymbol{s})}{\partial \boldsymbol{m}(\boldsymbol{y})} = \omega^2 \Bigg\{ \int \frac{\partial \left(\delta \boldsymbol{m}(\boldsymbol{x})\right)}{\partial \boldsymbol{m}(\boldsymbol{y})} G_0\left(\boldsymbol{r},\boldsymbol{x}\right) G\left(\boldsymbol{x},\boldsymbol{s}\right) d\boldsymbol{x} + \int \delta \boldsymbol{m}(\boldsymbol{x}) G_0\left(\boldsymbol{r},\boldsymbol{x}\right) \frac{\partial G\left(\boldsymbol{x},\boldsymbol{s}\right)}{\partial \boldsymbol{m}(\boldsymbol{y})} d\boldsymbol{x} \Bigg\} \quad (2.70)$$

where the coordinates $\boldsymbol{y}$ run concurrent to $\boldsymbol{x}$. Starting with the first integral term, if the model parameters m_i are independent of each other, then the derivative of the model perturbation reduces to the Kronecker delta δ_{ij}. This means that

$$\frac{\partial G(\boldsymbol{r},\boldsymbol{s})}{\partial \boldsymbol{m}(\boldsymbol{y})} = \omega^2 \Bigg\{ G_0\left(\boldsymbol{r},\boldsymbol{x}\right) G\left(\boldsymbol{x},\boldsymbol{s}\right) + \int \delta \boldsymbol{m}(\boldsymbol{x}) G_0\left(\boldsymbol{r},\boldsymbol{x}\right) \frac{\partial G\left(\boldsymbol{x},\boldsymbol{s}\right)}{\partial \boldsymbol{m}(\boldsymbol{y})} d\boldsymbol{x} \Bigg\} \quad (2.71)$$

To resolve the second integral term, we exploit the fact that the Lippmann-Schwinger equation is iterative. Substituting the Lippmann-Schwinger equation into the second integral term of equation 2.71, we find that

$$\begin{aligned} \frac{\partial G(\boldsymbol{r},\boldsymbol{s})}{\partial \boldsymbol{m}(\boldsymbol{y})} =& \omega^2 \Bigg\{ G_0\left(\boldsymbol{r},\boldsymbol{x}\right) G\left(\boldsymbol{x},\boldsymbol{s}\right) + \int \delta \boldsymbol{m}(\boldsymbol{x}) G_0\left(\boldsymbol{r},\boldsymbol{x}\right) \frac{\partial G_0\left(\boldsymbol{x},\boldsymbol{s}\right)}{\partial \boldsymbol{m}(\boldsymbol{y})} d\boldsymbol{x} \\ &+ \omega^2 \int \delta \boldsymbol{m}(\boldsymbol{x}) G_0\left(\boldsymbol{r},\boldsymbol{x}\right) d\boldsymbol{x} \int \frac{\partial \left(\delta \boldsymbol{m}(\boldsymbol{x}')\right)}{\partial \boldsymbol{m}(\boldsymbol{y})} G_0\left(\boldsymbol{x},\boldsymbol{x}'\right) G\left(\boldsymbol{x}',\boldsymbol{s}\right) d\boldsymbol{x}' \\ &+ \omega^2 \int \delta \boldsymbol{m}(\boldsymbol{x}) G_0\left(\boldsymbol{r},\boldsymbol{x}\right) d\boldsymbol{x} \int \delta \boldsymbol{m}(\boldsymbol{x}') G_0\left(\boldsymbol{x},\boldsymbol{x}'\right) \frac{\partial G\left(\boldsymbol{x}',\boldsymbol{s}\right)}{\partial \boldsymbol{m}(\boldsymbol{y})} d\boldsymbol{x}' \Bigg\} \end{aligned} \quad (2.72)$$

As previously, the independence of the background Green's functions G_0 from the true medium $\boldsymbol{m}$ means the second term on the R.H.S. of equation 2.72 is equal to zero; and the independence of the model parameters from each other reduces the derivative in the third term on the R.H.S. to the Kronecker delta. So after one

iteration we find

$$\begin{aligned}\frac{\partial G(\boldsymbol{r},\boldsymbol{s})}{\partial \boldsymbol{m}(\boldsymbol{y})} =&\omega^2\Big\{G_0\left(\boldsymbol{r},\boldsymbol{x}\right)G\left(\boldsymbol{x},\boldsymbol{s}\right)+\omega^2\int\delta\boldsymbol{m}(\boldsymbol{x})G_0\left(\boldsymbol{r},\boldsymbol{x}\right)G_0\left(\boldsymbol{x},\boldsymbol{x}'\right)G\left(\boldsymbol{x}',\boldsymbol{s}\right)d\boldsymbol{x}\\&+\omega^2\int\delta\boldsymbol{m}(\boldsymbol{x})G_0\left(\boldsymbol{r},\boldsymbol{x}\right)d\boldsymbol{x}\int\delta\boldsymbol{m}(\boldsymbol{x}')G_0\left(\boldsymbol{x},\boldsymbol{x}'\right)\frac{\partial G\left(\boldsymbol{x}',\boldsymbol{s}\right)}{\partial\boldsymbol{m}(\boldsymbol{y})}d\boldsymbol{x}'\Big\}\end{aligned} \tag{2.73}$$

This result contains another derivative of a full Green's function $G(x',s)$, if we iterate the Lippmann-Schwinger equation again then

$$\begin{aligned}\frac{\partial G(\boldsymbol{r},\boldsymbol{s})}{\partial \boldsymbol{m}(\boldsymbol{y})} =&\omega^2\Big\{G_0\left(\boldsymbol{r},\boldsymbol{x}\right)G\left(\boldsymbol{x},\boldsymbol{s}\right)+\omega^2\int\delta\boldsymbol{m}(\boldsymbol{x})G_0\left(\boldsymbol{r},\boldsymbol{x}\right)G_0\left(\boldsymbol{x},\boldsymbol{x}'\right)G\left(\boldsymbol{x}',\boldsymbol{s}\right)d\boldsymbol{x}\\&+\omega^4\int\delta\boldsymbol{m}(\boldsymbol{x})G_0\left(\boldsymbol{r},\boldsymbol{x}\right)d\boldsymbol{x}\int\delta\boldsymbol{m}(\boldsymbol{x}')G_0\left(\boldsymbol{x},\boldsymbol{x}'\right)G_0\left(\boldsymbol{x}',\boldsymbol{x}''\right)G\left(\boldsymbol{x}'',\boldsymbol{s}\right)d\boldsymbol{x}'\\&+\omega^4\int\delta\boldsymbol{m}(\boldsymbol{x})G_0\left(\boldsymbol{r},\boldsymbol{x}\right)d\boldsymbol{x}\int\delta\boldsymbol{m}(\boldsymbol{x}')G_0\left(\boldsymbol{x},\boldsymbol{x}'\right)d\boldsymbol{x}'\ldots\\&\ldots\int\delta\boldsymbol{m}(\boldsymbol{x}'')G_0\left(\boldsymbol{x}',\boldsymbol{x}''\right)\frac{\partial G\left(\boldsymbol{x}'',\boldsymbol{s}\right)}{\partial\boldsymbol{m}(\boldsymbol{y})}d\boldsymbol{x}''\Big\}\end{aligned} \tag{2.74}$$

It is clear that iterating the Lippmann-Schwinger equation *ad infinitum* will result in an infinite series:

$$\begin{aligned}\frac{\partial G(\boldsymbol{r},\boldsymbol{s})}{\partial \boldsymbol{m}(\boldsymbol{y})} =\omega^2\sum_{n=0}^{\infty}\delta G_n(\boldsymbol{r},\boldsymbol{x}^{(n)})G\left(\boldsymbol{x}^{(n)},\boldsymbol{s}\right)\\+\omega^2\int\delta\boldsymbol{m}(\boldsymbol{x})\delta G_\infty(\boldsymbol{r},\boldsymbol{x}^{(\infty)})\frac{\partial G(\boldsymbol{x}^{(\infty)},\boldsymbol{s})}{\partial\boldsymbol{m}(\boldsymbol{y})}d\boldsymbol{x}\end{aligned} \tag{2.75}$$

where

$$\begin{aligned}\delta G_0(\boldsymbol{r},\boldsymbol{x}^{(0)}) &= G_0\left(\boldsymbol{r},\boldsymbol{x}^0\right)\\ \delta G_1(\boldsymbol{r},\boldsymbol{x}^{(1)}) &= \omega^2\int G_0\left(\boldsymbol{r},\boldsymbol{x}^0\right)\delta\boldsymbol{m}(\boldsymbol{x}^0)G_0\left(\boldsymbol{x}^0,\boldsymbol{x}^1\right)d\boldsymbol{x}^0\\ \delta G_2(\boldsymbol{r},\boldsymbol{x}^{(2)}) &= \omega^4\int G_0\left(\boldsymbol{r},\boldsymbol{x}^0\right)\delta\boldsymbol{m}(\boldsymbol{x}^0)d\boldsymbol{x}^0\int G_0\left(\boldsymbol{x}^0,\boldsymbol{x}^1\right)\delta\boldsymbol{m}(\boldsymbol{x}^1)G_0\left(\boldsymbol{x}^1,\boldsymbol{x}^2\right)d\boldsymbol{x}^1\end{aligned}$$

etc.

We now make two observations: first, that after infinite orders of scattering, the concept of further scattering is meaningless, hence the integral term on the R.H.S.

of equation 2.75 must be equal to zero; and second, that the infinite sum over the terms δG is the sum over all scattering orders including the non-scattering wavefield G_0, which from the Lippmann-Schwinger equation is equal to the Green's function G. Thus, with some relabelling of coordinates, we find that the derivative of the Lippmann-Schwinger equation with respect to the model parameters $\boldsymbol{m}$ is

$$\frac{\partial G(\boldsymbol{r},\boldsymbol{s})}{\partial \boldsymbol{m}(\boldsymbol{x})} = \omega^2 G(\boldsymbol{r},\boldsymbol{x}) G(\boldsymbol{x},\boldsymbol{s}) \tag{2.76}$$

When the Green's functions on the R.H.S. of equation 2.76 are replaced by the background model Green's functions, equation 2.76 becomes the familiar equation for the derivative of the Born modelling equation.

2.4.2.2 Marchenko-Lippmann-Schwinger Full-Waveform Inversion

My claim is that inversion schemes based on the Born linearization of the Lippmann-Schwinger equation can be compatible with higher-order scattering data through the linearization of the Lippmann-Schwinger equation using Green's function calculated with a Marchenko method. The aim is to formulate an inverse problem so that the higher-order scattering is correctly accounted for in the calculation of the linear terms in the model, rather than using higher-order scattering to calculate the terms that are non-linear in the model. As a consequence, the scheme I propose is implemented almost identically to the standard Born implementation.

We again wish to determine the model which minimises the least-squares error function between observed data and modelled data, except that we now separate the data into a reflections-only component $\boldsymbol{d}^{refl}$, and a component containing all other events $\boldsymbol{d}^{div}$ (where "div" stands for diving waves) including direct waves and refractions, i.e.

$$\boldsymbol{d} = \boldsymbol{d}^{refl} + \boldsymbol{d}^{div} \tag{2.77}$$

If we assume that the reflections are separated from the other recorded arrivals in the time-series such that there is no cross-talk between reflections and any other types of events, then the least-squares error function between observed data and modelled

data can be written as

$$C(\boldsymbol{m}) = \frac{1}{2}\{\|\Delta \boldsymbol{d}^{refl}(\boldsymbol{m})\|^2 + \|\Delta \boldsymbol{d}^{div}(\boldsymbol{m})\|^2\} \tag{2.78a}$$

$$= \frac{1}{2}\{\|\boldsymbol{d}^{refl}_{obs} - \boldsymbol{d}^{refl}_{cal}(\boldsymbol{m})\|^2 + \|\boldsymbol{d}^{div}_{obs} - \boldsymbol{d}^{div}_{cal}(\boldsymbol{m})\|^2\} \tag{2.78b}$$

where again "obs" and "cal" denote the observed and calculated (or modelled) data respectively, of which only the latter is model-dependent. As in the standard FWI, the data (both $\boldsymbol{d}^{refl}$ and $\boldsymbol{d}^{div}$) are evaluated in the reference model $\boldsymbol{m}_0$. Following the same procedure as in section 2.2.1, the derivative of the cost function with respect to the model parameters is

$$\frac{\partial C(\boldsymbol{m})}{\partial \boldsymbol{m}} = -\left\{ \Re\left[\frac{\partial \left(\boldsymbol{d}^{refl}_{cal}(\boldsymbol{m})\right)^\dagger}{\partial \boldsymbol{m}} \Delta \boldsymbol{d}^{refl} \right] + \Re\left[\frac{\partial \left(\boldsymbol{d}^{div}_{cal}(\boldsymbol{m})\right)^\dagger}{\partial \boldsymbol{m}} \Delta \boldsymbol{d}^{div} \right] \right\} \tag{2.79}$$

The Lippmann-Schwinger equation says that scattering is a frequency-dependent phenomenon, scaling with the square of the angular frequency. Thus we expect the Born approximation to be valid only in the low-frequency part of the data where scattering is weakest, with the full Lippmann-Schwinger equation required to accurately model the higher-frequency data components. Conveniently, the majority of arrivals in the seismic data can be divided into two complementary regimes, with the diving waves (see Appendix B for an explanation of diving waves and refractions) prevalent at long offsets occupying the low-frequency range whilst the reflections exist in the high-frequency range and dominate the short- and mid-offsets in the data. This means that the reflected wave derivative $\partial \boldsymbol{d}^{refl}_{cal}(\boldsymbol{m})/\partial \boldsymbol{m}$ is given by the derivative of the Lippmann-Schwinger equation (equation 2.79), whilst the standard Born treatment can be applied to the diving wave derivative $\partial \boldsymbol{d}^{div}_{cal}(\boldsymbol{m})/\partial \boldsymbol{m}$. This gives the derivative of the cost function as

$$\frac{\partial C(\boldsymbol{m})}{\partial \boldsymbol{m}} = -\Re\Bigg\{ \left[\omega^2 \left(G(g, x_i) G(x_j, s) \right)^\dagger \Delta \boldsymbol{d}^{refl} \right] + \left[\omega^2 \left(G_0(g, x_i) G_0(x_j, s) \right)^\dagger \Delta \boldsymbol{d}^{div} \right] \Bigg\} \tag{2.80}$$

The Green's functions from the surface source and receiver arrays into the medium

are calculated by a Marchenko method, whilst the reference model Green's functions are modelled with a finite-difference wave-equation solver, for example.

This algorithm, which I call Marchenko-Lippmann-Schwinger Full-Waveform Inversion, can be summarised as follows:

1. Record data in the true Earth $\boldsymbol{d}_{obs}$ as well as its source function
2. Do forward modelling in the reference model $\boldsymbol{m}_0$ to record data $\boldsymbol{d}_{cal}(\boldsymbol{m}_0)$
3. Separate observed and modelled data into the reflection data $\boldsymbol{d}^{refl}$ plus the diving waves $\boldsymbol{d}^{div}$

   While cost function has not converged:

   (a) Calculate data residuals $\Delta\boldsymbol{d}^{refl} \equiv \boldsymbol{d}^{refl}_{obs} - \boldsymbol{d}^{refl}_{cal}$ and $\Delta\boldsymbol{d}^{div} \equiv \boldsymbol{d}^{div}_{obs} - \boldsymbol{d}^{div}_{cal}$

   (b) Back-propagate diving wave data residual from receiver locations

   (c) Convolve reflection wave data residual with adjoint Green's functions obtained by a Marchenko method.

   (d) Cross-correlate forward- and back-propagated wavefields to get $\partial C(\boldsymbol{m})/\partial \boldsymbol{m}$

   (e) Line-search to obtain optimal step length α

   (f) Update reference model according to $\boldsymbol{m}_k = \boldsymbol{m}_{k-1} - \alpha \frac{\partial C(\boldsymbol{m})}{\partial (\boldsymbol{m})}$

   (g) Do forward modelling in the reference model $\boldsymbol{m}_k$ to record data $\boldsymbol{d}_{cal}(\boldsymbol{m}_k)$
4. Accept reference model as solution

Chapter 3

Experimental Design & Computational Requirements for Implementation of Marchenko-Lippmann-Schwinger Methods

In this chapter I describe the experimental design that I used in each of my experiments, explaining the physical and computational factors that influenced my choices. I describe the workflow to calculate the Marchenko Green's functions for use in Marchenko-Lippmann-Schwinger methods and discuss how computational resource limitations may impact the applicability of such methods.

3.1 Experimental Design

3.1.1 P-wave velocity and density models

The velocity model I have used to represent the true Earth, shown in figure 3.1, is designed to represent a geological syncline infilled with several sedimentary layers. The syncline is over- and underlain by lower velocity media. The velocity in the

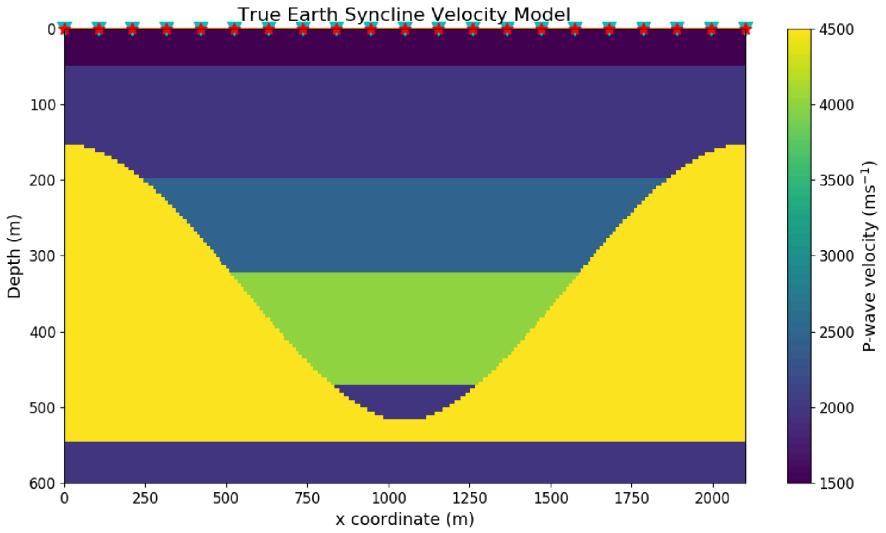


Figure 3.1: True Earth P-wave velocity model. Every 7th source (red star) and receiver (cyan triangle) of the surface arrays are shown.

uppermost layer is set at 1500ms^{-1} and is intended to be analogous to the water layer in the marine seismic environment. As a starting model for my experiments, I use the smooth model shown in figure 3.2. This model has been created from the true Earth model by passing a Gaussian smoothing window across the model in both x- and z-directions starting at a depth of 50m, it is assumed that the structure in the uppermost 50m representing a water layer is known exactly a priori. Since the physics of the problem I am testing assumes constant density, the density model used for the acoustic wavefield modelling is homogeneous with a uniform density of 1000 kg/m^3 – the approximate density of (fresh) water. As the density is homogeneous, the exact value is unimportant since it scales the wavefield amplitude uniformly across space. The model size is 600m deep x 2100m across with a pixel size of 5m x 5m, giving an aspect ratio of 1:3.5 in both physical and model parameter space.

3.1.1.1 Spatial and temporal sampling in the model

The choice of pixel size is a compromise between the need to prevent numerical instability and numerical dispersion (note: not the physical dispersion of seismic waves when the phase velocity does not equal the group velocity) whilst also modelling efficiently with respect to both time and computational resources, and the need to have

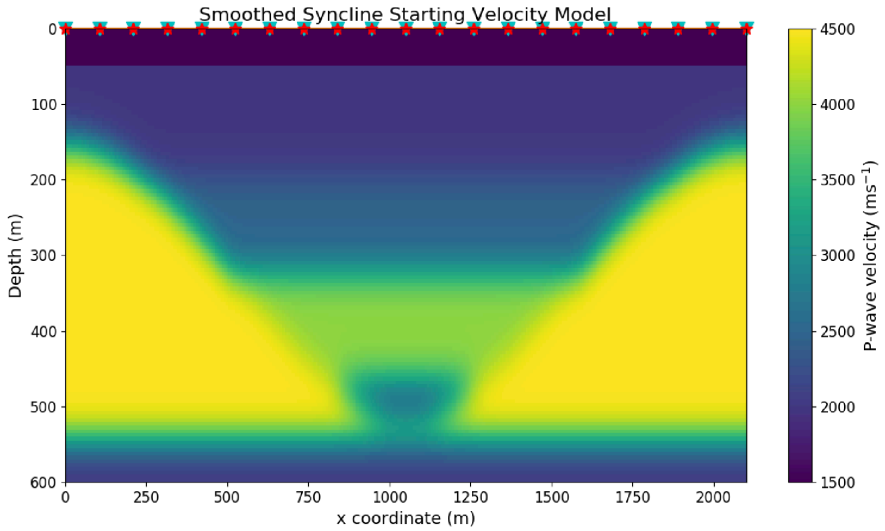


Figure 3.2: Smoothed P-wave starting velocity model.

an appropriate spatial resolution for the structures the model describes. The relationship between model pixel size, time sampling interval in the wavefield, and numerical stability of a finite-difference approximation is given by the Courant-Friedrich-Lewy condition (Courant et al., 1967):

$$\frac{c\Delta t}{\Delta x} < C \tag{3.1}$$

where c is the wavespeed, Δt the time sampling interval, Δx the pixel size, and C is the dimensionless Courant number describing how unstable the finite-difference solution is – the lower the value of C, the more stable the finite-difference solution. For a 4th-order finite-difference stencil, the Courant number is 0.606 (Sei, 1995). Noting that the distance travelled by the wavefront in a given time interval is greatest where the p-wave velocity is greatest, we can make the finite-difference approximation globally stable by requiring that the time sampling interval and pixel size satisfy the following inequality:

$$\Delta t < \frac{0.606}{c_{max}}\Delta x \tag{3.2}$$

where c_{max} is the maximum wavespeed in the model.

Finite-difference approximations are inherently dispersive; as a rule of thumb, it is generally considered that to reduce numerical dispersion, the wavefield should be

sampled in space at least 5 times per wavelength. For a given frequency f the wavelength is minimised where the p-wave velocity is slowest, thus we can define the spatial sampling limit to globally minimise numerical dispersion at a given frequency:

$$\Delta x < \frac{c_{min}}{5f} \tag{3.3}$$

where c_{min} is the minimum wavespeed in the model. Combining equations 3.2 and 3.3 for the maximum frequency we wish to model f_{max}, we can define an upper limit on the time sampling interval that is both numerically stable and minimises numerical dispersion for the full range of frequencies we expect to model; the result is that

$$\Delta t < \frac{0.606}{5f_{max}} \frac{c_{min}}{c_{max}} \tag{3.4}$$

With c_{min} = 1500ms^{-1}, c_{max} = 4500ms^{-1}, and maximum modelling frequency f_{max} = 60Hz, this gives an upper limit on Δt_{max} = 0.00067s. Based on this limit, I set Δt= 0.0005s. Putting this time sampling interval into equation 3.2 gives a lower limit on the pixel size Δx_{min} = 3.3m; based on this limit I set Δx = 5m.

An alternative to finite-difference modelling of the pressure wavefield is to use a pseudospectral method, whereby the spatial derivatives of the wave equation are calculated in the wavenumber domain. When implemented using the Fast Fourier Transform (FFT), pseudospectral methods can be highly accurate and computationally efficient (Kosloff and Baysal, 1982). When the rapid expansion method (REM) for calculating time integrals of the wave equation is used in combination with pseudospectral methods (Kosloff et al., 1989), the results can be highly accurate and numerically stable whilst requiring fewer calculations than the equivalent finite-difference method that is equally accurate (Pestana and Stoffa, 2010). In this way, pseudospectral methods may avoid the numerical instabilities of finite-difference methods caused by large time-steps.

3.1.1.2 Model design

I chose a syncline structure for several reasons; firstly, sedimentary synclines are seen in a number of geological settings, thus it provides a synthetic analogue to a potential

real-world experiment in the future. Secondly, the symmetry of the model about the z-axis from the centre of the x-axis aids in quality control for both seismic modelling and retrieved Marchenko Green's functions – differences between Green's functions from mirror image model points may indicate errors in a number of modelling parameters, including source and/or receiver array locations, the defined wavefield boundary, and model point indexing as examples. Thirdly, the varying gradient of the walls of the syncline mean that the problem cannot be reduced to a series of one-dimensional inversions using only zero-offset source-receiver pairs, which conversely means that any successful methods will need to be able to account for two-dimensional structures, as would be required in many real-world geological settings. Smoothly-varying interfaces have the advantage of being relatively straightforward to interpret in the scattered wavefield as compared to scattering from a rugose interface. Also, Marchenko methods in the form used herein do not work for diffractions, this is for the same reason that evanescent waves cannot be recovered - diffractors scatter energy in all directions, meaning energy is lost that cannot be separated into either up-going or down-going wavefields. These features mean that synclinal structures have been used successfully in a number of studies on Marchenko methods, for example; Wapenaar et al. (2014b), Meles et al. (2016), van der Neut and Wapenaar (2016), and Singh et al. (2017).

A characteristic property of the Earth is the prevalence of strong seismic reflectors in the subsurface, often resulting in strong multiple reflections in observed seismic data. Strong reflections occur where large contrasts in acoustic impedance occur due to changes in velocity, density, or both. In early Marchenko studies, it was common to use a complex density model and a homogeneous p-wave velocity model as analogues of the true Earth to generate strong multiple reflections whilst also simplifying the problem. The advantage of this approach is that one can have direct waves for use in the Marchenko algorithm that have exactly correct travel-times at all locations. By judicious choice of the uniform p-wavespeed, the direct arrival and the scattered wave coda can be made to be well-separated in time at almost all locations – the locations where this is most difficult to enforce are closest to the scattering interfaces. This is important because the direct arrival defines the windowing function used in the Marchenko algorithm, if this window is poorly defined

e.g. there are reflections inside the window, the focusing function will be incorrect and the Marchenko scheme may not converge on the actual Green's function G. Combined with a smooth estimate of the true complex density structure, the phase of wavelet will also be correct except when close to density interfaces in the true Earth, and so the iterative Marchenko scheme predicts waveforms at the exact arrival times they are expected to arrive at with correct wavelet phases, with increasing iterations correcting the amplitudes of the scattered waveforms. The disadvantage of such a model is that it is patently geologically absurd, and for this reason such models are no longer considered appropriate for most studies. Unlike most previous Marchenko studies, my aim is not to obtain an acoustic impedance image but to do an inversion for the model parameters themselves. The additional complexity of inversion is the reason for using simpler physics that assumes the Earth has constant density. The disadvantage of this assumption is that density variations cannot be used to increase the amplitudes of the reflected wavefield, including multiple reflections that can be predicted by Marchenko methods.

The absence of density contrasts requires strong velocity contrasts in the model to generate stronger multiple reflections. The problems with this approach are two-fold: firstly, increasing the downward reflection coefficient needed to encourage multiple-scattering decreases the upward transmission coefficient, thereby reducing the recorded amplitude of multiply-scattered events recorded at the surface; and secondly, increasing velocity contrasts in the true model increases the size of the region around velocity interfaces where the kinematics of the smooth model are incorrect. As a result, the direct wave arrival times in these regions is incorrect, in turn leading to an error in the predictions made by the Marchenko algorithm. This means that we expect a systematic error in the kernel of the inverse problem when using Green's functions obtained by a Marchenko method in the presence of strong velocity contrasts. The kinematics of the true velocity model can be preserved if smoothing is performed on the slowness model instead of the velocity model. The issue then is that no smooth velocity model derived from real data is ever likely to have the exact kinematics of the true velocity model, and so the synthetic experiment again does not represent the problem when using real-world data.

3.1.2 Wavelet design

The wavelet used in my experiments, shown figure 3.3, is the derivative of a Gaussian function with a peak frequency of 15Hz. As a function of frequency f, this wavelet has the form

$$S_1(f) = \frac{f}{\sqrt{2}f_p} \exp\left(- \frac{f^2}{2f_p^2} \right) \tag{3.5}$$

for a peak frequency f_p. The amplitude spectrum of my wavelet, also shown figure 3.3, has been cut-off to zero amplitude at a minimum frequency of 0Hz and a maximum frequency of 60Hz. The length of the wavelet in time is set to 3.1s at a sampling interval of 0.0005s (or sampling rate of 2000Hz), which equates to 3s of wavefield recording from the anti-symmetry point of the wavelet. The additional 0.1s allows the peak of the wavelet to be shifted so that the entire waveform occurs at times $t > 0$ i.e. the wavelet is causal. Since real seismic sources are causal, it is important to that our synthetic experiments verify our methods are valid for causal signals. A more typical choice of wavelet for seismic wavefield modelling is the Ricker wavelet (also figure 3.3), which as a zero-phase wavelet with distinctive sidelobes and a prominent central amplitude peak, is desirable for the relative ease of interpreting waveform arrivals in the data. As a function of frequency, the Ricker wavelet can be written as

$$S_2(f) = \frac{f^2}{f_p^2} \exp\left(- \frac{f^2}{f_p^2} \right) \tag{3.6}$$

The Ricker wavelet is the second derivative of a Gaussian function, making my wavelet choice the integral of a Ricker function with a peak frequency of 15Hz.

The advantage of using the first Gaussian derivative over the second Gaussian derivative is that for the same parameters (peak frequency and time sampling interval), the first Gaussian derivative has a broader frequency spectrum than its second derivative counterpart. The advantage of the Ricker wavelet's zero-phase property over the anti-symmetric form of my wavelet is diminished when modelling in two-dimensions, since the two-dimensional analytic seismic Green's function incurs a $\pi/4$ phase rotation compared to the three-dimensional analytic Green's function (Berkhout, 1987) (see Appendix C for an explanation). For either wavelet then, the resulting waveform

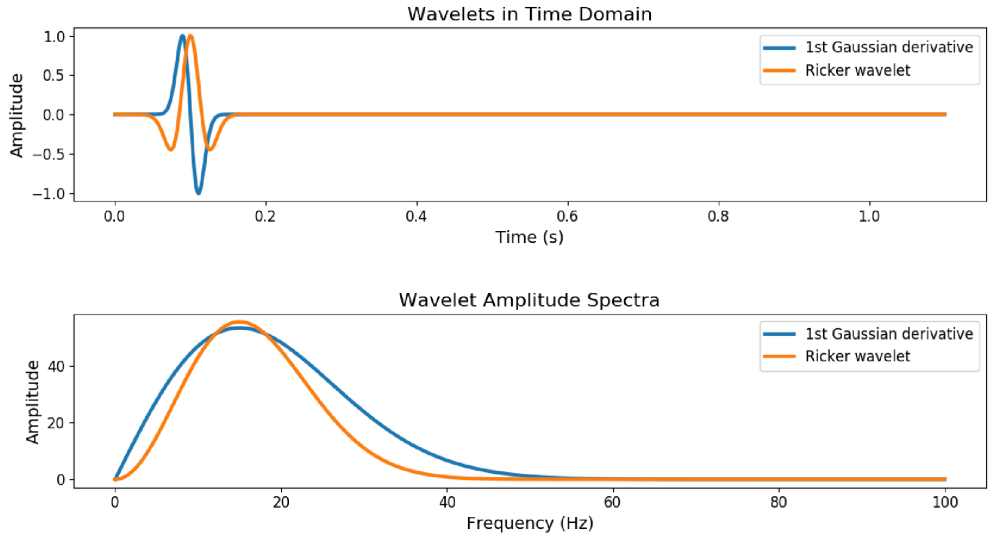


Figure 3.3: Comparison of 1st derivative of a Gaussian wavelet and the Ricker wavelet at peak frequency = 15Hz. Top: Wavelets in the time domain; and bottom: their (frequency) amplitude spectra.

observed in the data will be asymmetric in time.

3.1.3 Wavelet design for Marchenko methods

Without involving ourselves in the details of how Marchenko methods work here (for more information, see section 2.3), the Green's function solution between surface and subsurface locations is found by an iterative series of convolutions between the reflection response recorded at the surface due to the physical experiment and a modelled function called a "focusing function". If the focusing function was modelled with the source wavelet of the physical experiment, the result would not be the Green's function but the Green's function convolved with the source wavelet many times over. In a real data scenario, the procedure would be to record the source signature in the field and deconvolve the recorded reflection response of this wavelet; in practice, deconvolution is difficult to achieve well due to the zero division problem, and is largely avoided unless absolutely necessary.

In synthetic seismic experiments, the same result can be achieved by modelling the reflection response using a wavelet with a flat frequency response across the non-zero bandwidth of the physical experiment source signature; in this instance, the Gaussian

derivative wavelet with peak frequency at 15Hz. Such a wavelet can be created as one would make a boxcar filter with cosine-tapered edges. Mathematically, the flat spectrum wavelet can be written

$$S_{FW}(f) = \begin{cases} 0 & f < f_{min} \\ \left(\cos\left(\pi \frac{f - f_{min}}{f_{min} - f_{left}} \right) + 1 \right) / 2 & f_{min} \leq f < f_{left} \\ 1 & f_{left} \leq f < f_{right} \\ \left(\cos\left(\pi \frac{f - f_{right}}{f_{max} - f_{right}} \right) + 1 \right) / 2 & f_{right} \leq f \leq f_{max} \\ 0 & f > f_{max} \end{cases} \tag{3.7}$$

where f_{min} is the minimum cut-off frequency, f_{left} is the lower tapering off frequency, f_{right} is the upper tapering off frequency, and f_{max} is the maximum cut-off frequency. In the time domain, this results in a sinc function. For my experiments, I used the settings in table 3.1 to define a flat spectrum wavelet.

Setting	Frequency (Hz)
f_{min}	0
f_{left}	2
f_{right}	55
f_{max}	65

Table 3.1: Wavelet settings for the flat spectrum wavelet in figure 3.4

The resulting wavelet and its amplitude spectrum are shown in figure 3.4. Because Marchenko-Lippmann-Schwinger full-waveform inversion requires the backpropagation of the data-residual using the full Green's functions $G^{\dagger}(x, r)$ (equation 2.80), it was decided to create an additional Marchenko Green's function with the flat wavelet as its source signature instead of removing the wavelet signature from the Marchenko Green's function by deconvolution. The amplitude spectrum of the autoconvolution of the flat spectrum wavelet is also shown to demonstrate that the spectrum is almost identical to that of the flat spectrum wavelet, and thus we can expect the wavefield produced by convolution of the Marchenko Green's functions with the data-residuals at the receivers to faithfully reproduce the wavefield one might produce by modelling

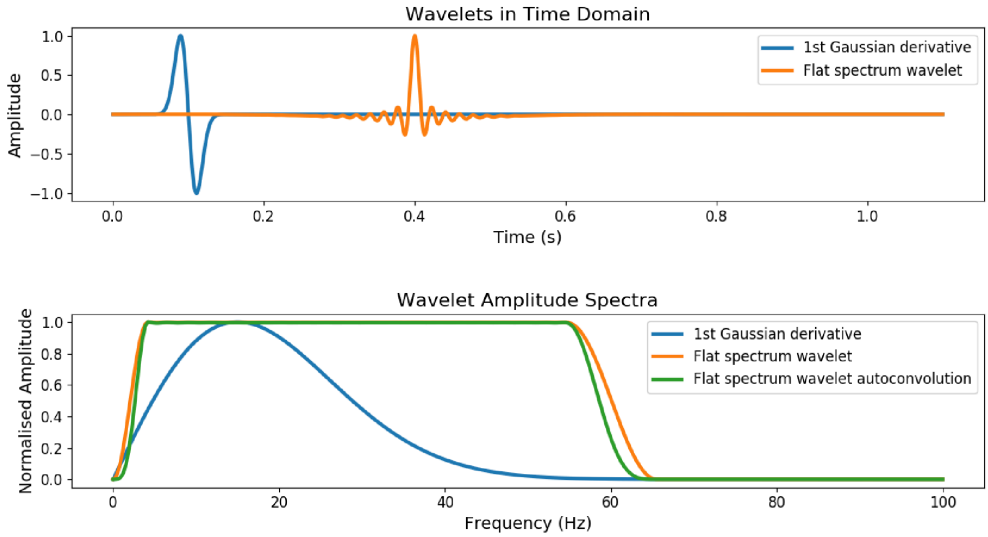


Figure 3.4: Comparison of the 1st derivative of a Gaussian wavelet and the flat spectrum wavelets. Top: The Gaussian 1st derivative and normalized amplitude flat spectrum wavelets in the time domain. Bottom: the normalized amplitude spectra of the wavelets compared against the amplitude spectrum of the autoconvolution of the flat spectrum wavelet.

with finite differences using the same source function. Errors may only be expected for the very highest and lowest frequencies, which in general also exhibit the lowest amplitudes in the spectrum.

3.1.4 Source and receiver arrays

In my experiments I use an array of 141 co-located sources and receivers placed at the surface of the model and spaced at 15m intervals across the entire 2100m width of the model, this corresponds to placing a source and receiver at every 3rd grid point on the finite-difference grid along the surface. This configuration of sources and receivers is symmetrical about the centre of the x-axis, which when combined with the symmetry of the model structure implies an expected symmetry in the results of my experiments that should simplify interpretation of the results. The sources and receivers being co-located is a requirement of the Marchenko algorithm of Wapenaar et al. (2014b).

The number and spacing of the sources and receivers was chosen as a compromise between having adequate data sampling and the computational and resource require-

ments of calculating Marchenko Green's functions for a greater number of sources. The most efficient means to modelling the necessary direct waves from every model point to every real source location is to model the wavefields created by every real source at every real source location at the surface to every point in the medium; and then re-order the wavefield datacubes into many shot gathers, with a virtual source at each model point. The wavefield for a given source at any point in the medium is a seismic data trace identical to the trace that would be recorded if a receiver were placed at that location. Using source-receiver reciprocity of the wavefield and the co-located source-receiver geometry imposed by the Marchenko method, combining the traces from each single-source wavefield at said point in the medium into a single gather is identical to the gather that would be recorded if a source were placed at the point inside the medium and recorded by receivers at our source locations. The simplest method to do this requires having twice the free memory as the combined size of all the modelled wavefields. Assuming that all data are stored as 32-bit floats, the size of each wavefield in memory with no inefficiency for the data structure recorded for 3s across the entire model space is 1.14GB, meaning that even for a relatively small experiment such as this the free memory space requirement rapidly approaches an order of 1TB once all sources have been considered.

The simplest choice would be to place sources and receivers at each model point along the surface, this would give 421 co-located sources and receivers. Lomas and Curtis (2019) found that as a rule of thumb, the source & receiver spacing should be less than $1/4$ of the wavelength at the peak frequency at the receiver in order to obtain an acceptable result from a Marchenko scheme. For a surface velocity of 1500ms^{-1} the observed wavelength at the peak frequency of 15Hz would be 100m. This gives an approximate upper limit on the source-receiver spacing of 25m or every 5th grid point i.e. a five-fold reduction in the number of required sources. However, when a test shot was modelled at this grid spacing the data was found to be spatially aliased. The maximum source-receiver spacing that did not spatially alias the data was found to be 15m or every 3rd model point, resulting in 141 co-located sources and receivers. This gives a three-fold reduction in the number of modelled wavefields compared to the simplest design.

To further reduce the amount of memory required to store the wavefields, I also subsampled the wavefield in time 8-fold, increasing the time sampling interval to 0.004s. The choice to subsample 8-fold was made so that the maximum modelled frequency of 60Hz would remain well below the Nyquist frequency to prevent temporal aliasing of the data.

3.2 Seismic modelling and Marchenko routines

3.2.1 Choice of programs

To model all the seismic data and wavefields used in my experiments, I used the "`fdelmodc`" package developed by Jan Thorbecke. The `fdelmodc` package is an open source 2D & 3D acoustic and visco-elastic finite-difference wavefield modelling program that has been peer reviewed in Thorbecke and Draganov (2011), and has previously been and continues to be used in research into Marchenko methods, notably at Delft University of Technology. This makes it an obvious choice for research using Marchenko methods.

The program I used to implement the Marchenko method of Wapenaar et al. (2014b) is the "`marchenko`" program, also by Jan Thorbecke. This program has also been peer reviewed in Thorbecke et al. (2017) and is also open source, allowing the results of that paper and other results from the research group at Delft University of Technology to be reproduced by anyone as examples included in the package. To avoid confusion, whenever "`marchenko`" is written in the lower case and in the `typewriter` typeface, it should be understood that I am referring to the software program that implements the Marchenko algorithm.

The key benefit of using these two programs in combination is that the `marchenko` program was written to use data produced by the `fdelmodc` program, which minimises the amount of data manipulation and processing of the wavefield data needed to apply the Marchenko algorithm. This both improves the efficiency with which data may be sent between the programs and eliminates potential errors in reformatting data from another program that may be stored in an unsuitable format for the `marchenko`

program to read. Both programs read and write data in the Seismic Unix or "SU" format (Cohen and Stockwell, 2017), an open source seismic processing suite that employs a combined data & header data format to store seismic data together with its metadata. This data format is similar to the Society of Exploration Geophysicists (SEG) industry standard formats for seismic data, and the later Regularly Sampled Format (RSF) open source format by Fomel et al. (2013). Whilst such data formats may have built-in redundancies in their header field options that are liable to user error, the `fdelmodc` and `marchenko` programs are written to require or recognise only a small number of the total number of possible header fields. This reduces the likelihood of finding or creating data with incorrect data headers. Seismic Unix allows for existing data stored in native formats to be converted to the SU format, and the SU format is also sufficiently widespread that tools to handle data in this format exist for scientific programming languages such as Python & MATLAB®.

As an alternative to finite-difference wavefield modelling, it is possible to model the direct waves needed for the Marchenko method by using an Eikonal equation solver to generate a table of traveltimes from each source location to every point in the model. Using the traveltime to each model point, one can create a time series with a single spike of unit amplitude at the time sample corresponding to the traveltime to that location in the already determined traveltime table, and convolve that time series with the source wavelet to create a waveform arrival at the correct time. The advantage of this method is that it is computationally fast compared to solving the wave equation by finite-differences, since the Eikonal equation is only a 1st-order differential equation compared to the wave equation which is 2nd-order. This allows algorithms such as the fast-marching method (Sethian, 1996; Sethian and Popovici, 1999) to be applied to the Eikonal equation to find the travel times. The disadvantage, however, is that traveltime-based methods for generating direct waves are less accurate than finite-difference methods can be. Whilst a geometric spreading correction can be applied to the resulting direct waves, and a $\pi/4$ phase rotation can be applied to account for the 2D physics of the problem (see Appendix C), traveltime-based methods cannot reliably account for changes in amplitude and phase in the waveform as the wavefront propagates through regions of different wavespeed and density. Since my experiments

are synthetic tests to prove or disprove the feasibility of the methods being tested, it makes sense to use the most accurate modelling possible so that modelling error can be discounted when interpreting the results of the studies.

3.2.2 Workflow for obtaining Marchenko Green's functions

In this subsection, I describe the workflow for generating Marchenko Green's functions starting with no data. I will assume that one is modelling by finite-difference methods only, rather than using an Eikonal solver to model direct waves.

To begin, one starts by modelling data in the true Earth model using a suitable flat spectrum wavelet for all source and receiver locations to obtain the full reflection response including the direct waves between sources and receivers. By modelling data in a homogeneous velocity model with velocity equal to the surface layer velocity in the true Earth for all source and receiver locations and subtracting from the previously modelled data in the true Earth, in the ideal case one is left with data only containing reflections including all the multiple reflections. By using a flat spectrum wavelet, these data are the band-limited reflection Green's functions. When working with real data, this step would be replaced by post-acquisition processing to remove direct waves and refractions from the data before deconvolution with the recorded source wavelet. Since the Marchenko method I used does not take surface-related multiples (e.g. seabed multiple reflections) into account, I modelled the data without a free surface; in the real data case, one would apply surface-related multiple elimination (SRME) (Verschuur et al., 1992) or some other multiple-removal method before deconvolution (assuming deconvolution and wavelet estimation have not already happened as part of the seismic data-processing workflow).

From this point onwards, the procedure is the same for both synthetic datasets and real datasets. Using the (recorded) source wavelet, the wavefield in the smooth starting model from every source location is recorded at every model point. Since there are generally far fewer sources than model points, it is significantly more efficient to model fewer wavefields and rearrange the real wavefields into virtual shot gathers from sources placed in the subsurface to real receivers at the surface than it is to

model a real source at each point in the subsurface and only record data at the locations of the real receivers on the surface. In synthetic experiments, one may obtain more accurate direct waves by modelling in the true Earth model, especially as the kinematics of the smooth model maybe be very different to the true model close to interfaces in the model. This will result in an additional processing step later to window out any refractions before the direct wave or arrivals after the direct wave.

The next step is to rearrange the wavefield datacubes into shot gathers from sources in the subsurface. This step requires loading the direct wavefields into memory to create a hypercube of size [# sources, # x coordinates, # z coordinates, # time samples], indexing over all x and z coordinates in the model, and writing shot gathers of size [# sources, # time samples] at each coordinate. Since the sources and receivers are co-located in our experiment, by source-receiver reciprocity of Green's functions the source number for each wavefield becomes the receiver number in each shot gather.

Wavefields can be read into MATLAB® from SU format directly by reading the files as native binaries and passing over the headers. The MATLAB® parallel processing toolbox allows for efficient use of memory when looping over coordinate indices, in this way I was able to achieve 10x parallelisation. Writing out the shot gathers is the slowest part of the Marchenko process; as the size of the gathers increases, the time to reload the gather, append additional traces, and re-write the gather to disk increases. For this reason, as many wavefields as possible should be simultaneously loaded into memory to reduce the computational time required writing gathers.

At this point, if the direct waves shot gathers were obtained from the true Earth model then the appropriate filtering should be applied now. This could be done using any seismic processing package, however the `marchenko` program is packaged with another function "`fmute`" for the purpose of muting above, below, or around the direct wave; a data file containing the mute window may be provided to the function, otherwise the function will determine the first arrival itself by default. This is done by assuming the highest magnitude arrival in the zero-offset trace is also the peak of the first arrival, from this point a trace-by-trace search away from this datum for the highest amplitude within a specified time-window is used to follow the first arrival

trajectory through the data gather. Thus it is implicitly assumed that the velocity model is not dipping, particularly in the near-surface. As the `fmute` function reads data in SU format, the shot gathers obtained before will need to have SU headers with the correct source location and receiver locations attached to the binaries beforehand. If alternate processing software was used to filter the direct waves or the direct wave gathers were already ready for use, correct SU headers will need to be attached to the binaries before proceeding. Conceptually, we should now have twice as much data in storage as the original modelled wavefields, since we have copied traces from one file to another in a different order. In reality, we are likely using significantly more memory due to the additional header files on each direct wave shot gather.

We are now ready to use the `marchenko` program. As this program can only process a single model point at a time i.e. a single direct wave shot gather, it requires relatively little memory, and has a short run-time of seconds to minutes depending on the number of iterations of the Marchenko algorithm, the number of traces, and the length of those traces in time. Consequently, this process is well-suited to running in parallel on large clusters. After we have obtained our Marchenko GFs, we now have 3x our initially conceived data stored to disk. If the Marchenko GFs are required in the form of wavefields, as they are for my Marchenko-Lippmann-Schwinger Full-Waveform Inversion method, then the reverse of the procedure to create shot gathers from single-source wavefields can be used to reshape Marchenko shot gathers into pseudo-wavefields (pseudo in the sense that such a wavefield never actually existed!) In this case, we would have at least 4x the amount of data in storage as we originally modelled (the original smooth model wavefields, the direct-wave gathers, the Marchenko Green's function gathers, and the Marchenko GF pseudo-wavefields) – I highlight this to emphasise the need to be actively managing the amount of data created when using big data methods such as these, and to allow others to factor in such considerations if reproducing this work. This workflow is summarised in figure 3.5.

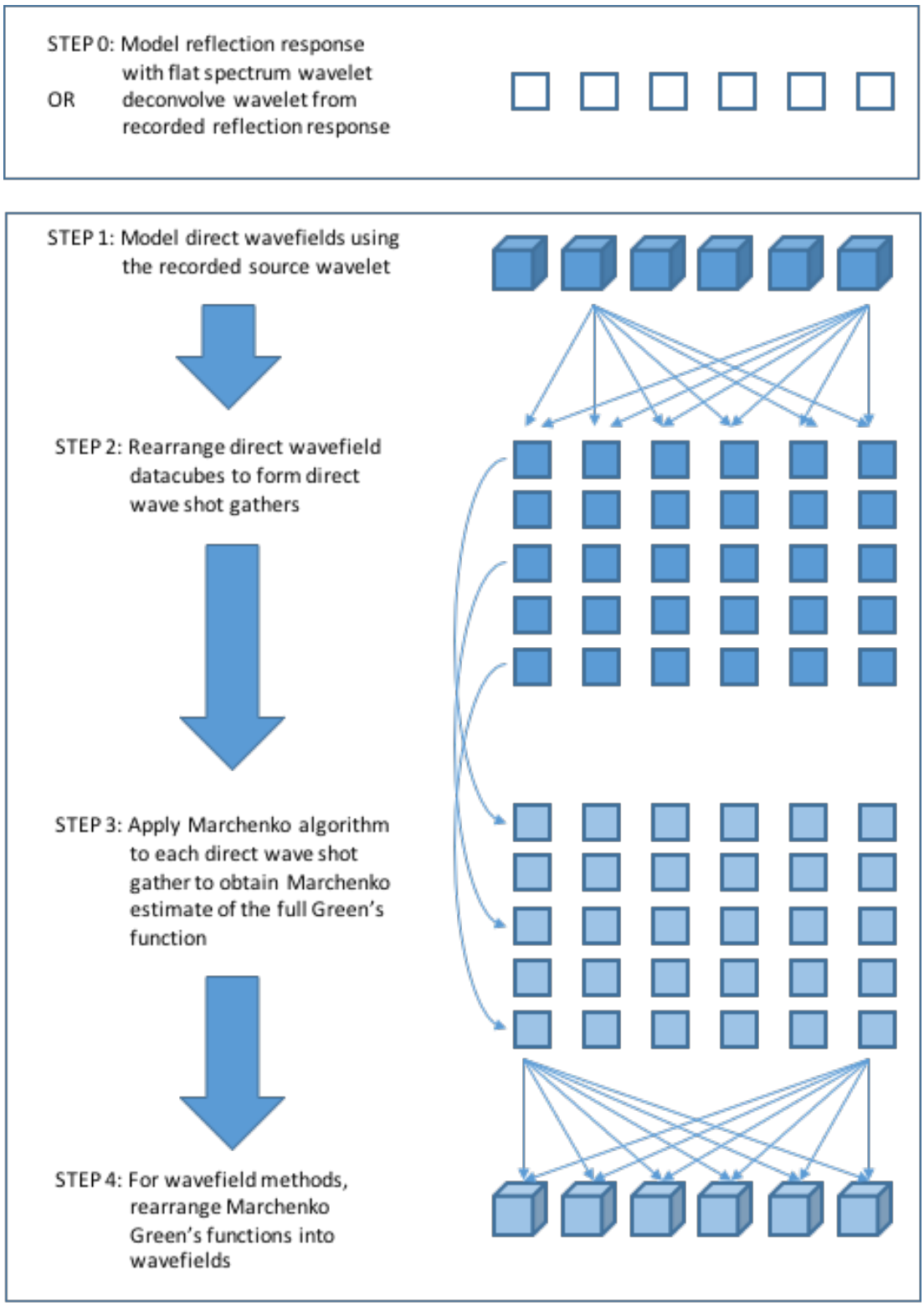


Figure 3.5: A pictorial representation of how data is transformed during the workflow of producing Marchenko Green's functions and subsequently wavefields.

3.3 Summary

In this chapter, I explain the experimental design used to test Marchenko-Lippmann-Schwinger methods. I begin with why a symmetric syncline structure is a good choice for a seismic wavefield experiment, then use computational limits imposed by the use of a finite-difference solver to optimise the parameters governing the size of the model and the time sampling rate used. I explain my choice of wavelet, and my choice of source and receiver arrays, which are controlled by the requirements of the Marchenko algorithm used and the need to manage the amount of data created and stored to disk. I then explain my choices for a finite-difference wavefield solver and a Marchenko algorithm routine, and why the combination chosen is superior to other options that may be faster to implement. I finish by describing the workflow from having no data to having wavefields of Green's functions obtained from a Marchenko method.

Chapter 4

Benchmarking Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger-Inversions in One Dimension

The Lippmann-Schwinger equation is the canonical equation of acoustic scattering. It relates the observed scattered wavefield to the medium parameters of the Earth through the scattering wavefield in the true medium and the non-scattering wavefield in the background medium. With only a smooth representation of the Earth, we cannot know the scattering wavefield. Thus the observed scattered wavefield as described by the Lippmann-Schwinger equation is a non-linear function of the medium parameters.

Non-linear problems are challenging to solve because of the unpredictability of the relationship between the model and the data, and the solution is generally highly dependent on the starting model. The accepted way to circumnavigate the non-linearity of the Lippmann-Schwinger equation is to linearize the problem, by replacing the scattering wavefield in the kernel with the background wavefield, so that only the medium parameters remain unknown. This results in an ill-posed linear problem called the Born scattering problem that relates the observed singly-scattered wavefield to the medium parameters. This compromise fundamentally changes the physics of

the system and thus the solution we should expect from the inverse problem. The Earth is a highly scattering medium, so the solution must create sufficiently strong scattering that it reproduces our observations of single scattering events. In assuming the Earth is a singly-scattering medium, the same solution must also be sufficiently weakly scattering that no further reflections are observed. There can be no guarantee that this "Goldilocks solution" is an accurate representation of the true medium, or that such a solution even exists.

Since the development of Marchenko methods, it has been possible to access the previously unknown scattering wavefield without knowing the true medium parameters. Consequently, the Lippmann-Schwinger equation is a linear function of the medium parameters. This means that we can evaluate whether the alternate physics of the Born equation changes the solution of the inverse problem, by comparing the Born inverse solution to the solution of the linear Lippmann-Schwinger inverse problem.

In this chapter I present the results of a linear inversion of the Lippmann-Schwinger equation. I begin by reproducing the results of Beydoun and Tarantola (1988) who solved the 1D inverse Born problem using analytical Green's functions and exact representations of the inverse problem for a boxcar perturbation, as a benchmark to compare the solution to the inverse Born problem using finite-differences and matrix inverses. I then compare the benchmarked solution of the Born inverse problem to the solutions to Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inverse problems using matrix methods.

4.1 Benchmarking of one-dimensional inversion of finite-difference Green's functions against an analytic result

The 1D inverse Born problem for a boxcar perturbation to a homogeneous slowness-squared model was solved by Beydoun and Tarantola (1988), figure 4.1 shows the definitions of the true (perturbed) and background (unperturbed) media in terms of

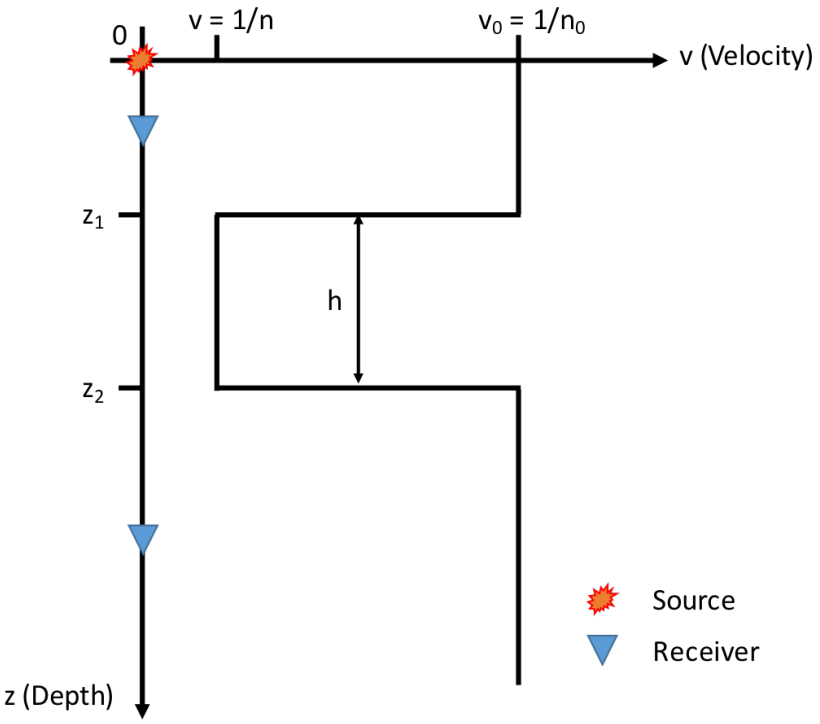


Figure 4.1: The 1D slab model. Adapted from Beydoun and Tarantola (1988).

velocity. The background medium has constant slowness $n_0 = 1/v_0$, while the true medium is perturbed between depths z_1 and z_2 with constant slowness $n = 1/v$, where $h = z_2 - z_1$ is the thickness of the perturbation. The source is located at depth $z = 0$, with a single receiver at depths $0 \leq z_g \leq z_1$ or $z_g \geq z_2$ above or below the perturbation. Using these definitions, the problem is defined by the following equations:

$$\Delta u_B(z) = \frac{2n_0^2}{\pi} \int_{-\infty}^{+\infty} e^{-4\pi i\omega n_0(z-z_g)} \frac{\Delta P(z_g,\omega)}{2\pi i\omega} d\omega$$

(Eq. A16 (Beydoun and Tarantola, 1988)) (4.1)

$$\Delta P(z_g,\omega) = \frac{n-n_0}{2n_0} e^{-4\pi i\omega n_0 z_g} \left(-e^{4\pi i\omega n_0 z_1} + e^{4\pi i\omega n_0(z_2+(n-n_0)h)}\right) + \mathcal{O}([\delta n]^2)$$

(Eq. A19 (Beydoun and Tarantola, 1988)) (4.2)

where $\Delta u_B = n^2 - n_0^2$ is the perturbation to the slowness-squared model under the Born approximation, and $\Delta P(z_g, \omega)$ is the scattered pressure wavefield with frequency

ω observed at receiver depth z_g.

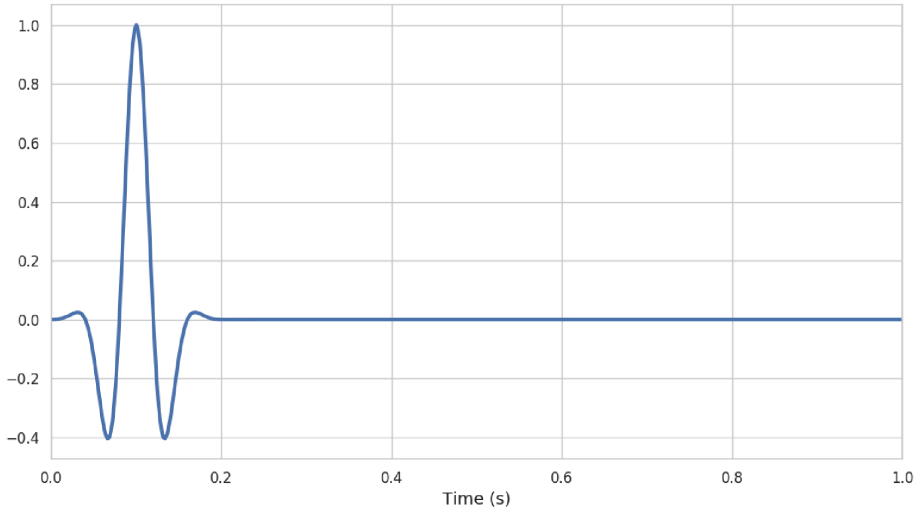
To solve the inverse problem for the slowness-squared perturbation defined at a given depth z, one first solves the forward problem for the scattered wavefield given by equation 4.2 for a given frequency. The forward modelling process is repeated for all frequencies over a theoretically infinite bandwidth (in practice, a finite-bandwidth with as many frequency bins as is reasonably possible) and substituted into equation 4.1. The forward modelling and inversion process is then repeated for every depth in the model. To obtain the updated velocity model, the slowness-squared perturbations from equation 4.1 are added to the homogeneous background slowness model, then the reciprocal of the square root of the sum gives the final velocity model.

As written, equations 4.1 and 4.2 are for an infinite frequency bandwidth source; however for their example a Gabor wavelet (Červený, 1985) is used. The Gabor wavelet is defined as

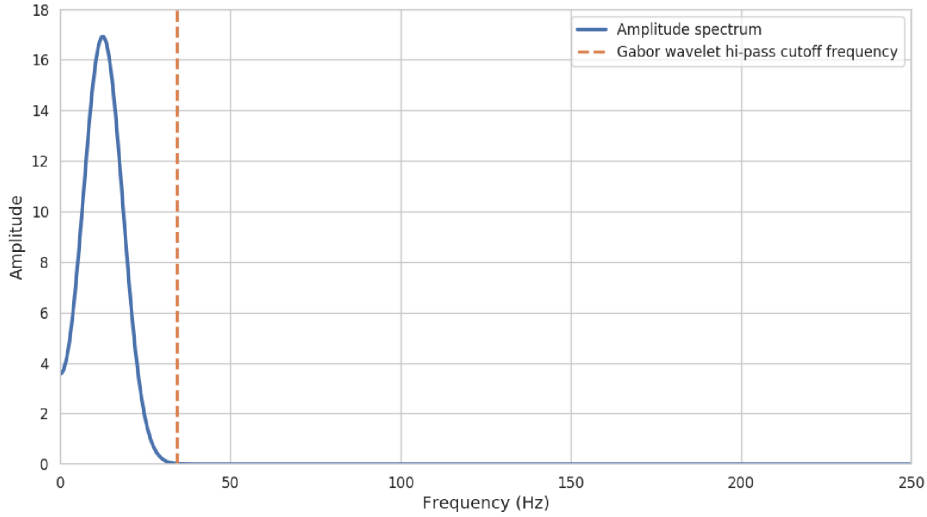
$$S_{Gabor}(t) = e^{-[2\pi f_M(t-t_0)/\gamma]^2}\cos(2\pi f_M(t-t_0)+\nu) \tag{4.3}$$

where f_M is the dominant frequency, γ the half-width parameter, t_0 the impulse time, and ν the initial phase of the wavelet. Here, $f_M = 12.5$Hz, $\gamma = 3$, $t_0 = 0$s, and $\nu = 0$. Figure 4.2 shows this wavelet in the time domain and its amplitude spectrum; unlike other Gaussian wavelets (e.g. the Ricker wavelet), the Gabor wavelet has a significant amplitude at zero frequency, however this energy cannot be used in the inversion since the inversion (equation 4.1) is inversely proportional to the frequency.

Figure 4.3 shows the true and background models used by Beydoun and Tarantola (1988), the model is 3.5km deep with a 1km deep inclusion ($h = 1$km) at a depth of 1km ($z_1 = 1$km). Since those authors do not specify the spatial and temporal sampling used in their analysis, I have chosen $dx = 0.01$km and $dt = 0.002$s with $nt = 8001$ samples since this satisfies the Courant numerical stability condition (equation 3.1) and numerical dispersion requirements, and simplifies the discretization of the model. Figures 4.4a and 4.4b show the Born inversion results for the model in figure 4.3 by Beydoun and Tarantola (1988) and myself respectively for a source and receiver co-located at the surface and using the wavelet specified by the previous authors



(a) Time domain representation



(b) Amplitude spectrum

Figure 4.2: The Gabor wavelet with $f_M = 12.5$Hz, $\gamma = 3$, $t_0 = 0.1$s, and $\nu = 0$.

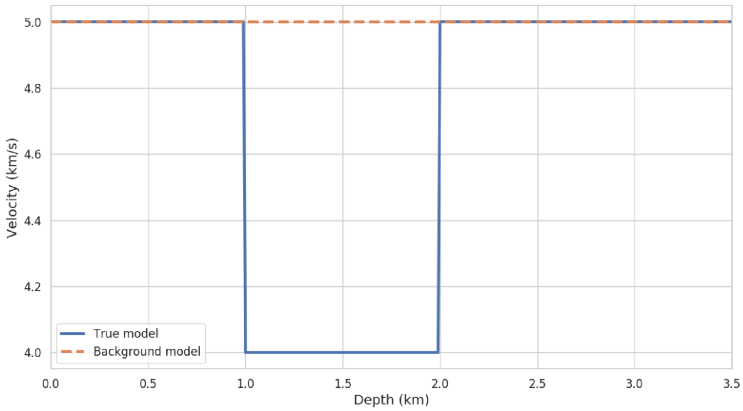


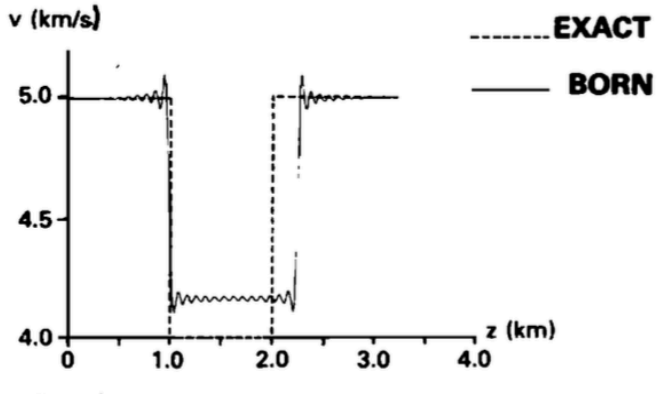
Figure 4.3: True and initial models for 1D Born inversion

(figure 4.2 and equation 4.3). Figure 4.4c shows the same inversion using a white spectrum bandlimited to the frequency bandwidth of the 12.5Hz Gabor wavelet.

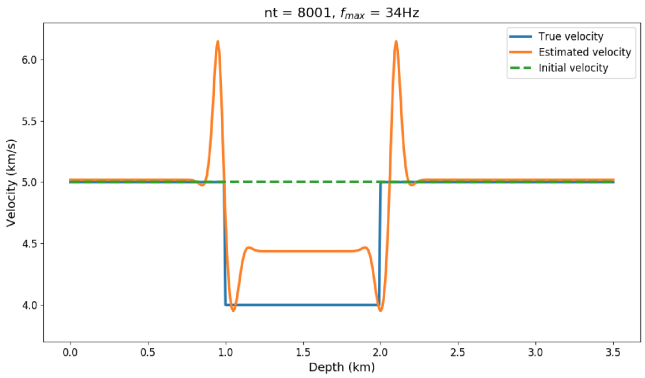
From these figures it is clear that to reproduce the results of the original authors, it is necessary to use a white spectrum in the inversion . When a wavelet is applied to the inversion kernel of equation 4.1 it is imprinted onto the inversion result. Figures 4.4b and 4.4c both show a DC shift in the inversion results compared to the original authors' results.

Figure 4.4d shows the Born inversion of figure 4.3 using the full frequency bandwidth permitted by the chosen time sampling, giving a maximum frequency $f_{max} \approx 250$Hz. This shows that model fit improves with wider frequency bandwidth. By comparison to figure 4.4c however it appears that extremely high frequencies are not necessary to invert for horizontal structures, rather the main benefit of higher frequency data is in reducing the Gibbs phenomenon ("ringing") of the solution around these structures.

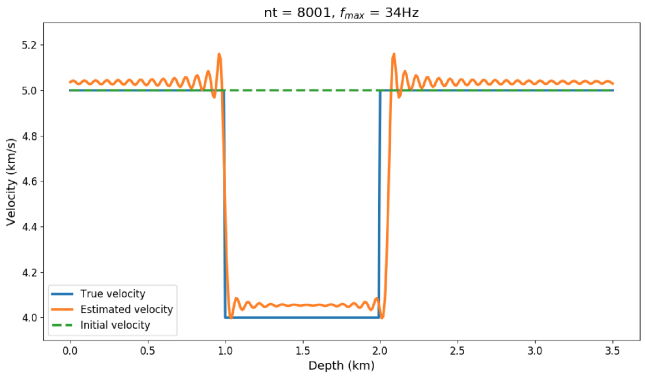
Figure 4.5 shows the effect of the number of time samples in the forward modelled wavefields (equation 4.2) on the analytical Born inversion result. It is evident that increasing the number of time samples reduces the magnitude of the DC shift on the inversion result. Since the forward and inverse problems are defined in the frequency domain and not the time domain, it is not the dependency on the number of time samples that is being tested. Rather, because the temporal sampling rate is constant, the Nyquist frequency is fixed while the number of frequency bins increases, which



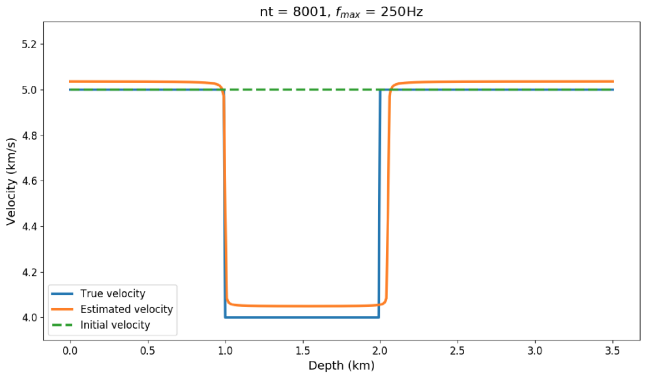
(a) Reproduced from Beydoun and Tarantola (1988)



(b) With 12.5Hz Gabor wavelet



(c) With white spectrum in bandwidth of 12.5Hz Gabor wavelet



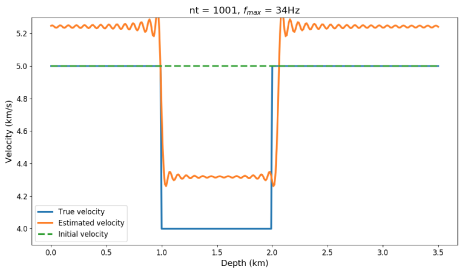
(d) With white spectrum up to Nyquist frequency

Figure 4.4: 1D Born inversion of Beydoun and Tarantola (1988) slab model.

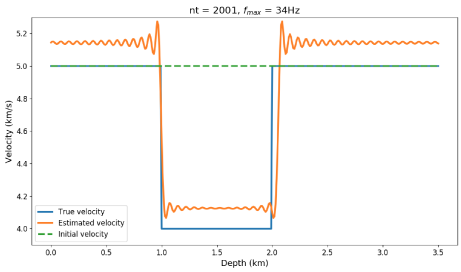
in turn increases the frequency-sampling density. Thus it is more accurate to say increasing the frequency-sampling density decreases the magnitude of the DC shift.

Figure 4.6 shows the comparison of the reflection data as would be recorded in the model to reflection data recorded in the updated model post-Born inversion, modelled using 1D finite-differences and the 12.5Hz Gabor wavelet, and windowed up to 2s. Comparison is not provided against the initial model since constant velocity models are not reflective so the corresponding scattered-wave data would simply be a set of zeros. The four traces shown are from the inversions using 2001 time samples (a & b), and using 32001 time samples (c & d); where the inversion models for (a) and (c) used flat spectrum Green's functions including all frequencies up to the Nyquist frequency, and (b) and (d) used Green's functions modelled with the 12.5Hz Gabor wavelet. Circled red on all traces is the artefact caused by the co-location of source and receiver - since the amplitude of the wavefield is related to the velocity of the medium, the small difference in the velocities of the initial and recovered velocity model results in a difference of amplitudes in the wavefield at the time of source-injection, which is instantaneously recorded to the source-receiver co-location; this artefact is not present in the true data since the true model is equal to the initial model at the source location.

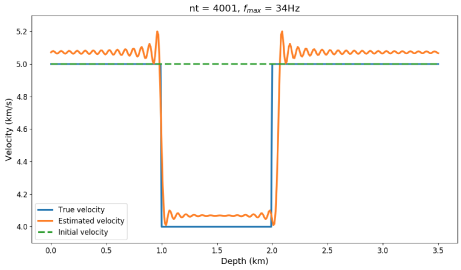
Looking first at the effect of time redundancy, or more accurately the frequency-sampling density, by comparing the inversion models that used flat spectrum GFs and all frequencies up to the Nyquist frequency (figures 4.6a & 4.6c), increasing the frequency-sampling density improves the phase match of the first reflection (corresponding to the uppermost reflector for a source and receiver co-located at the Earth's surface), reduces the amplitude of the source-receiver co-location artefact, and reduces the DC shift along the length of the trace. When the Green's functions are convolved with a wavelet (figures 4.6b & 4.6d), the source-receiver co-location artefact is again mitigated by increasing the frequency-sampling density in the data recording; however the observed reduction of the DC shift along the data trace is limited to the time windows before the arrival of the first reflection and after the arrival of the second reflection, whilst the improved phase match of the first reflection is not observed. For both the flat spectrum Green's functions and those convolved with



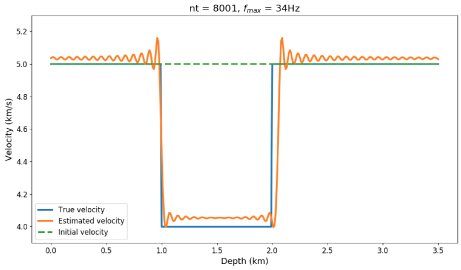
(a) nt=1001 ($d\omega = 0.5$Hz)



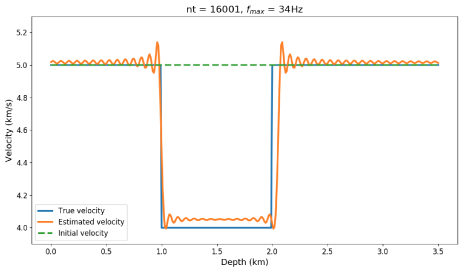
(b) nt=2001 ($d\omega = 0.25$Hz)



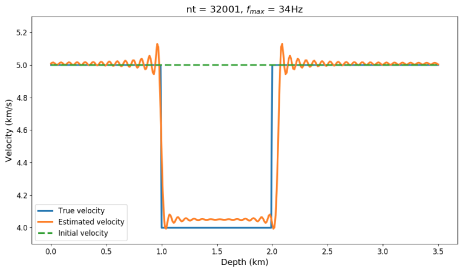
(c) nt=4001 ($d\omega = 0.125$Hz)



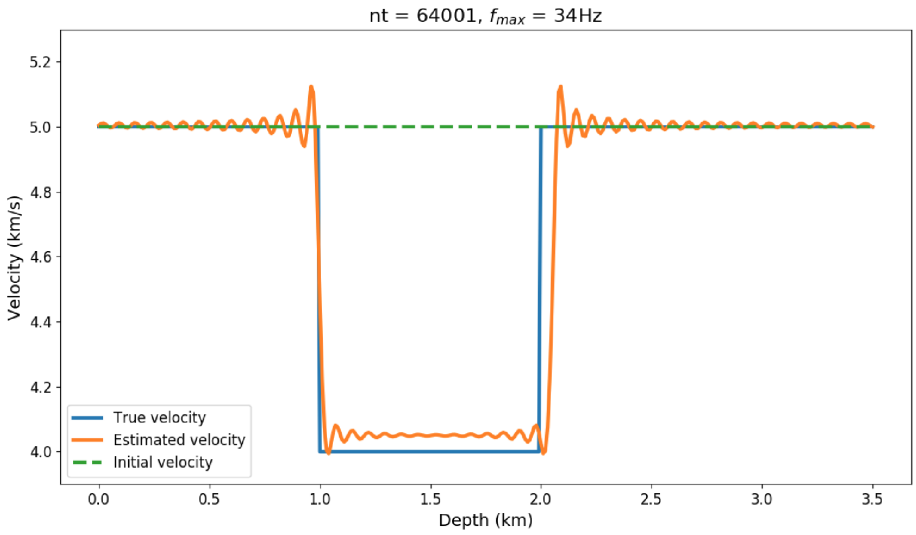
(d) nt=8001 ($d\omega = 0.0625$Hz)



(e) nt=16001 ($d\omega = 0.03125$Hz)



(f) nt=32001 ($d\omega = 0.015625$Hz)



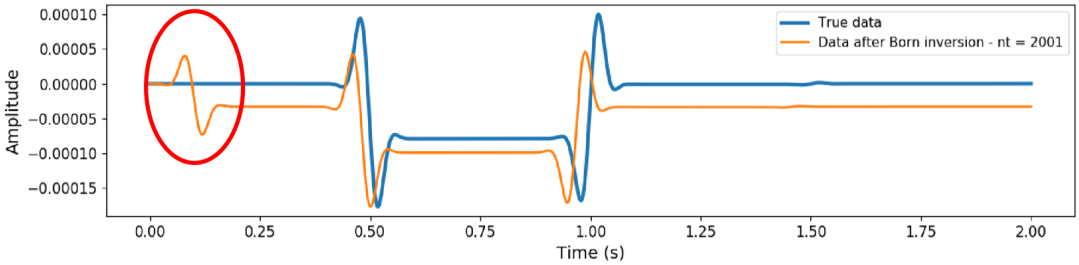
(g) nt=64001 ($d\omega = 0.0078125$Hz)

Figure 4.5: Analytical Born inversion of the 1D slab model for varying lengths of time series, increasing in powers of two.

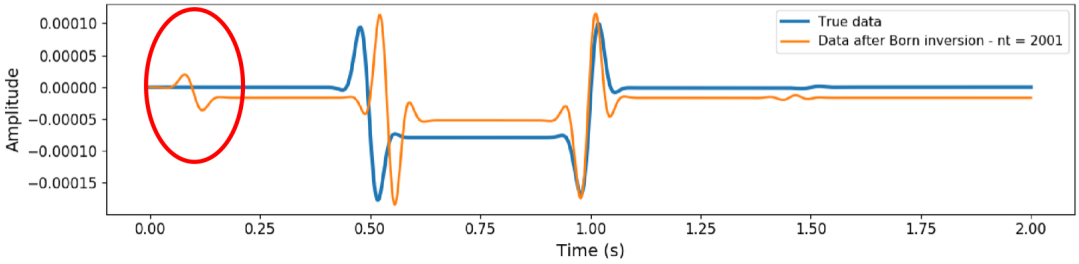
the (Gabor) wavelet, increased frequency-sampling density results in time retardation of the second reflection (from the bottom of the slab) and all subsequent reflections thereafter. It is worth noting that it is not realistic to observe a DC component to the data, since this cannot be accurately recorded using mass/spring transducers (e.g. geophones, hydrophones, seismometers).

To study the effect of the sign of the velocity perturbation on the inversion result, the velocities of the constant background model and the perturbation in figure 4.3 were reversed and the experiment repeated. Figure 4.7 shows the inversion results for the cases (a) where the wavefield is convolved with the 12.5Hz Gabor wavelet, and (b) using a white spectrum bandlimited to the frequency range of the 12.5Hz Gabor wavelet. The results are approximately equivalent to the those for the negative velocity perturbation (figure 4.4), although the magnitude of the DC shift is slightly reduced for the case of a positive velocity perturbation.

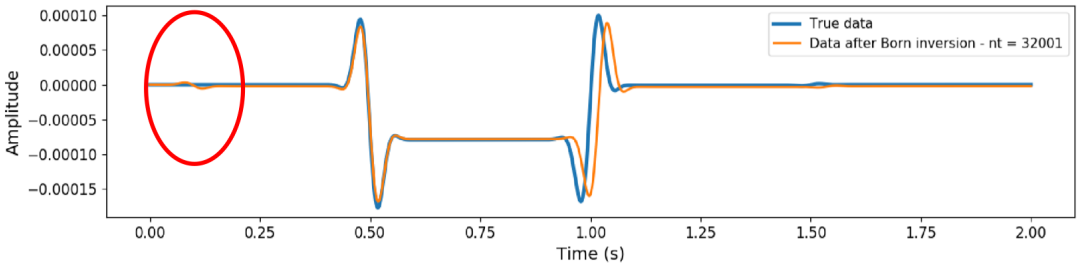
Comparing figures 4.5d and 4.7b, the error in the depth estimation of the second interface can be explained by the absence of the 0Hz frequency bin in the inversion. For the case of a negative velocity perturbation, or equivalently a positive slowness(-squared) perturbation (figure 4.5d), underestimation of the slowness-squared in the perturbed region corresponds to overestimation of the velocity. Thus to match the arrival-time of the reflection from the second interface, the depth to the second interface must increase. Conversely, in the case of a positive velocity perturbation or negative slowness(-squared) perturbation (figure 4.7b), overestimation of the slowness-squared corresponds to underestimation of the velocity requiring underestimation of the depth to the second interface for the data to match. For a negative slowness-squared perturbation, over-estimating the slowness-squared in the perturbed region is equivalent to under-estimating the amplitude of the slowness-squared perturbation i.e. the magnitude of the slowness-squared perturbation is under-estimated regardless of the sign of the perturbation.



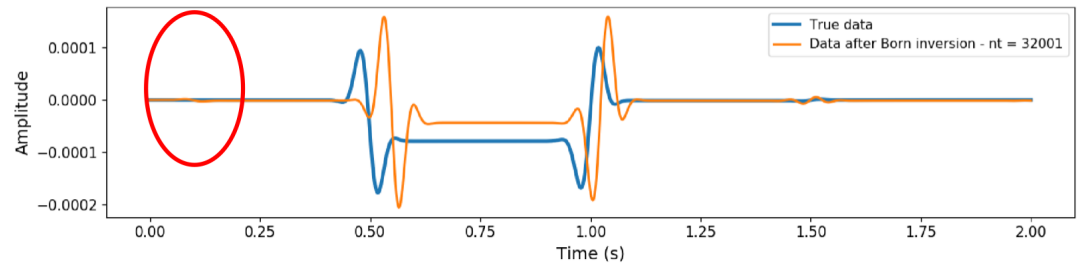
(a) Following Born inversion with white spectrum up to Nyquist frequency and nt=2001



(b) Following Born inversion with 12.5Hz Gabor wavelet and nt=2001

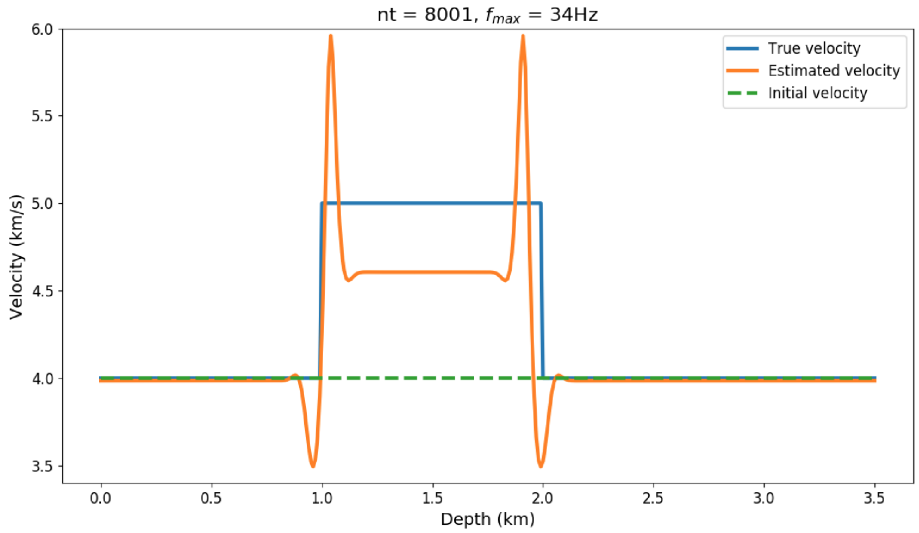


(c) Following Born inversion with white spectrum up to Nyquist frequency and nt=32001

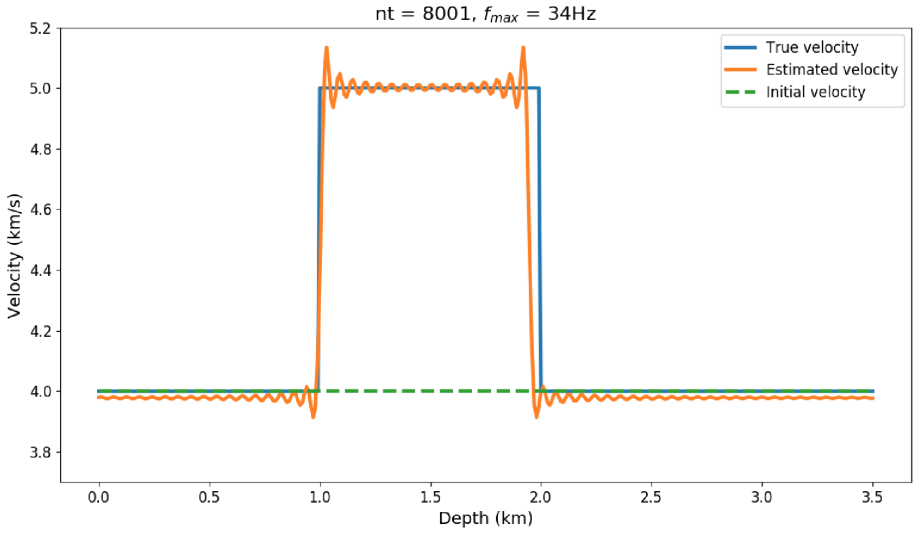


(d) Following Born inversion with 12.5Hz Gabor wavelet and nt=32001

Figure 4.6: Data following 1D Born inversion of Beydoun and Tarantola (1988) slab model.



(a) With 12.5Hz Gabor wavelet



(b) With white spectrum in bandwidth of 12.5Hz Gabor wavelet

Figure 4.7: 1D Born inversion of a slab model with a positive velocity perturbation.

4.2 Comparison of Born inversions using analytical Green's function solutions to Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inverse solutions using finite-difference wavefields

In the previous section, I established the accuracy achievable with Born inversion using the analytical framework of Beydoun and Tarantola (1988). By comparing this analytical solution to a Born inversion of the same model using finite-difference wavefields and the linear least-squares Born inverse solution (equation 2.38), one should be able to qualitatively predict the achievable accuracy of an inversion of a similar model with no known analytical solution.

If the two Born inversion solutions are comparable, and if the hypothesis of this thesis is true, that the inclusion of physics describing all orders of scattering should improve the accuracy of the inverse scattering problem solution; then the Born inversion result will serve as a lower bound on the model accuracy expected from Marchenko-Lippmann-Schwinger inversion. As a result of errors expected in the Green's function retrieval process using Marchenko methods, the upper bound on model accuracy would be provided by Lippmann-Schwinger inversion.

For this comparison, I use the analytical inversion results where the Green's functions are convolved with the source wavelet. This same source wavelet is used for finite-difference modelling. A 50-point cosine taper has been applied to both ends of the inversion results to prevent issues with solution scaling due to exploding amplitudes at the top and bottom of the model.

Figure 4.8 compares the results of Born inversion using the analytical method of Beydoun and Tarantola (1988) to (top) linear least-squares Born inversion using equation 2.38, (middle) Lippmann-Schwinger inversion using equation 2.67, and (bottom) Marchenko-Lippmann-Schwinger inversion using equation 2.67. The difference between Lippmann-Schwinger inversion and Marchenko-Lippmann-Schwinger inversion

is that Lippmann-Schwinger inversion uses Green's functions G modelled in the true medium in equation 2.67, whereas Marchenko-Lippmann-Schwinger inversion uses Green's functions estimated with a Marchenko method. It is evident that the analytical inversion has estimated the correct velocity model structure (albeit with incorrect amplitude) whereas the matrix-based inversions utilising finite-difference wavefields have not. Whilst the inverse scattering problem is known to be ill-posed, the extent of this ill-posedness was grossly underestimated *a priori* when this work was initiated. One possible remedy for ill-posedness is to regularise the inverse problem.

To deal with the ill-posedness, I considered the following strategies:

1. iteration of the unregularised inverse problem;
2. regularisation by Tikhonov damping of the kernel matrix;
3. regularisation by applying smoothing to the kernel matrix; and
4. a combination of iteration with Tikhonov damping or smoothing.

It should be noted that the analytical solution cannot be iterated, since it assumes that the initial velocity model is homogeneous.

Beginning with iteration of the unregularised problem, figure 4.9 shows the velocity models found after 50 iterations of inversion using Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversion. It is clear that iterating the inverse problem cannot stabilise the solution by itself - the resultant models returned by all three methods are still dominated by random noise, with the only significant difference being an increase in the amplitude of the noise associated with the Marchenko-Lippmann-Schwinger result.

Figure 4.10 shows the effect of applying Tikhonov damping to each inversion method, the amount of damping applied is that which optimises the L2-fit of each inversion model to the true model. Both the Lippmann-Schwinger and Marchenko-Lippmann-Schwinger results show a strong response in the region of the first reflector followed by a decrease in velocity below the reflector and a less precise response in the region of the second reflector. In contrast, the Born inversion result displays no clear trend and, like the unregularised solution, is dominated by noise.

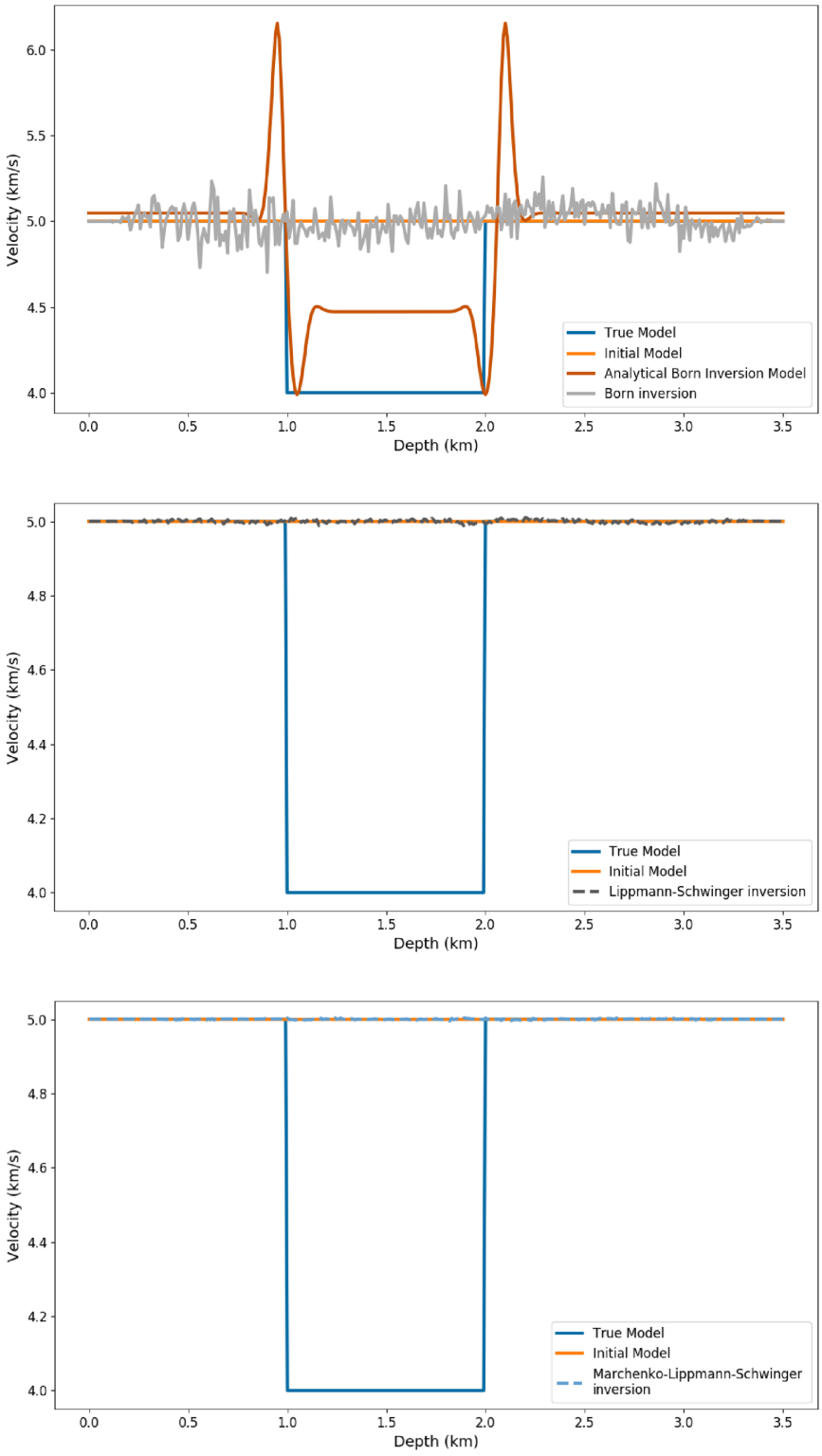


Figure 4.8: Comparison of analytic inversion to least-squares Born inversion (top), Lippmann-Schwinger inversion (middle), and Marchenko-Lippmann-Schwinger inversion (bottom) for the slab model of Beydoun and Tarantola (1988).

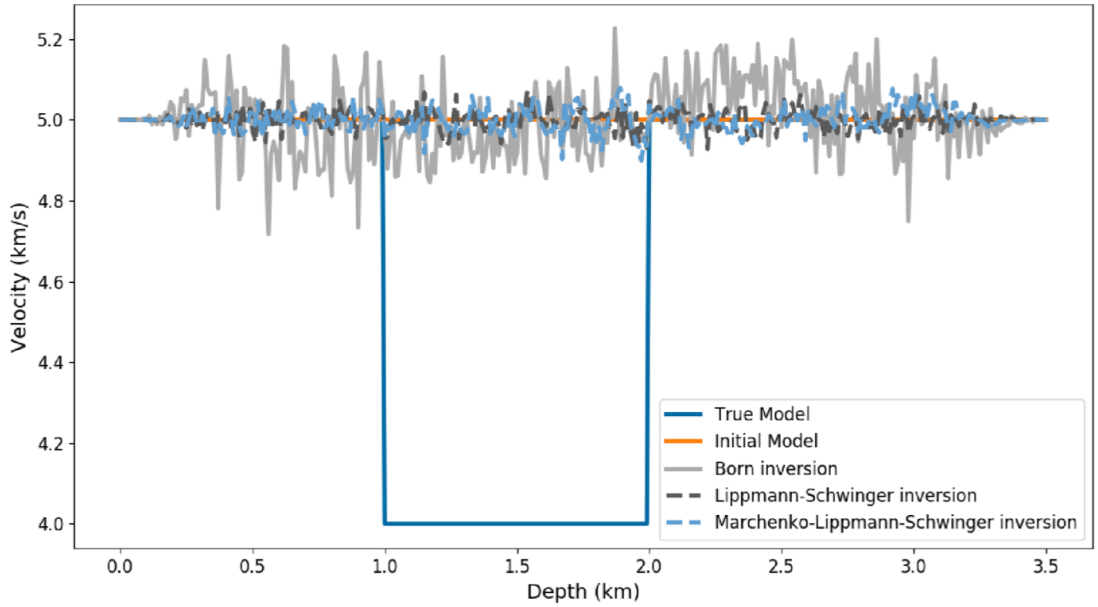


Figure 4.9: Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversion models after 50 iterations of each scheme.

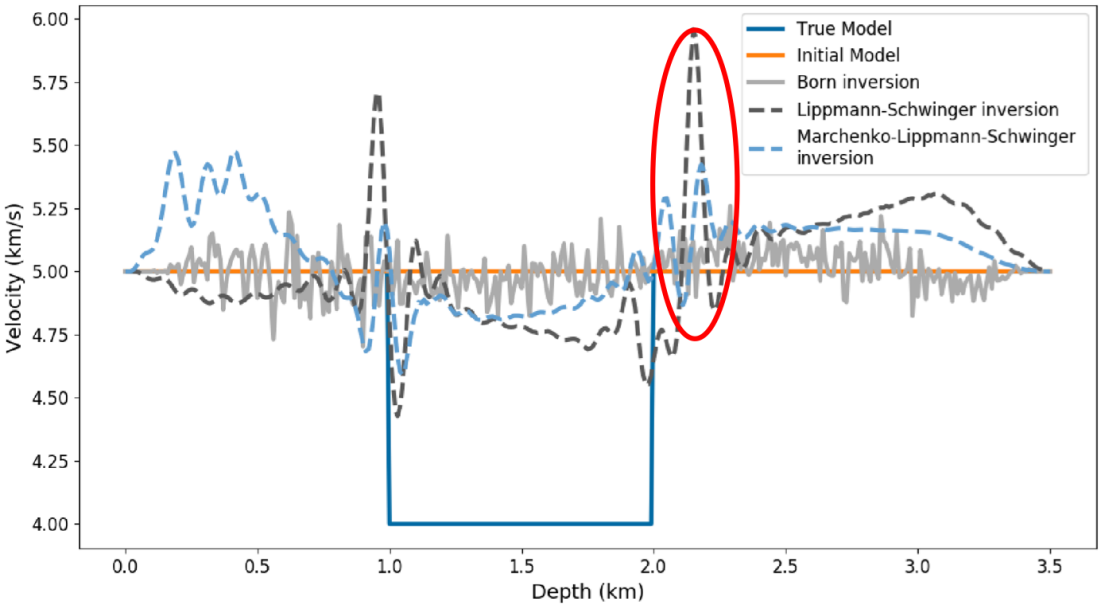


Figure 4.10: Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversion models after a single iteration of each scheme with Tikhonov damping.

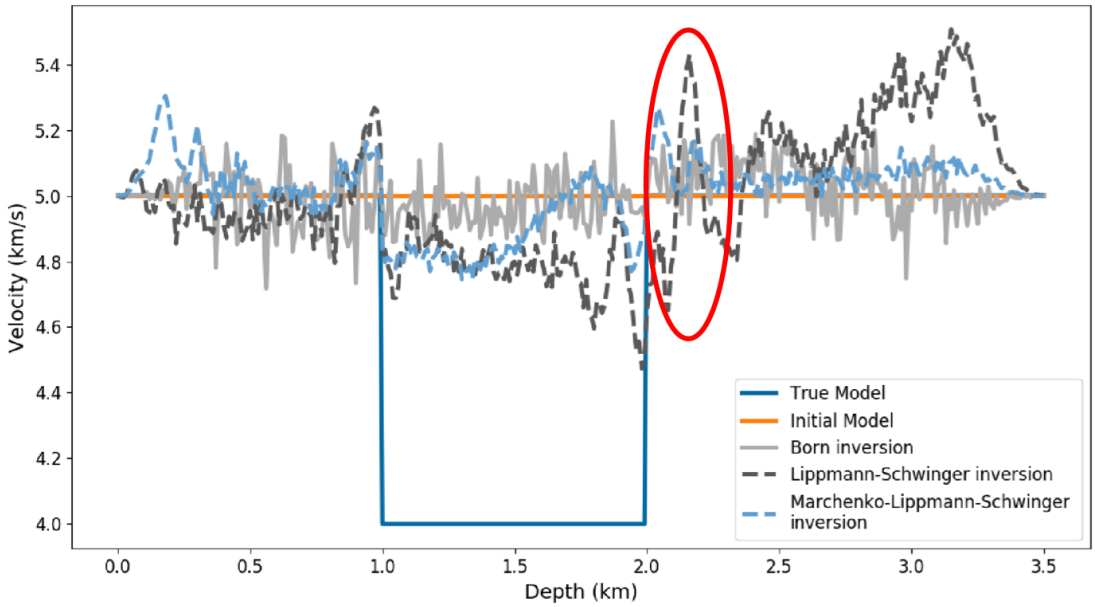


Figure 4.11: Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversion models after 50 iterations of each scheme with Tikhonov damping optimised at every iteration.

Iterating the previous result, figure 4.11 shows the models found after 50 iterations of Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversion with Tikhonov damping - for each inversion method, damping has been optimized at every iteration. The damped Lippmann-Schwinger and Marchenko-Lippmann-Schwinger methods have refined the initial inversion result, largely removing the imprint of the seismic source wavelet on the inversion result at the first reflector. At the second reflector, the Marchenko-Lippmann-Schwinger model has a stronger response, while the strong response below the reflector in the Lippmann-Schwinger solution has been reduced in amplitude. A combination of Tikhonov damping and iteration appears to have had no impact on the Born inversion result.

Figure 4.12 shows the result of a single iteration of inversion with smoothing applied to the kernel matrix. Smoothing is implemented by appending a spatial first-derivative stencil to the kernel of the Lippmann-Schwinger equation and appending zeros to the data vector so that the linear system of equations remains consistent. The strength of the smoothing is controlled by linear scaling of the spatial derivative stencil, and is optimized by line-search. Marchenko-Lippmann-Schwinger inversion has identified the first interface, with a significantly weaker response close to the

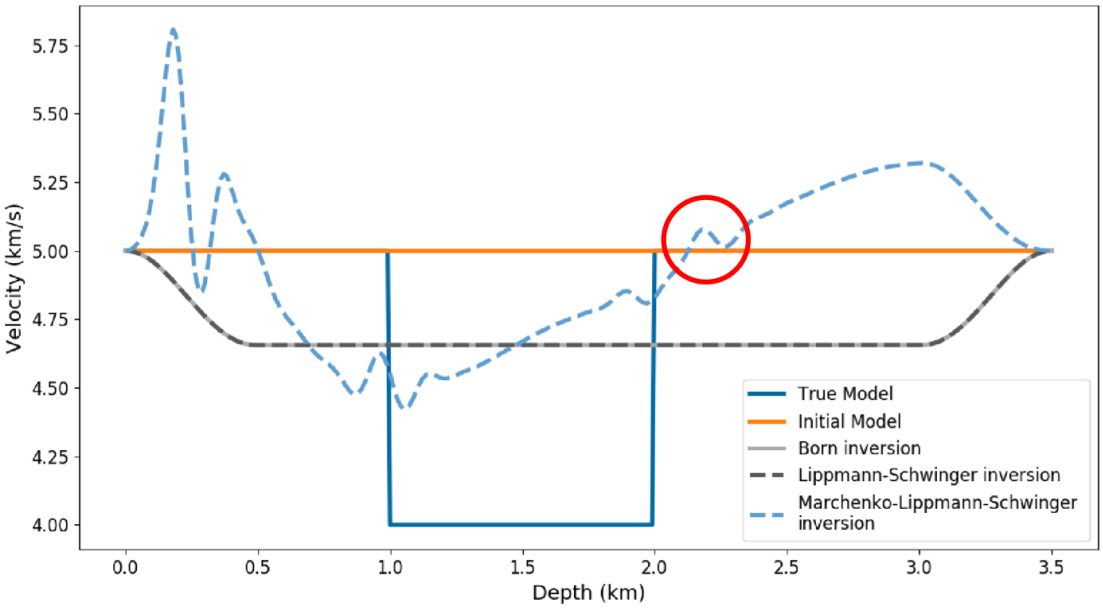


Figure 4.12: Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversion models after a single iteration of each scheme with smoothing applied.

second interface and a clear velocity gradient between the two reflectors. Interestingly, the Born and Lippmann-Schwinger methods find almost identical inaccurate models despite clear differences in physics. The results after 50 iterations of smoothed inversion are shown in figure 4.13. Now the Born and Lippmann-Schwinger inversion models have diverged, with the Born inversions having found a single wide and smoothly varying perturbation, while the Lippmann-Schwinger and Marchenko-Lippmann-Schwinger inversions both agree on the probable locations of two reflectors and an overall decrease in velocity in the area between them. However, the iteration process appears to have smoothed and flattened the velocity structure between the two interfaces into an interpretable single unit in the Marchenko-Lippmann-Schwinger case, whilst this has not happened in the Lippmann-Schwinger case.

4.3 Discussion

By studying an analytical analogue of the problem this thesis aims to solve, we are able to draw a number of conclusions about the ideal conditions for which this problem can be solved. Comparing figure 4.4b, which is a band-limited inversion within the bandwidth of the source wavelet and has a maximum frequency of 34Hz, with figure

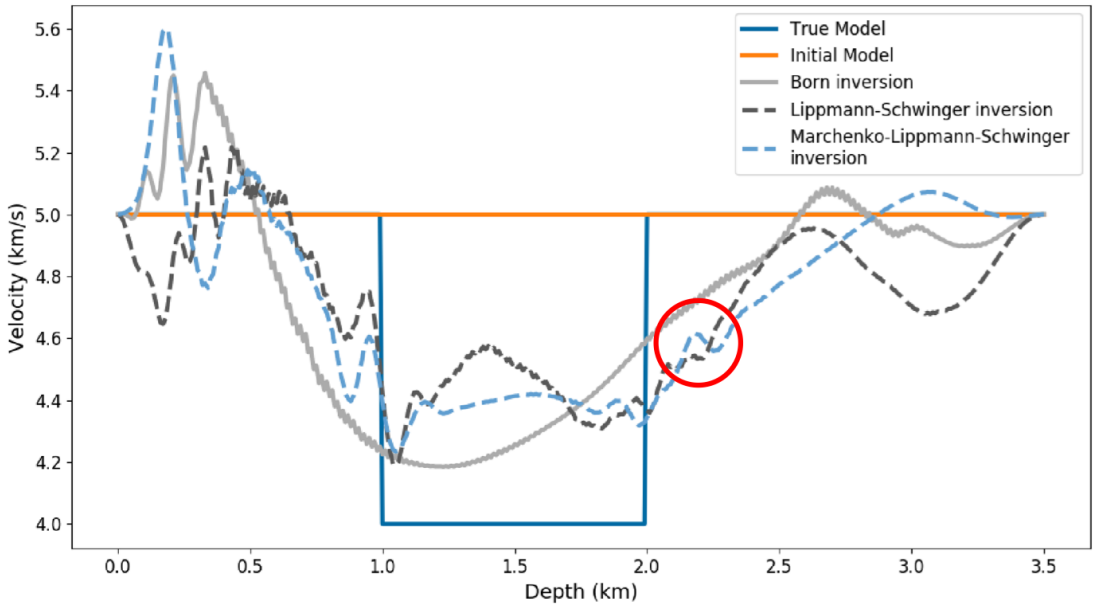


Figure 4.13: Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversion models after 50 iterations of each scheme with applied smoothing optimised at every iteration.

4.4d, which employed a white spectrum up to the Nyquist frequency of 250Hz, we see that while the amplitude error in the velocity perturbation from the band-limited inversion is greater than the full-bandwidth inversion, both inversions have accurately demarcated two vertical interfaces corresponding to the vertical sides of the velocity perturbation present in the true model. From this, we can conclude that extremely high seismic frequencies are not required to image vertical structures using Born inversion - typical frequency bandwidths used in exploration seismology are sufficient to image sharp interfaces. Further comparison with figure 4.4c shows that by employing a white spectrum band-limited to the frequency range of the source wavelet deployed, the amplitude error in the perturbation can be reduced to the same level as if all frequencies up to Nyquist had been used. This tells us that it is the lowest frequencies that are responsible for correctly determining amplitudes.

The "ringing" present in the band-limited inversion in figure 4.4c but not present in figure 4.4d may partially be explained by the Gibbs phenomenon, the rapidly decaying oscillations around large vertical steps in the Fourier representation of a time series due to the periodicity assumption of the Fourier transform. By bandpassing the spectra of the Green's functions, one or more step functions are introduced to

the spectrum. These step functions are represented in the Fourier domain as a finite sum of periodic functions that overshoot and undershoot the amplitude of the step function, resulting in the characteristic oscillations.

In the original paper by Beydoun and Tarantola (1988), the number of time samples used in their analytic modelling was not specified, thus it was necessary to investigate the role of time redundancy on the inversion. Since the modelling procedure assumes waves are only singly-scattering, any time after the arrival of the second reflection (from the bottom of the slab) may be considered redundant. In doing so, I showed in figure 4.5 that increasing the time redundancy of the modelled Green's functions reduced the amplitude error in the inversion; specifically, it reduced the DC shift present in the result. Combined with the previous observation that the lowest frequency components are responsible to obtaining correct amplitudes, I conclude that increasing time redundancy reduces the amplitude error in the inversion result by improving the resolution of low frequencies in the inversion. This can be seen by plotting the amplitude spectrum of the wavelet using increasingly long time series (and thus increasing time redundancy), as in figure 4.14. As the time redundancy increases, the size of each frequency bin decreases because the number of bins increases while time sampling rate remains fixed. This increased sampling density in the frequency domain allows for rapidly varying low frequency features of the spectrum to be correctly represented. Consideration should be taken of the increase in computational expense that comes with increased time redundancy, whilst the cost is trivial for this toy model and this analytical inversion scheme, for larger models employing more expensive modelling techniques such as finite-differences and matrix operations for the inversion, the cost can grow rapidly. Time-domain data modelling expense scales linearly with time modelled, and modelling time periods with no signal would be considered poor practice when compared to padding time series with zeroes, which is standard practice in data processing in many fields. Since the increase in cost of matrix operations, especially complex operations such as calculating matrix inverse, scale super-linearly with the increase in time or frequency samples, care should be taken to limit the time redundancy to that necessary to achieve the level of accuracy

needed in the result to avoid unnecessary expense.

Using figures 4.4b-4.4d and 4.5, we can conclude that Beydoun and Tarantola (1988) were likely using band-limited white spectrum wavefields with large amounts of time redundancy. This is based on the ringing character around the steps of the perturbation, the minimal DC shift in the solution outside of the perturbed velocity region, and the absence of the wavelet shape imprinted on the inversion. This implies that to obtain the "ideal" Born inversion result (i.e. consistent with the analytic solution), it is necessary to deconvolve the wavefield of any wavelet used. In practicality, deconvolution is difficult to perform without introducing artefacts that will carry through any further processing, and the decision to perform deconvolution or not is thus based on the structure you expect to be imaging and the expected use for the result. If accurate true-amplitude images are not required, or if high vertical resolution is not a priority, it may be better to accept the reduction in vertical resolution caused by the imprint of the wavelet to prevent noise or artefacts being introduced by a poor deconvolution.

Considering all the evidence from the analytic Born inversion together, I conclude that it is not possible to invert for the exact structure of the true model using Born inversion using the method of Beydoun and Tarantola (1988), even when the ideal conditions for Born inversion are satisfied. If their method is a true analytic representation of the Born forward and inverse scattering equations, then any Born inversion result is inherently inaccurate. From figure 4.6, we can see that the velocity error introduced by the inversion results leads to an error in the data domain. As a consequence of these errors, one can only expect to match the first reflection arrival when ideal inversion conditions apply, any subsequent reflections will be time advanced or retarded relative to their equivalent arrival in the true data based on whether the velocity perturbation was positive or negative. Figures 4.4 and 4.5 show that the velocity and the depth to target are overestimated when the velocity perturbation is negatively-valued; as shown in figure 4.7, the velocity and depth to target are underestimated when the velocity perturbation is positively-valued. If the velocity error is large enough that the subsequently modelled data is more than half the time period of the wavelet advanced or retarded relative to the true data, that model would be

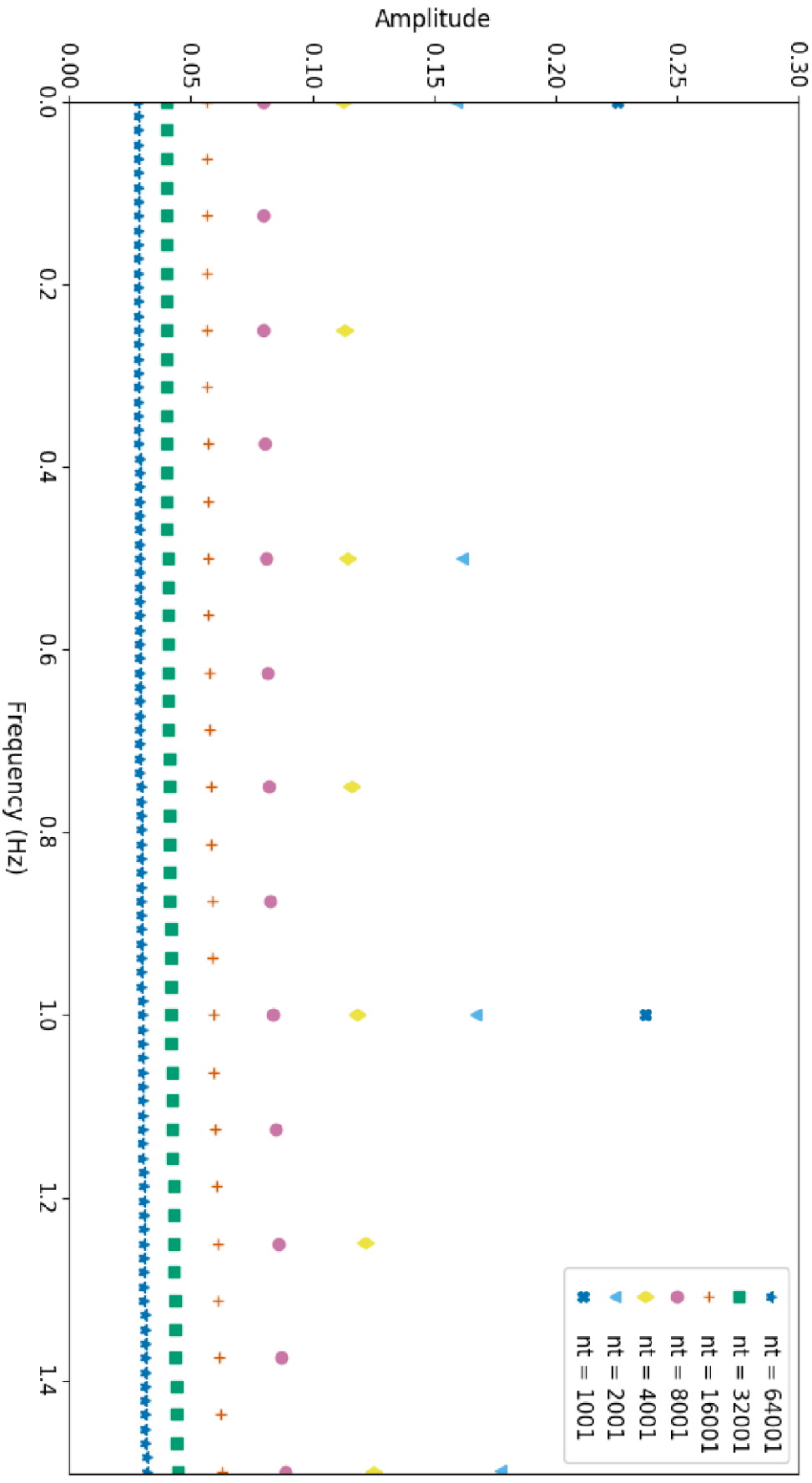


Figure 4.14: Close-up plot of the amplitude spectrum of the 12.5Hz Gabor wavelet in the range 0 - 1.5Hz as a function of time series length in samples.

unsuitable for use as a starting model in other least-squares inversion methods such as full waveform inversion, limiting the potential applications of Born inversion. This problem is called cycle-skipping.

When we compare the result of the analytical inversion to the Born inversion by matrix inversion (figure 4.8), we see that the analytic inversion correctly finds the location of the upper interface of the slab, and identifies the lower interface with a small location error; the amplitude of the velocity perturbation is overestimated, but remains clearly identifiable. In contrast, the Born inversion by matrix inversion is contaminated with random noise and presents no clear structures. When also compared to Lippmann-Schwinger and Marchenko-Lippmann-Schwinger inversion, the amplitudes of the model updates are so small as to be almost indistinguishable from the initial model; if one were to zoom in on the models, one would find that they are also contaminated with random noise at all depths. Given the simplicity of the structure to be inverted for, it is clear that I had underestimated the level of ill-posedness of the problem and the likely need for some kind of regularisation.

The methods chosen to reduce the ill-posedness (Tikhonov damping, smoothing, and iteration) were chosen for their simplicity to implement and their existing ubiquity in geophysical inversion. Figure 4.9 shows that after 50 iterations of inversion, there was no significant change in the character of any different inversion results, except to increase the amplitude of the noisiness on the Lippmann-Schwinger and Marchenko-Lippmann-Schwinger inversions.

Whilst the results remain noisy, Tikhonov damping significantly improves the outcomes of Lippmann-Schwinger and Marchenko-Lippmann-Schwinger inversion with a single iteration while showing no improvement to the Born inversion. As shown in figure 4.10, Lippmann-Schwinger and Marchenko-Lippmann-Schwinger both identify the upper and lower interfaces, although there are two possible locations for the second interfaces given. Comparing after 50 iterations of damped inversions (figure 4.11), the overshooting of the amplitudes at the slab edges has been reduced. Focusing on the deeper slab edge and observing that the amplitude spikes circled red in figures 4.10 and 4.11 have been suppressed by iteration whilst the smaller amplitude spike at the depth of the slab edge has increased, I conclude that the circled structure is

an artefact of the inversion whilst the shallower amplitude spike is due to correlation of signal that has constructively stacked over many iterations.

As expected, smoothing has the effect of removing noise from the solution. The Marchenko-Lippmann-Schwinger solution shown in figure 4.12 identifies both slab edges, with reduced noise but a lower amplitude response as compared to the Tikhonov damping solution after a single iteration. Interestingly, the Born and Lippmann-Schwinger inversions appear to be identical after a single iteration. The extremely smooth profile of the inversion models and the clear imprint of the model edge tapers suggests these models may be the result of a combination of over-smoothing combined with the model taper. Since I optimised the smoothing by selecting the smoothing level that minimised the model misfit and did not attempt to balance the model domain effect against the data fit (e.g. using the L-curve method (Hansen, 1992)), it is not unreasonable that the best fitting model is one that smooths out all perturbations. After 50 iterations (figure 4.13), the Born and Lippmann-Schwinger models have diverged, suggesting that by iterating with smoothing, it is possible to escape from an initial local minimum in the solution space of the inverse scattering problem by reducing the number of local minima and converging iteratively in linear steps to a solution closer to the global minimum. The iterated Born solution has a single broad perturbation with smooth edges, whilst the Lippmann-Schwinger and Marchenko-Lippmann-Schwinger inversions have converged on similar solutions with a sharp interface at the top edge of the slab and a smoothed out trailing edge. The amplitude of the perturbation for all three inversion methods is superior when smoothing is applied compared to when damping is applied, however it is not clear whether this is a consequence of the simplicity of the true model permitting a simple broad perturbation, or whether this benefit will extend to more complex, many-layered models. By comparing the events circled red in figures 4.10, 4.11, 4.12, and 4.13 that have been identified as artefacts; I conclude that smoothing and iteration combined does not act to suppress imaging artefacts where the combination of damping and iteration does.

It may be possible with a combination of damping and smoothing to reduce the noisiness of the damped solutions, however the complication of optimising both types

of regularisation simultaneously adds to the complexity and computational cost of the problem without any guarantee that it will improve the solution.

Despite the existing application of a taper at the model edges, strong artefacts are still created by both the smoothed and damped inversions at the model edges and more aggressive tapering may be required depending on the model.

Given that the difference between Lippmann-Schwinger inversion and Marchenko-Lippmann-Schwinger inversion is that Lippmann-Schwinger inversion uses the true-model wavefield modelled by finite-differences whilst Marchenko-Lippmann-Schwinger inversion uses an estimate of the true-model wavefield using Marchenko-derived Green's functions, the difference in the inversion results between these methods must be due to the difference between these two wavefields. Figure 4.15 shows the difference between the true-model wavefield and the Marchenko estimate of that wavefield before the first iteration of inversion. Looking at the wavefields across all times and depths, figures 4.15a and 4.15b shows that the Marchenko estimate has a negative amplitude bias applied almost homogeneous across all depths after the arrival of the direct wave at that depth. Comparing the wavefield at different depths, figures 4.15c, 4.15d, and 4.15e show a consistent, negatively-valued DC shift in the amplitude of the Green's functions estimated with a Marchenko method as compared to the true-model Green's functions. The Marchenko estimates of the Green's functions are generally a close match in both amplitude and phase to the true Green's functions, with the exception of figure 4.15e, which shows in the enlarged panel that the first reflected arrival in the Marchenko estimate of the Green's function is phase-retarded compared to the true Green's function.

The DC shift in the Marchenko estimate may be explained by the additional processing necessary for the Marchenko scheme to converge in this example. To prevent the amplitude of the Marchenko estimate from exploding, I decided to apply the Marchenko method to direct arrivals using the time-derivative of the 12.5Hz Gabor wavelet used in Beydoun and Tarantola (1988) as their source signature, and then integrated the resultant Green's functions. The reason I believed this would work is that wave-equation modelling in 1D results in a wavefield with a source wavelet equal to the time integral of injected source wavelet, but without any phase rotations. So,

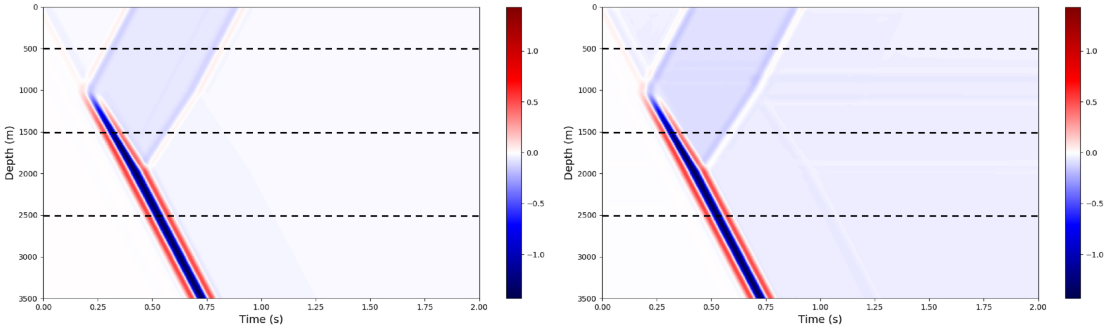
in principle, the resultant Marchenko-estimates of the Green's functions should have been the time-derivatives of the signal I required, and integrating these signals in time should have then resulted in the desired Green's functions scaled in amplitude. The limitation in this method is that the real part of the integral of the time-derivative of the Gabor wavelet does not exactly match the real part of the original wavelet (the Gabor wavelet is a complex function, of which only the real part is used). As shown in figure 4.16, the time derivative does not have the DC component present in the original waveform. Thus the Marchenko estimates of the Green's functions using the integral of the time derivative are also missing the DC component.

To summarise, in this chapter I examined the analytical Born inversion of a 1D slab model by Beydoun and Tarantola (1988) and reached the following conclusions:

1. Increasing the frequency-sampling density in the Green's functions improves the retrieved amplitudes of the model perturbation and reduces the DC shift by improving the sampling of low frequencies, and
2. Even under ideal conditions of having analytically exact single scattering wavefields and an analytic solution to the inverse problem, Born inversion cannot retrieve the exact velocity model.

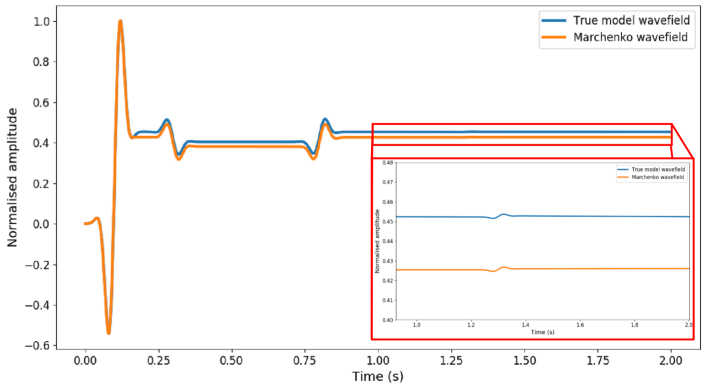
I then compared the results of the analytic Born inversion of the slab model to the results of Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversions using matrix inversion and Green's functions modelled by finite-differences. I found that

1. For the slab model, regularisation is essential for the three matrix inversion methods tested. Iterating the problem is not sufficient to regularise the problem.
2. Tikhonov damping is effective at identifying interfaces when using Lippmann-Schwinger and Marchenko-Lippmann-Schwinger inversion. With iteration, amplitude errors are reduced.
3. Smoothing, especially when combined with iteration, is more effective than Tikhonov damping at estimating the amplitude of the model perturbation.

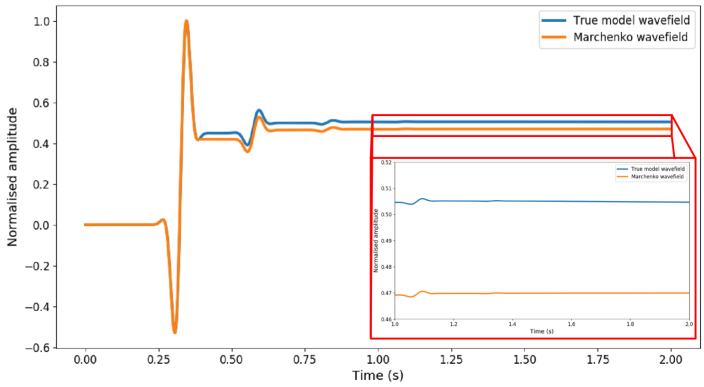


(a) True-model wavefield.

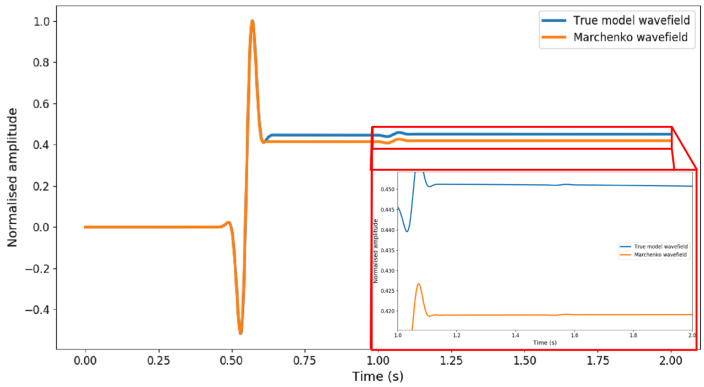
(b) Marchenko estimate of the true-model wavefield.



(c) Wavefields at z=500m

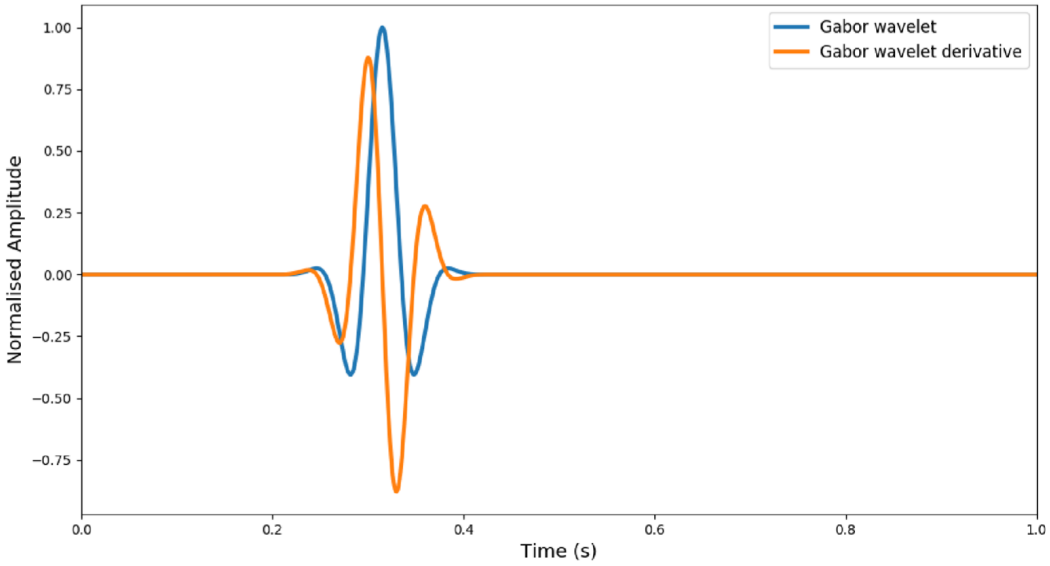


(d) Wavefields at z=1500m

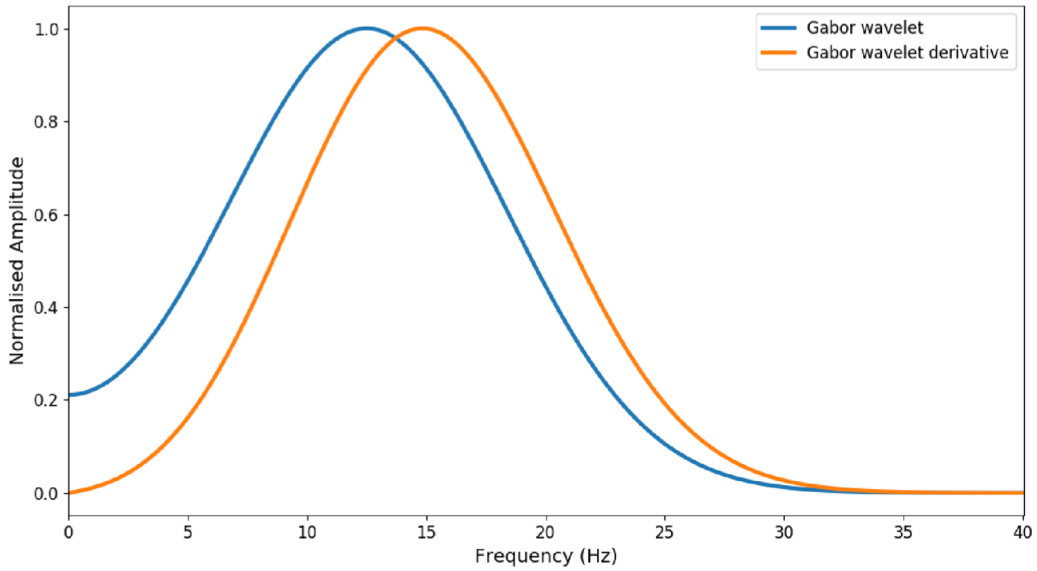


(e) Wavefields at z=2500m

Figure 4.15: Comparison of the wavefield in the true model and its estimate using a Marchenko method before the first iteration of inversion. Both wavefields have been amplitude normalised for comparison. Depth sections (c)-(e) are marked by dashed black lines in panels (a) & (b).



(a) The Gabor wavelet and its time-derivative, truncated at one second.



(b) Amplitude spectra of the wavelets in (a) truncated at 40Hz.

Figure 4.16: Comparison of the 12.5Hz Gabor wavelet used by Beydoun and Tarantola (1988) to its time derivative, in both time and frequency domains.

Chapter 5

Linear Inversion of a Many-Layered Velocity Model in One Dimension

In this chapter, I apply the conclusions from the inversion of the boxcar model to the inversion of a many-layered 1D model using Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversions. This model is a 1D section from the 2D model that I will analyse in the next chapter. I show the results of inversions over 50 iterations employing both Tikhonov damping and smoothing regularisations. I do this in two ways: by optimising the inversion in the model domain, i.e. by answering the question "which combination of methodology and regularisation gives the gradient direction that points most directly towards the true structure?";and by optimising the inversion in the data domain, i.e. "which combination of methodology and regularisation gives the model with the best fit to the true model data?". The first test would never happen in a reality: if you already knew the answer you would not be looking for it! It does, however, give insight into how the different methods seek to update the velocity model. The second test is a more realistic test of how these inversions methods would perform in an actual field experiment, using only information that would be available in such circumstances.

Given the unimpressive results presented previously, especially those without regularisation applied, a fair question to ask is "what is the value of testing a more complex model given the disappointing previous results?" to which one must question how complexity is defined. From a 1D perspective, I propose there are only two types

of complexity: the degree of layering, and and the lack of prior information.

The accepted logic is that more layers in the true model equals more complexity, and that thin layers are more complicated than thick layers. While the first statement is unquestionable, the second is more nuanced. To image a thick layer, the high-frequency information needs to be correct to precisely locate the layer interface, but the low frequency information must also be correct to set an accurate and consistent amplitude along the depth of the layer. In contrast, to image a thin layer, the low frequencies are arguably less important because the length scales over which the amplitude needs to be consistent are also smaller.

The challenge thin layers present is in identifying them, assuming the data has the correct frequencies to resolve them. Here, both layering and prior information play a role in creating difficulty in the inverse problem. Thin layers result in smaller differences in travel paths between reflected arrivals in our data. If the additional distance is small enough, the difference in traveltimes may be small enough for the waveforms to overlap. In this case, two overlapping reflections can be interpreted as a single event. Prior information could help us resolve this, but is often insufficient for this task. For our purposes, prior information takes the form of the initial velocity model. A small change in velocity at or near the location of the thin layer could imply the existence of a thin layer. If the prior information is overly smooth or lacking in resolution, as our initial velocity models often are, then we may not be able to resolve thin layers.

A third aspect of layering to consider is the amplitude of each layer, or more specifically, the magnitude of the change in amplitude between layers. This is closely related to the second type of complexity, prior information. The more accurate is the prior model, the less information needs to be resolved by the inversion and the better the result is expected to be. Conversely, the less accurate is the initial model, the more information needs to be resolved by the inversion. Where the change in amplitude is strong, prior information tends to be less accurate. This is assuming that the data contain sufficiently high frequencies to be able to resolve the structures in the true model.

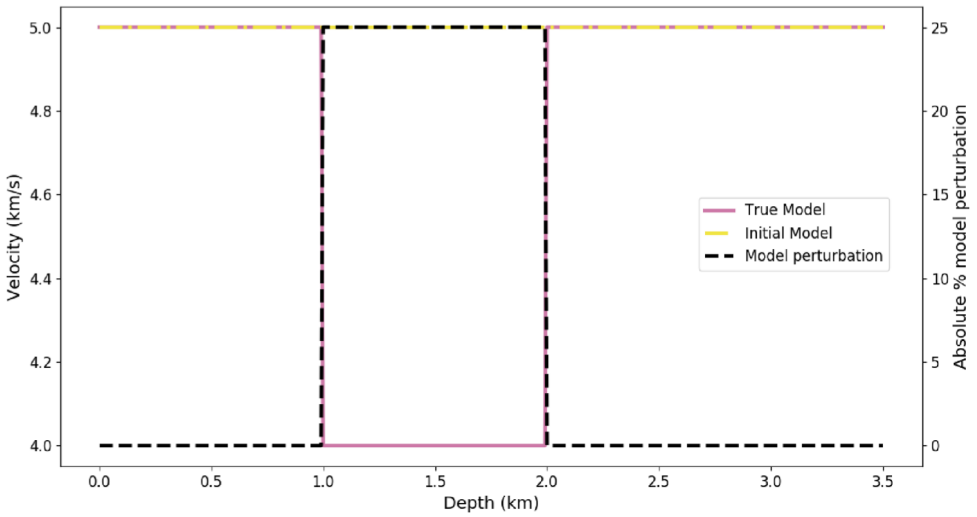


Figure 5.1: 1D slab model of Beydoun and Tarantola (1988) plotted against the absolute percentage error of the initial model.

Figure 5.1 shows the true and initial models from the previous experiment plotted against the absolute error between the two models as a percentage of the amplitude of the true model. The maximum error in the initial model is 25% of the true model amplitude, which is not insignificant. The shape of the model perturbation is also important: it is not a gentle gradient but a steep increase from minimum to maximum error. As discussed previously, step functions are difficult to represent in the Fourier domain due to the periodicity assumption which results in a series of over-corrected and under-corrected amplitudes, an effect known as the Gibbs phenomenon. If we considered the model perturbation in the same way as we would a filter, then we would be less surprised to see noise or ringing if we had applied a strong step function as a filter to data. If we compare that to the many-layered model that I show in this subsection in figure 5.2, we see that the new model has smaller and larger amplitude perturbations than the previous model, and that all perturbations are significantly thinner than the previous model, but because of the prior information provided by the smooth initial model, the model perturbations themselves also ramp up smoothly to their maximum values. So to answer the question posed above, the value comes from the fact that the slab model is a more complex challenge than it first appears, and that any failures in that context do not mean that any or all of the inversion methods tested are universally inadequate.

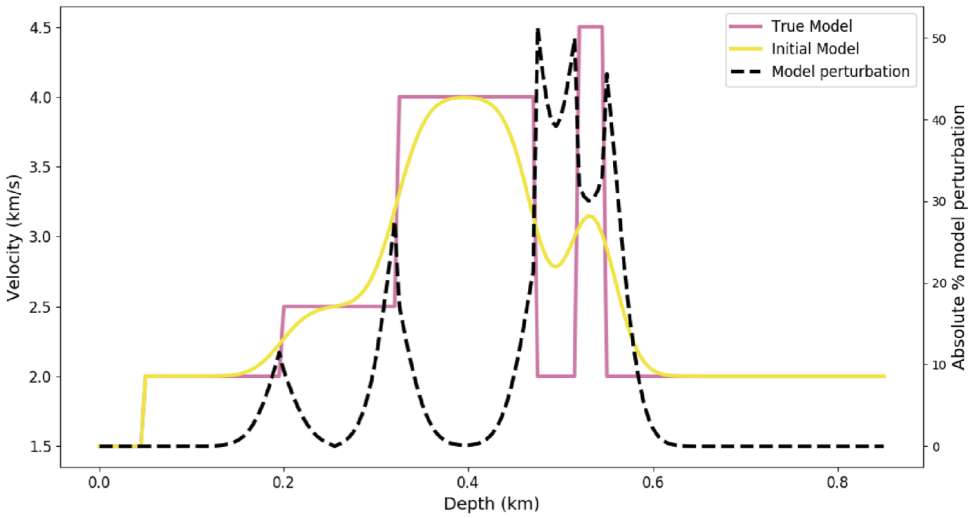


Figure 5.2: 1D many-layered model plotted against the absolute percentage error of the initial model.

5.1 Inversion optimised in the model domain

I now compare the performance of our three inversion methods, Born inversion, Lippmann-Schwinger inversion, and Marchenko-Lippmann-Schwinger inversion; on the many-layered model, using the least-squares residual between the true model and the inversion model to optimise the inversion result at each iteration. For this study, I use a maximum of 50 iterations.

Figure 5.3 shows the result of all three inversion methods after a single iteration and without any regularisation, serving as the benchmark against which the performance of regularised inversion should be measured. Born and Lippmann-Schwinger inversions show improved resolution of the interface between layers 3 and 4, and in improvement in amplitude in layer 6, but at the expense of adding noise to the inversion model. Marchenko-Lippmann-Schwinger inversion shows almost no change compared to the initial smooth model.

Figure 5.4 shows the subsequent models after 50 iterations of inversion. All three models are largely consistent with the initial smooth velocity model except for the addition of noise at all depths except where a taper has been applied at the model edges. Lippmann-Schwinger inversion shows some improvement in resolving the interface between layers 2 and 3 of the true model (the smallest model perturbation at approximately 10% of the true model velocity), as well as improvements in amplitude

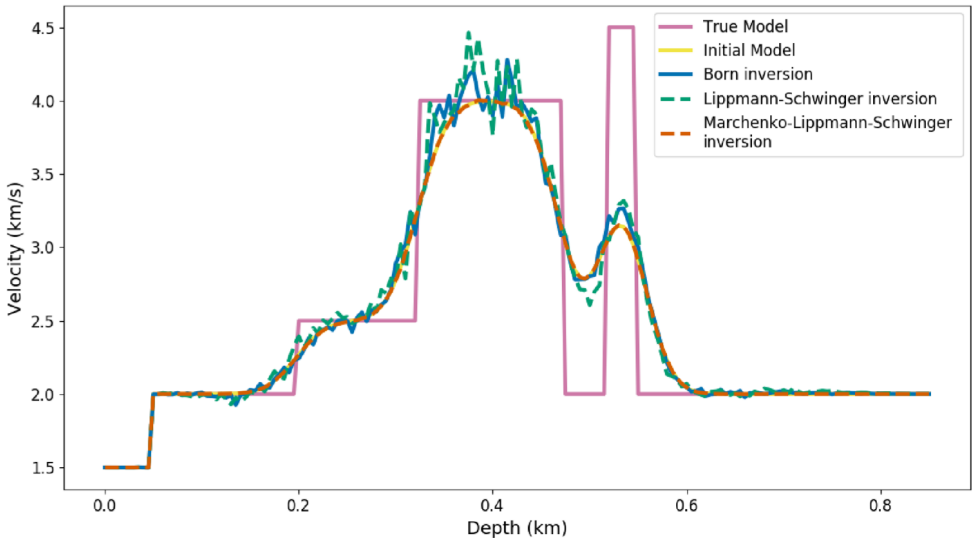


Figure 5.3: Velocity models after a single iteration of inversion without regularisation.

in layers 5 and 6 (the largest model perturbations at approximately 50% of the true model velocity). Marchenko-Lippmann-Schwinger inversion offers the most improvement in resolving the interfaces above and below layer 4, while Born inversion makes the best estimation of the velocity of layer 6.

The change in model residual over the 50 iterations is shown figure 5.5, and the corresponding change in the data residual in figure 5.6. Starting from the same model (and thus the same value of the model residual), Lippmann-Schwinger inversion shows strong immediate improvement and finishes the 50 iterations with the largest reduction in the model residual, followed by Born inversion. Marchenko-Lippmann-Schwinger inversion shows weak but consistent improvement in the model residual across all iterations. Comparing to the corresponding data residuals, only Marchenko-Lippmann-Schwinger inversion shows any improvement in data fit, and this improvement is marginal at best. After 50 iterations of inversion, the improvement is undone and the data misfit is worse than the starting model. Born and Lippmann-Schwinger inversions both show a worse data fit at all iterations of inversion, with Born being the worst performing by this metric.

Moving onto inversion with Tikhonov damping, figure 5.7 shows the resultant models after a single iteration of inversion. The level of applied damping and the scaling applied to the model update (a.k.a. the step-length in a gradient descent)

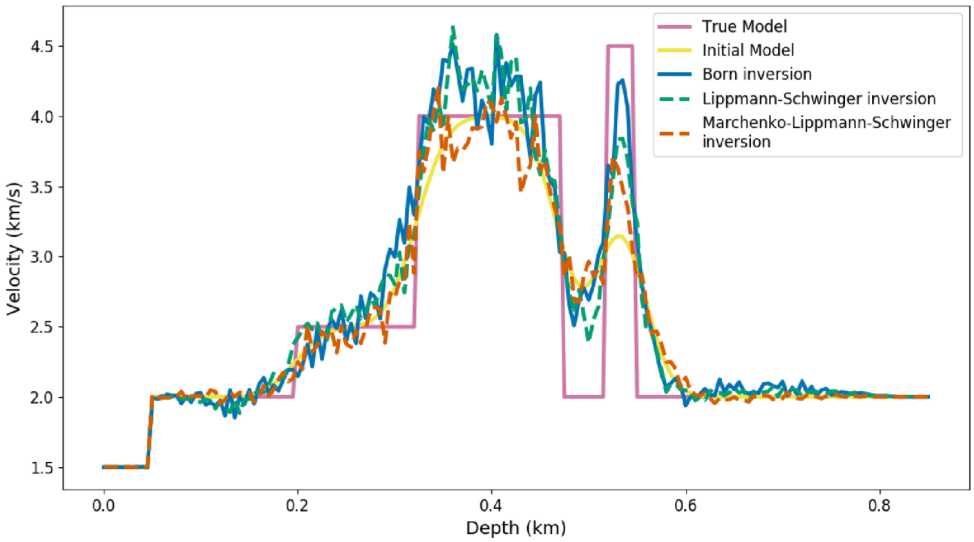


Figure 5.4: Velocity models after 50 iterations of inversion without regularisation.

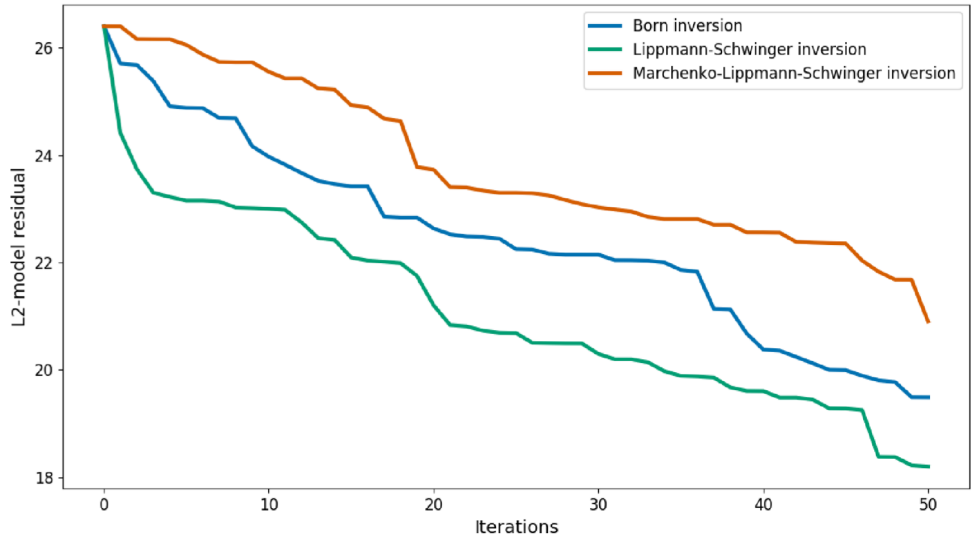


Figure 5.5: Model misfit trend over 50 iterations of inversion.

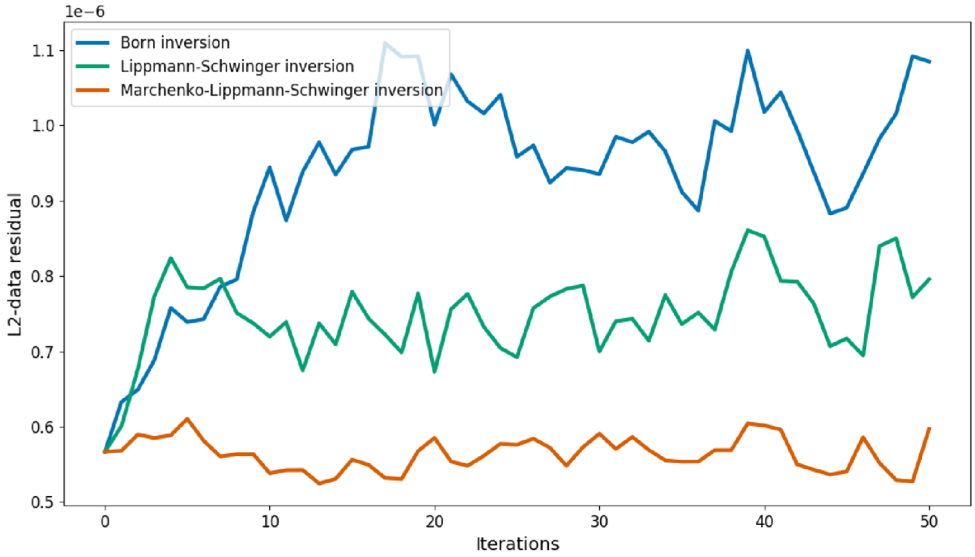


Figure 5.6: Data misfit trend over 50 iterations of inversion.

have been optimised independently. Marchenko-Lippmann-Schwinger inversion shows the least change compared to the initial model, with the least improvement in the velocities in layers 5 and 6, and a bulk over-correction of the velocity in layer 4. Born inversion results in a greater over-correction in layer 4 than Marchenko-Lippmann-Schwinger inversion, but offers a slight improvement over Marchenko-Lippmann-Schwinger inversion in layers 5 and 6. Lippmann-Schwinger inversion shows improved velocities in layers 5 and 6 as compared to the Born inversion result, as well as a strong response to the interfaces above and below layer 4, and a weaker response at the interface between layers 2 and 3. Noise due to the inversion process has been largely eliminated by applying damping with all three methods. After 50 iterations with Tikhonov damping (optimised at each iteration), the resulting models are shown figure 5.8. The models from all three methods are now largely consistent with each other, with noise increased significantly for all inversion methods.

Looking at the updates in the model residual, shown figure 5.9, we see that the trends in the model residuals agree with our observations on the qualities of the inversion models. Lippmann-Schwinger inversion shows the most immediate improvement, with Born and Marchenko-Lippmann-Schwinger inversions lagging behind, and then all methods show consistent improvement with increasing iterations. From the data residuals, shown figure 5.10, we see that all inversion methods eventually show an

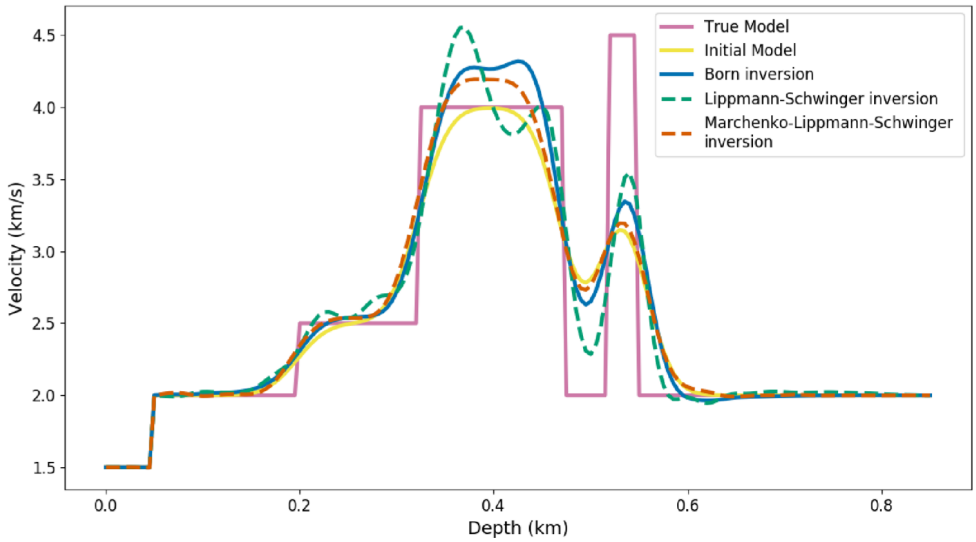


Figure 5.7: Velocity models after a single iteration of inversion regularised with Tikhonov damping.

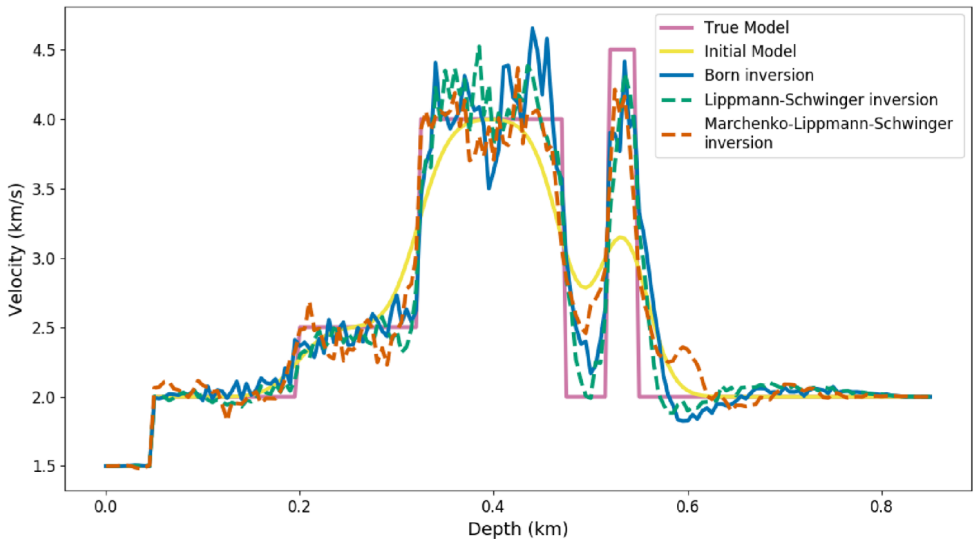


Figure 5.8: Velocity models after 50 iterations of inversion regularised with Tikhonov damping.

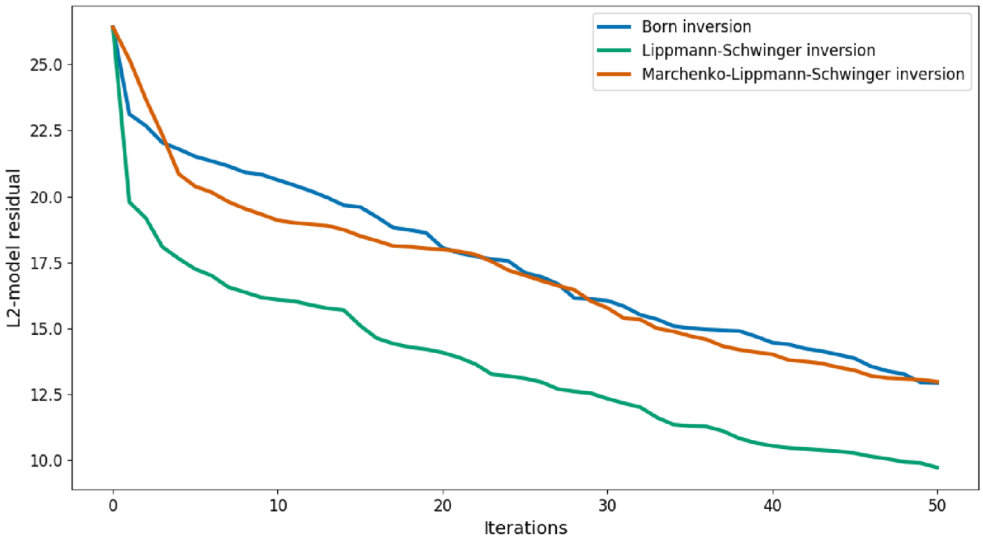


Figure 5.9: Model misfit trend over 50 iterations of inversion regularised with Tikhonov damping.

improvement in the data fit as well as the model fit after initially showing a worse fit. Lippmann-Schwinger inversion rapidly improves the data fit before flattening off, while Marchenko-Lippmann-Schwinger inversion slowly improves the data fit after initially moving in the wrong direction until iteration 16, at which point the data misfit begins to increase with increasing iterations. The Born inversion behaves similarly to the Lippmann-Schwinger inversion until iteration 11, at which point it suddenly increases before eventually falling into an exponential decay-like trend.

Lastly, the results of applying smoothing to all three inversion methods are shown in figure 5.11. All three models are largely consistent with the initial smooth model except around layer 4; here, Born inversion gives a smoothly-varying over-correction to the velocity, while the Lippmann-Schwinger and Marchenko-Lippmann-Schwinger inversions have a weak amplitude change on the interface between layers 3 and 4. Lippmann-Schwinger inversion squares out the round profile of the initial model in layer 4 with a noisy sine wave-like profile, whilst the Marchenko-Lippmann-Schwinger inversion has a much smoother profile. After 50 iterations with smoothing (figure 5.12), the resulting models bear great similarity to those without any regularisation, in that the models are universally dominated by spiking noise and largely adhere to the initial model in layers 1,2,3, and 7. The main differences to the unregularised

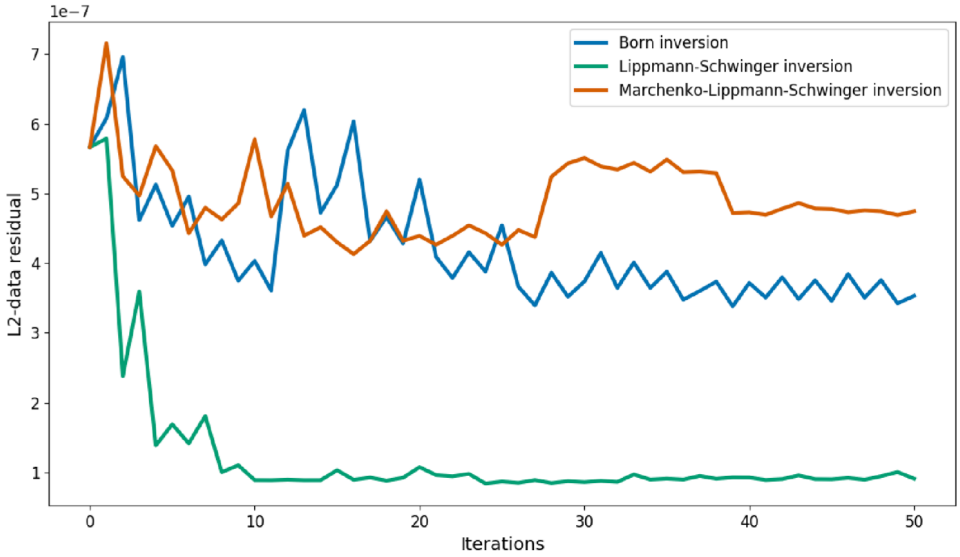


Figure 5.10: Data misfit trend over 50 iterations of inversion regularised with Tikhonov damping.

inversion appear in layer 6, with both Born and Lippmann-Schwinger approximating the thin layer with a high amplitude spike, while the Marchenko-Lippmann-Schwinger inversion remains close to the initial model.

Looking again at the metrics, we see in figure 5.13 that Marchenko-Lippmann-Schwinger inversion shows slow but consistent improvement in the model fit. In contrast, Lippmann-Schwinger inversion shows initial rapid improvement before tapering off, with the Born inversion falling between these two cases - initially improving quickly before eventually slowing down but overtaking the Lippmann-Schwinger inversion. In the data domain, figure 5.14 shows a completely different picture, with all three methods behaving erratically.

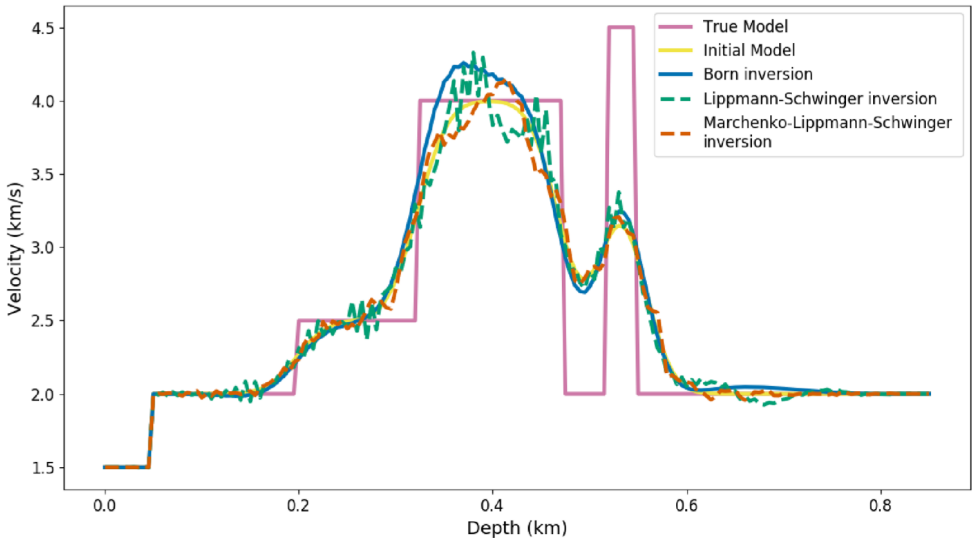


Figure 5.11: Velocity models after a single iteration of inversion with smoothing regularisation.

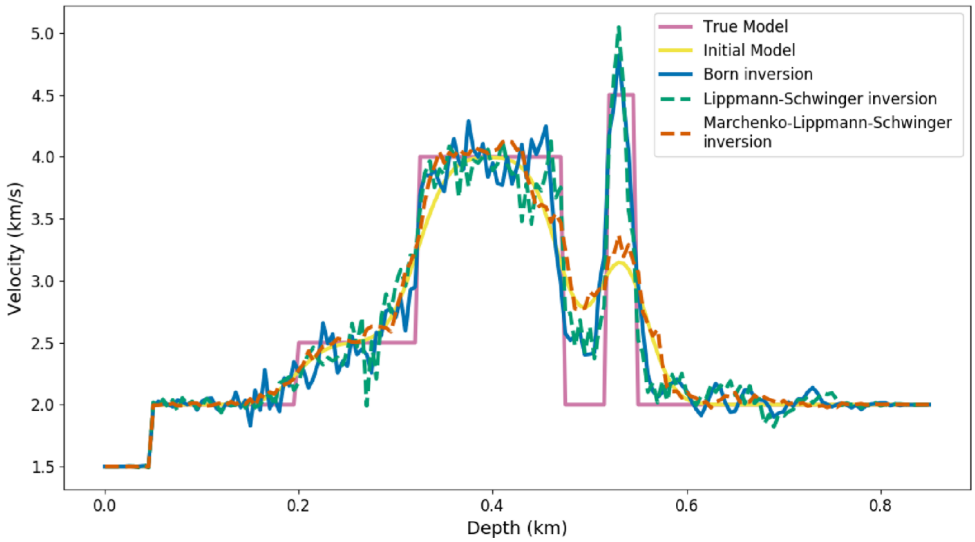


Figure 5.12: Velocity models after 50 iterations of inversion with smoothing regularisation.

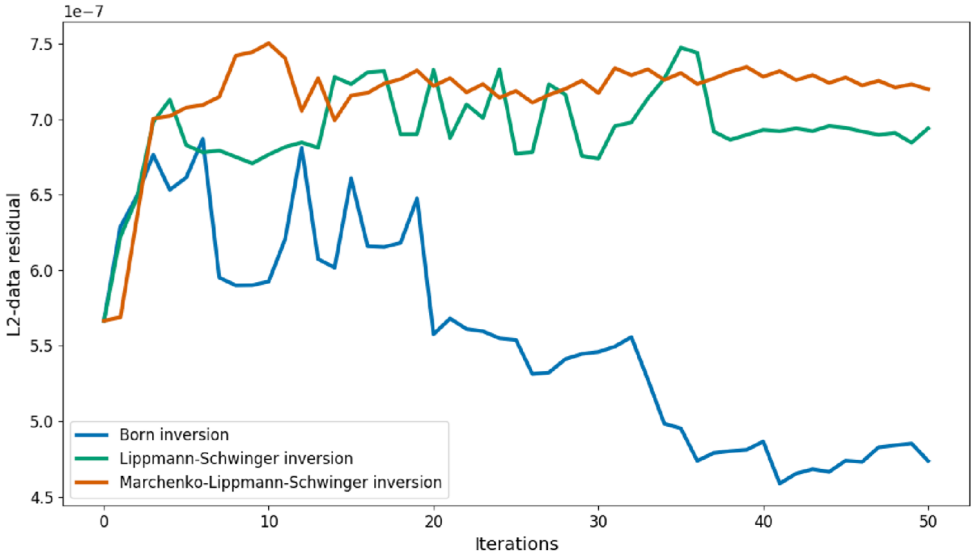


Figure 5.14: Data misfit trend over 50 iterations of inversion with smoothing inversion.

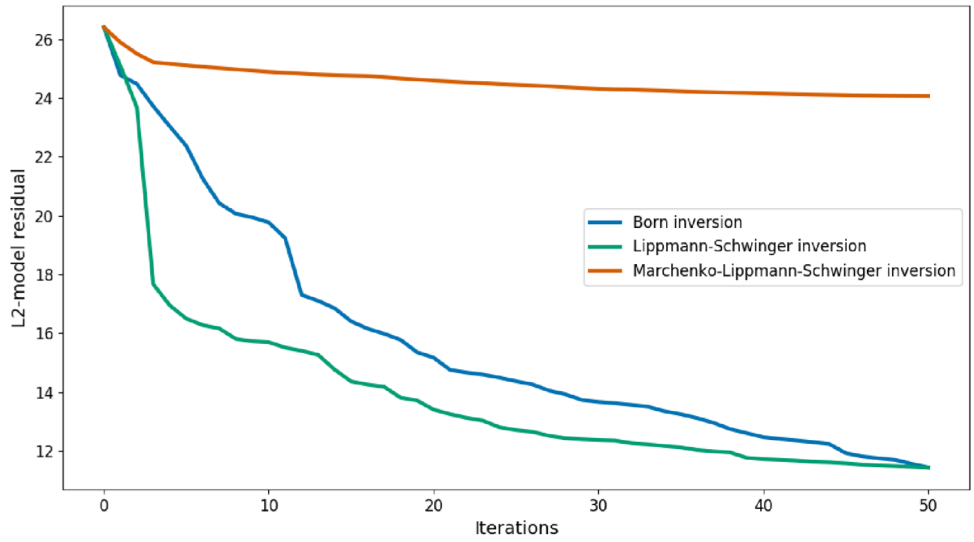


Figure 5.13: Model misfit trend over 50 iterations of inversion with smoothing regularisation.

5.2 Inversion optimised in the data domain

I now present the results for inversions optimised using the L2-norm of the data residual between the data from the true model and the data from the inversion model. As before, the regularisation level is optimised at each iteration, and the regularisation level and the step length are optimised independently.

Figure 5.15 shows the result of a single iteration of inversion without any regularisation when optimised in the data domain. Marchenko-Lippmann-Schwinger inversion shows a strong response at the interface between layers 3 and 4, however there is also significant noise added in the inversion process and it is unclear whether the lower interface between layers 5 and 6 has been identified or whether this is noise added to the velocity profile of the smooth initial model. There is also a strong spike immediately above the interface between layers 3 and 4. Given that there is also a spike immediately below the interface as well, this could either be a particularly strong instance of the spiking noise or an artefact in the inversion. It is possible that this is the result of an error in the Marchenko process, but it could equally be a location error of the type seen in the Born inversions of chapter 4. The Born inversion results in a similar profile to the initial model with noise added at a lower level than the Marchenko-Lippmann-Schwinger inversion, while the Lippmann-Schwinger inversion shows no visible difference to the initial model. As shown in figure 5.16, with iteration the noise added by the inversion has increased in magnitude across all three methods, but especially so with the Born and Marchenko-Lippmann-Schwinger inversions.

Looking at the data residual trend over all 50 iterations, figure 5.17 shows that the Born and Marchenko-Lippmann-Schwinger inversions, i.e. the methods that deviated most from the initial model, behaved similarly over the 50 iterations, going through periods of sudden decrease in the data residual followed by a slow decay period. In contrast, the Lippmann-Schwinger inversion models, which stayed close to the initial model, decreased slowly with a quasi-linear trend. In the model domain, the opposite picture is true, as shown in figure 5.18. Here, none of the resulting models are better fitting than the smooth initial model. As might be expected, since the Lippmann-Schwinger models deviated least from the initial model, those models showed almost

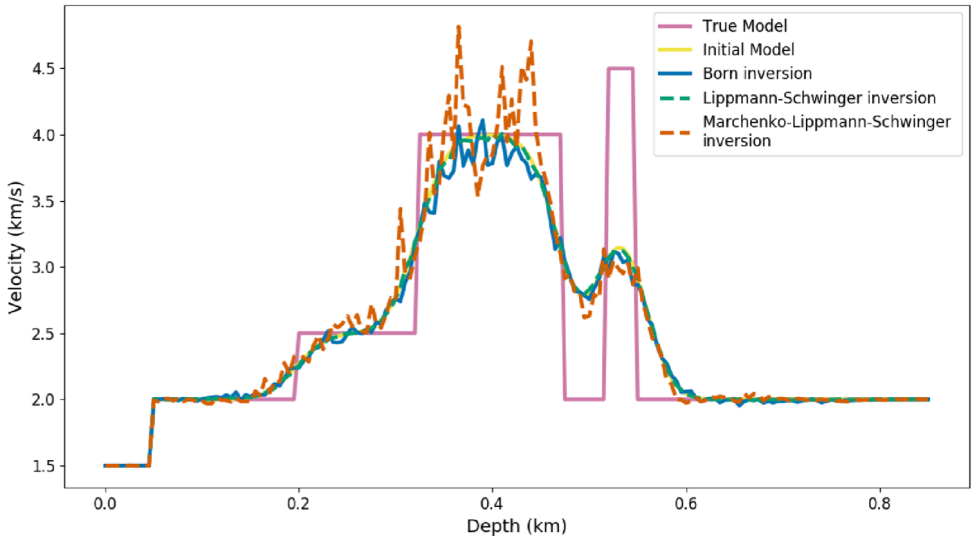


Figure 5.15: Velocity models after a single iteration of inversion optimised with the least-squares data misfit.

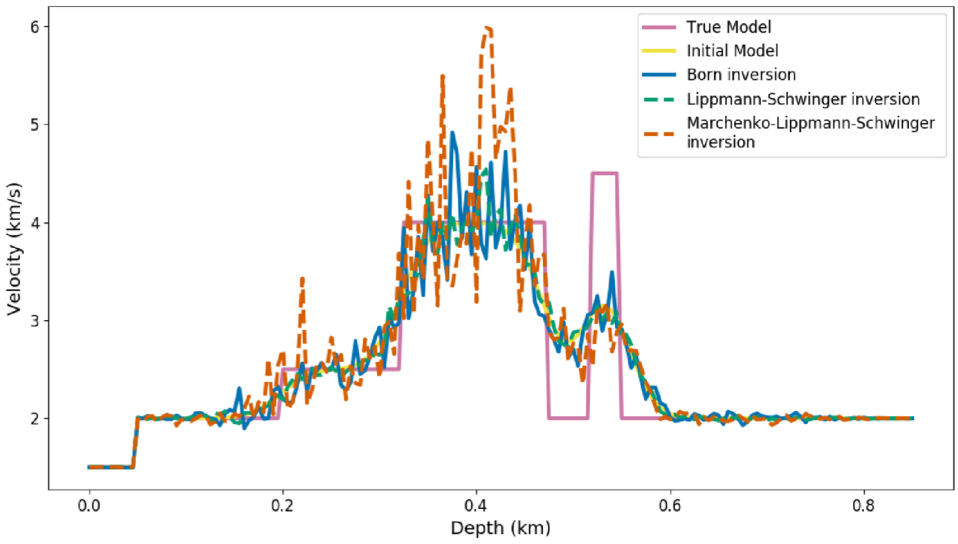


Figure 5.16: Velocity models after 50 iterations of inversion optimised with the least-squares data misfit.

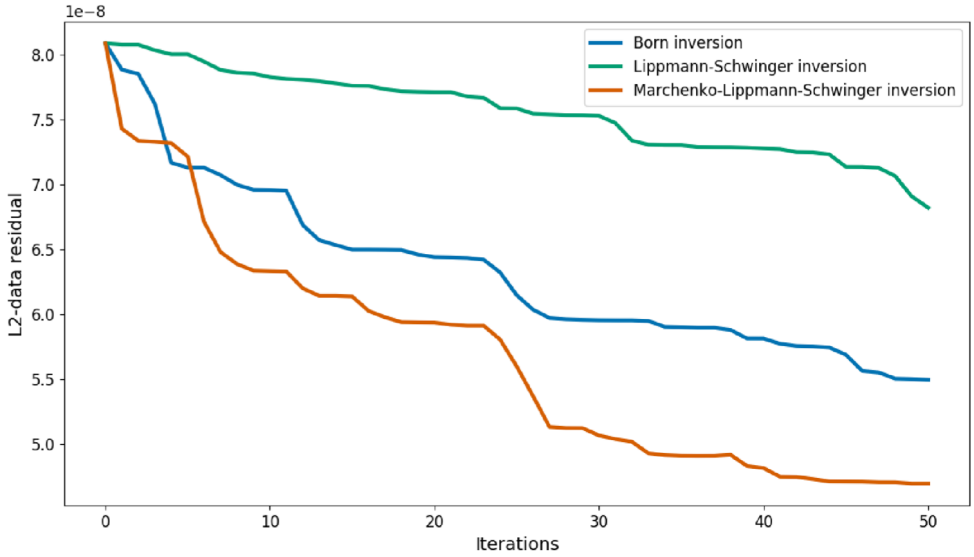


Figure 5.17: Data misfit trend over 50 iterations of inversion optimised with the least-squares data misfit.

no increase in model misfit. In almost contradictory fashion, the two noisy results follow very different trends in the model domain. The Born inversion models follow closely to the trend set by the Lippmann-Schwinger inversion models, which only slightly deviated from the initial model, whilst the Marchenko-Lippmann-Schwinger model misfit more than doubles over the first 21 iterations before continuing to grow at a slower rate for the remaining iterations.

Applying Tikhonov damping, figure 5.19 shows that Lippmann-Schwinger inversion is able to resolve thin layering with strong velocity contrasts, accurately retrieving the low velocity in layer 5 and improving the velocity estimate in layer 6 compared to the initial model, whilst simultaneously identifying two interfaces close to the upper and lower edges of layer 4. Born inversion shows a small improvement in the estimated velocities in layers 5 and 6, less than the improvement from Lippmann-Schwinger inversion, but also changes the velocity in the wrong direction in layer 4. Marchenko-Lippmann-Schwinger inversion is extremely noisy in layer 4, but otherwise deviates little from the initial model. After 50 iterations, there has been minimal change in the results except in layer 4. Born and Lippmann-Schwinger inversions have slightly lessened the overly large velocity estimates in layer 4, whilst the spiking noise in the Marchenko-Lippmann-Schwinger inversion has increased in amplitude, as

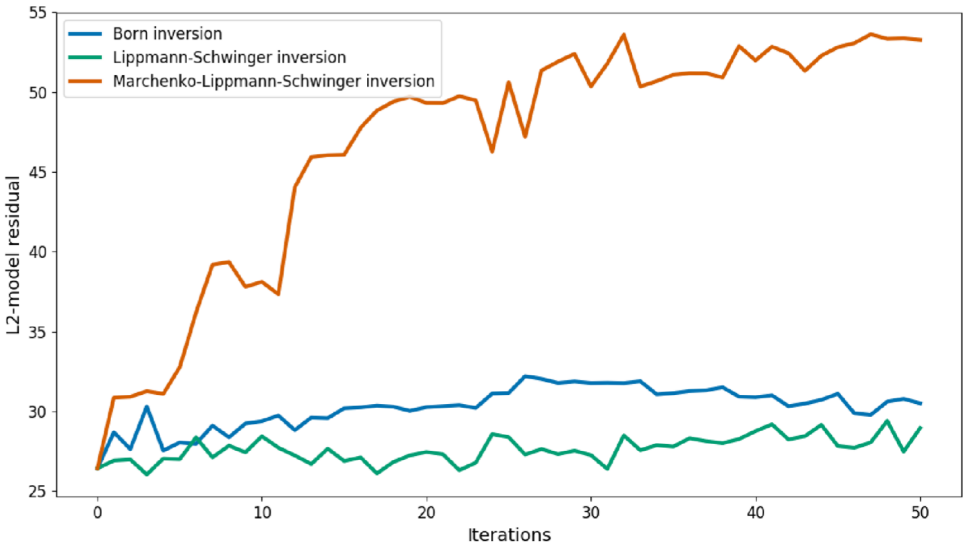


Figure 5.18: Model misfit trend over 50 iterations of inversion optimised with the least-squares data misfit.

shown in figure 5.20.

From the data residual trends, figure 5.21, we can see that all three methods showed initial reductions in the data misfit, including a 4-fold reduction for the Lippmann-Schwinger model and a 2-fold reduction for the Born model. Born and Lippmann-Schwinger inversions then abruptly halt after between 2 and 3 iterations, whilst the Marchenko-Lippmann-Schwinger inversion decreases the data misfit slowly across all iterations. Born and Lippmann-Schwinger inversions show similar behaviour in the data misfits, as seen in figure 5.22. After a first iteration improvement for the Born inversion model and a worse fitting model for the Lippmann-Schwinger inversion, the data misfit trends flatten after 2 to 3 iterations. The Marchenko-Lippmann-Schwinger model misfits increase with iteration, initially rapidly and then more slowly after 10 iterations of inversion .

Applying smoothing to the problem, figure 5.23 shows that the resultant models are largely consistent after a single iteration of inversion. All the inversion models follow a smooth trend with significant oscillatory noise added. Both the Born and Lippmann-Schwinger inversions compensate for lower velocities in layer 5 with significant increases to the velocity of layer 4, overshooting the true model velocity. The Marchenko-Lippmann-Schwinger inversion is extremely noisy, especially in

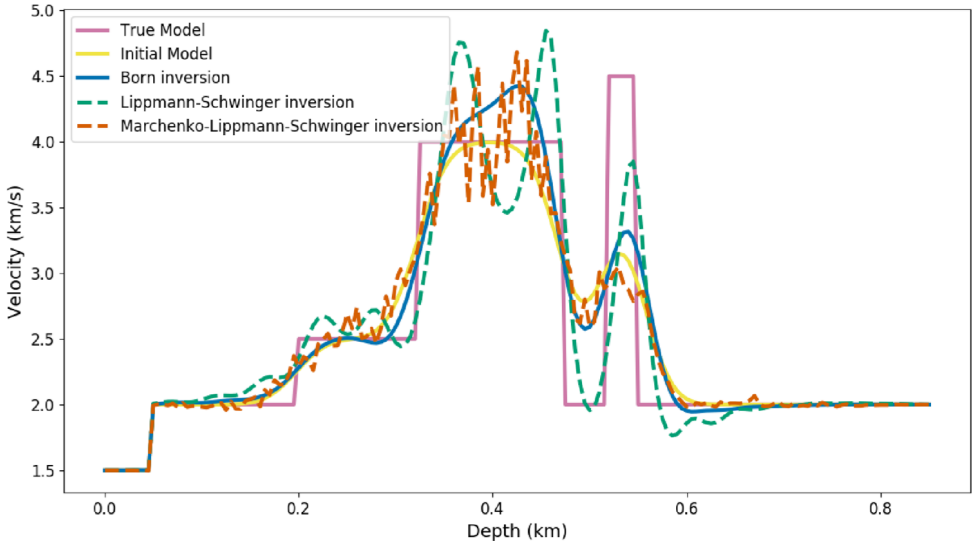


Figure 5.19: Velocity models after a single iteration of inversion, regularised with Tikhonov damping and optimised with the least-squares data misfit.

layer 4 where the amplitude of the noise is comparable to the amplitude of the true model velocity. The impact of iteration, as shown in figure 5.24, is markedly different in Marchenko-Lippmann-Schwinger inversion compared to Born and Lippmann-Schwinger inversions. With iteration, the noise level in the Born and Lippmann-Schwinger has been reduced so much that the inversion models could be described as essentially noise-free after 50 iterations. The Marchenko-Lippmann-Schwinger model is very different from the inversion model after 1 iteration. The noise in the lower half of the fourth layer has largely dissipated while the noise remains in the upper half. Noise has also increased in amplitude in the other layers it was present in the model after 1 iteration, and there is a large spike in layer 6. Given the noise level in the solution, it is difficult to determine if this spike is noise or the intended update direction.

Examining the data residual trend shown in figure 5.25, the behaviour of Born and Lippmann-Schwinger inversions is distinct from Marchenko-Lippmann-Schwinger inversion. Born and Lippmann-Schwinger inversion show immediate improvement in the data fit before ceasing to improve. In the case of Born inversion, which showed the greatest improvement, improvement ceases within 5 iterations, whilst Lippmann-Schwinger inversion essentially stops after the first iteration. Extraordi-

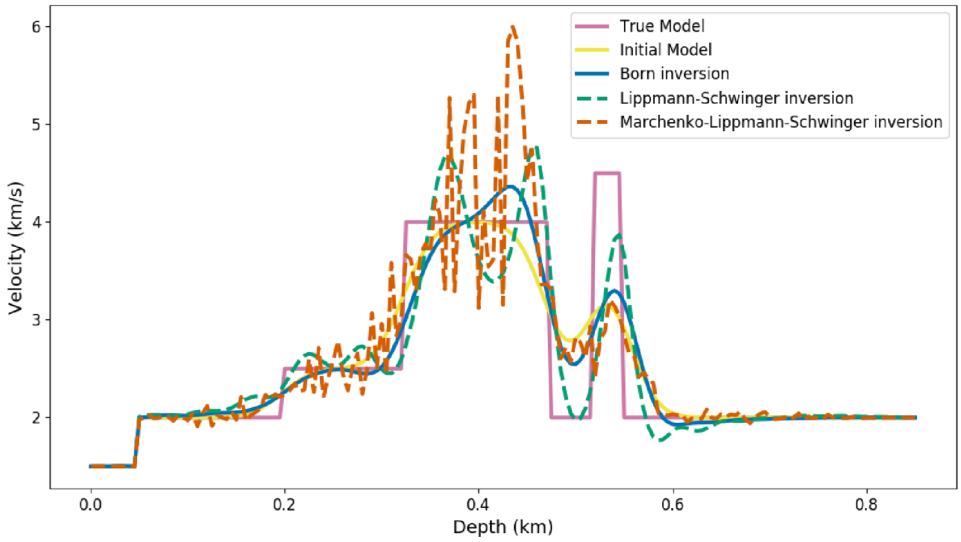


Figure 5.20: Velocity models after 50 iterations of inversion, regularised with Tikhonov damping and optimised with the least-squares data misfit.

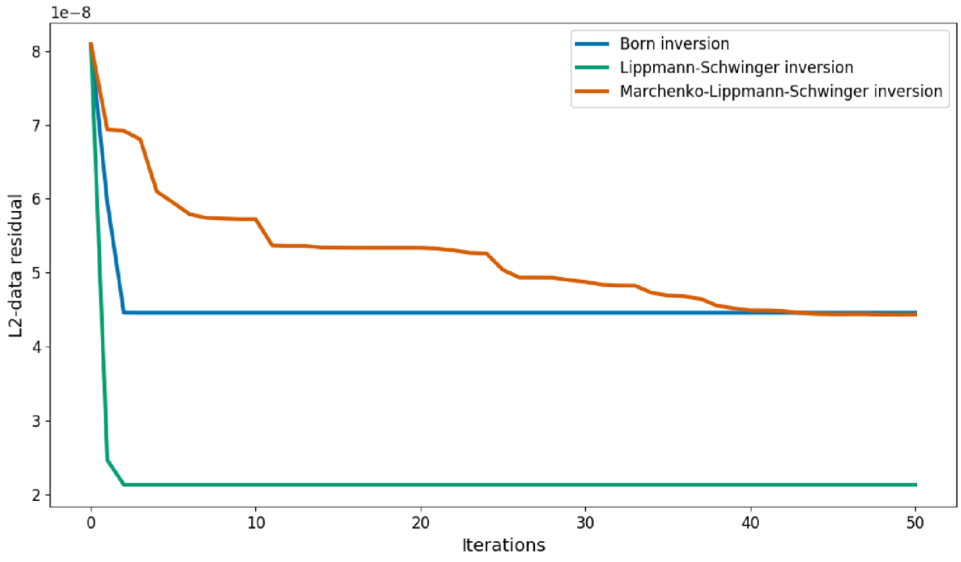


Figure 5.21: Data misfit trend over 50 iterations of inversion regularised with Tikhonov damping and optimised by the least-squares data misfit.



Figure 5.22: Model misfit trend over 50 iterations of inversion regularised with Tikhonov damping and optimised by the least-squares data misfit.

narily, Marchenko-Lippmann-Schwinger inversion outperforms Lippmann-Schwinger inversion by this metric, showing steady decrease in the data misfit over the first 15 iterations before ceasing to improve. In terms of the model misfit, figure 5.26 shows the model misfit increases for all inversion methods. Similarly to the data misfit, the trend can be separated into a variable phase followed by a stationary phase. Born and Lippmann-Schwinger inversions rapidly flatten out, while the Marchenko-Lippmann-Schwinger inversion model misfit stops changing after 15 iterations, coinciding with the cessation of change in the data misfit. The Marchenko-Lippmann-Schwinger model misfit increases rapidly, peaking at more than 3x the initial model misfit before settling at a lower value.

Looking at the data modelled in the final inversion models, figure 5.27 shows the data from models produced without any regularisation and with smoothing regularisation are comparable across all inversion methods. Only with Tikhonov inversion are we able to see a clear difference in the modelled data after inversion by the different methods. With Tikhonov damping, there is a clear improvement with Lippmann-Schwinger inversion in the fitting of the 3rd and 4th reflections (circled figure 5.27(middle)), with Born inversion achieving less accuracy in retrieving the amplitudes of these arrivals, and Marchenko-Lippmann-Schwinger inversion falling behind Born inversion.

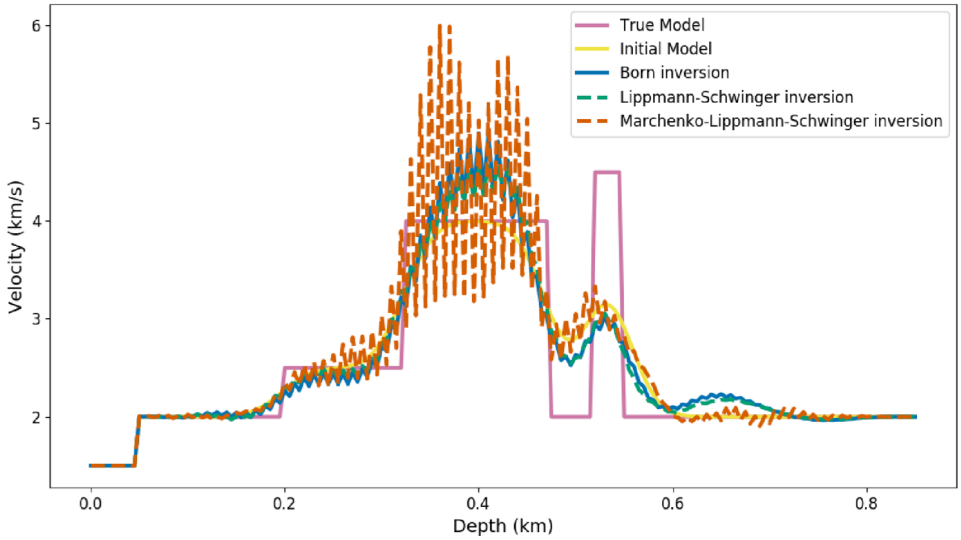


Figure 5.23: Velocity models after a single iteration of inversion with smoothing regularisation and optimised with the least-squares data misfit.

5.3 Discussion

Beginning with the inversions optimised in the model domain, from figures 5.3 and 5.4 it is clear that the unregularised inversions, both before and after the solution has been iterated, are significantly less noisy than those for the 1D slab model (figures 4.8 and 4.9). Importantly, the noise that does exist in the models shown in figures 5.3 and 5.4 appears to be correlated to the magnitude of the model misfit (figure 5.2), with less noise where the misfit is smaller (layers 2 and 3 of the true model), and more noise where the misfit is greater (layers 4, 5, and 6). In contrast, the noise pattern in the unregularised slab model inversions appears to be uncorrelated to the model perturbation. Prior information in the form of the initial model may have an important role in limiting the magnitude of this noise and in correlating the noise to the model misfit, even though the maximum amplitudes of the model perturbation in the many-layered model is twice that of the perturbation in the slab model, the noise remained sufficiently constrained in the many-layered case for the underlying model structure to remain visible. Another possibility is that artefacts due to the Gibbs effect interfere and cancel out. Thinking of the velocity model as we would a filter, each step in velocity results in spatial oscillations before and after the change in velocity. Since the boxcar model has only two reflectors, these oscillations are distributed throughout

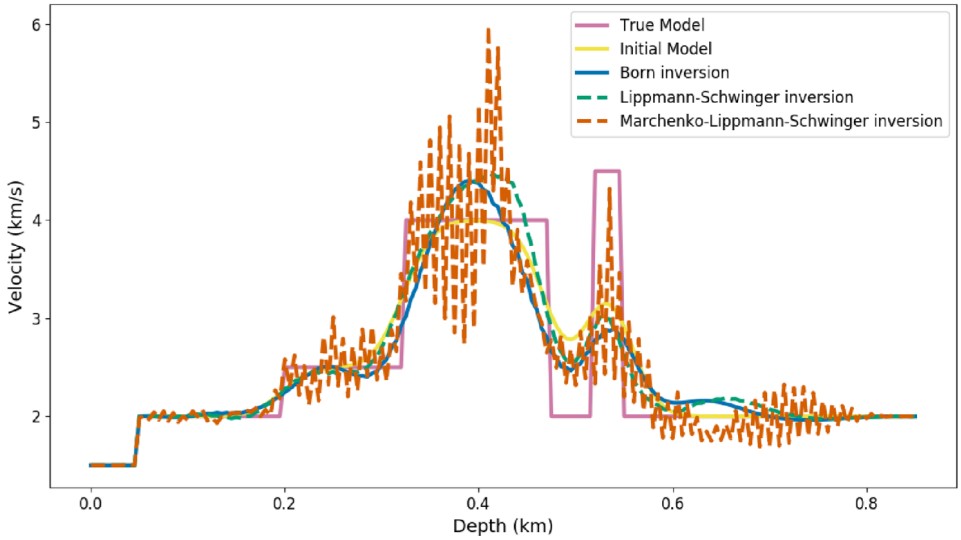


Figure 5.24: Velocity models after 50 iterations of inversion with smoothing regularisation and optimised with the least-squares data misfit.

the model unimpeded. These are added to the background velocity model and become their own model perturbations in future iterations of the inversion process, creating their own scattered waves. In contrast, the many-layered model has 6 reflectors, each of which creates spatial oscillations due to the Gibbs effect. These Gibbs oscillations may interfere to reduce the number and amplitude of oscillations carried through to the updated velocity model, whilst the closer spacing of the interfaces restricts the region in which the oscillations are unimpeded by the higher amplitude reflectors. These two effects, the addition of prior information and the mutual cancellation of oscillations due to the Gibbs effect, may be in effect individually or in combination with each other. Discriminating the individual effects of either cause is complicated due to the fact that reducing the amount of prior information by worsening the initial velocity model increases the size of the model perturbation, which in turn increases the amplitude of the Gibbs effect.

When regularisation is applied, we see that model retrieval improves in the many-layered case as well. Figure 5.12 shows that when smoothing is applied, we are still able to resolve high amplitude, non-smooth structures with both Born and Lippmann-Schwinger inversions when the inverse problem is iterated. Comparing the final inversion models with smoothing applied to the models after a single iteration (figure

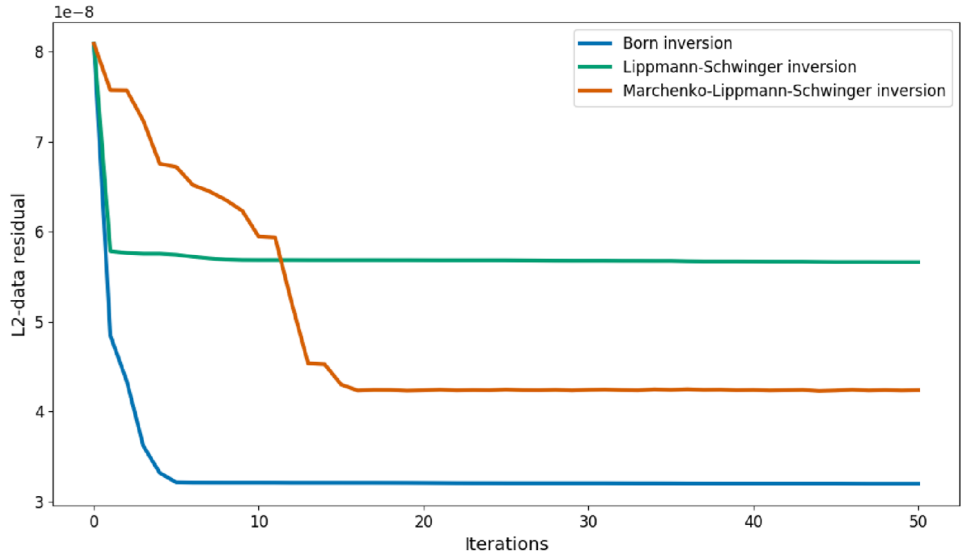


Figure 5.25: Data misfit trend over 50 iterations of inversion with smoothing regularisation and optimised by the least-squares data misfit.

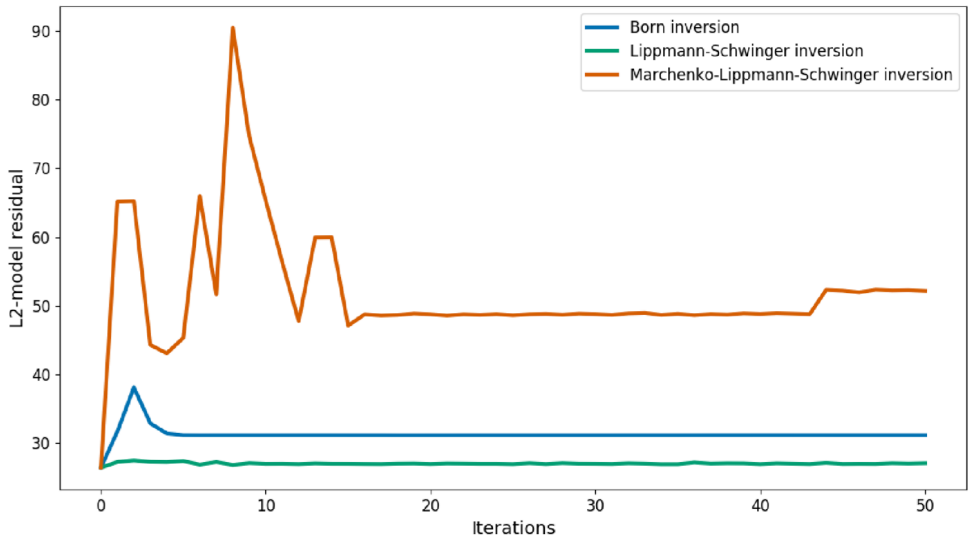


Figure 5.26: Model misfit trend over 50 iterations of inversion with smoothing regularisation and optimised by the least-squares data misfit.

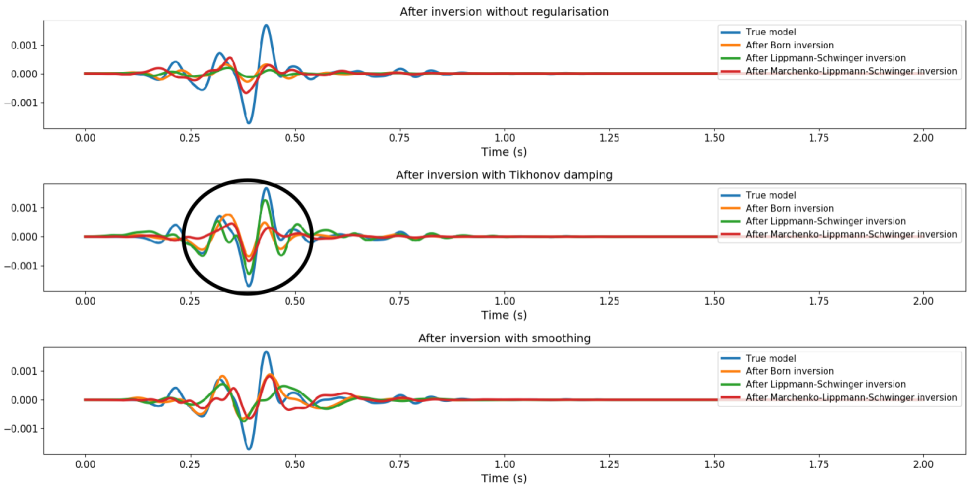


Figure 5.27: Comparison of data modelled in the true model to data modelled after 50 iterations of inversion (top) without regularisation, (middle) with Tikhonov damping, and (bottom) with smoothing.

5.11), as well as those models with no applied regularisation, we see that when regularisation is not applied or where smoothing is applied but the underlying model perturbation is weak, the tendency of the inversion result is to stick to the initial model trend rather than perturbing the structure. It is only in the deepest layers of the many-layered model, where the model perturbation is greatest, that the smoothed inversion is able to see the underlying structure. We can see that at some point in the iterative process, the optimal amount of smoothing to best fit the model has decreased as, especially with the Born inversion, there is a clear change from a smooth, noise-free velocity trend to a noisier, sharply-changing velocity structure.

Figures 5.7 and 5.8, showing the resultant models after a single iteration of inversion with Tikhonov damping and after 50 iterations of Tikhonov-damped inversion, we see that Tikhonov damping is superior to both smoothed and unreglarised inversion at determining both the amplitudes and non-smooth character of the true velocity model across all three methods of inversion tested. Looking closely at layers 2 and 3 in figure 5.8, we see that compared to smoothed and unregularised inversions, Tikhonov damping allows the step-function interface between the two layers to be observed, demonstrating that in my experiments Tikhonov damping performs better than smoothed or unregularised inversions when the model perturbation is small. This result may be true more generally than this specific instance, although testing with a

wider range of true and initial models would be necessary to prove this.

Looking at all the inversion models optimised in the model domain, by comparing the single iteration models to the models after many iterations of inversion, we see that iteration plays an important role in determination of the correct amplitudes in the models, even though iteration also leads to the amplification of noise in the models.

In terms of quantifiable metrics, figures 5.5, 5.9, and 5.13 show that for the many-layered model, Lippmann-Schwinger inversion always outperform Born inversion in terms of the model misfit regardless of whether regularisation is applied or not, and that when optimal Tikhonov damping is applied, Marchenko-Lippmann-Schwinger also outperforms Born inversion. That Lippmann-Schwinger and Marchenko-Lippmann-Schwinger inversion show the greatest reduction in model misfit over any other combination of method and regularisation agrees with our interpretation of the inversion models as being superior when optimal Tikhonov damping is applied.

When we compare the model misfit trends to the data misfit trends (figures 5.6, 5.10, & 5.14), we see that there is no clear correlation between model misfit and data misfit, indicating that the mathematically optimal model, when optimised using the model misfit, does not necessarily result in an optimal, or even superior, data misfit.

Returning now to data domain optimisation, from figures 5.3 and 5.4, it is clear that the 1D inversion without any applied regularisation is unstable, leading to high levels of noise in the inversion result regardless of the inversion method used. None of the tested inversion methods have updated the model away from the initial model. That is, to say, the noise amplitude is significantly larger than the causal model update, resulting in an optimal combined update that serves primarily to minimise the added noise. As with the model domain optimisation, iterating an already noisy velocity model results in the amplification of the noise but a velocity trend that largely matches the initial model.

The effect of smoothing regularisation on the many-layered model is minimal, as shown figure 5.23. With a single iteration of inversion, all three inversion methods have increased the model perturbation in layers 3, 4, and 6 of the true velocity model. Given the noise introduced in the inversion, it is clear that the optimal model update was undersmoothed at the first iteration, and thus additional smoothing would only

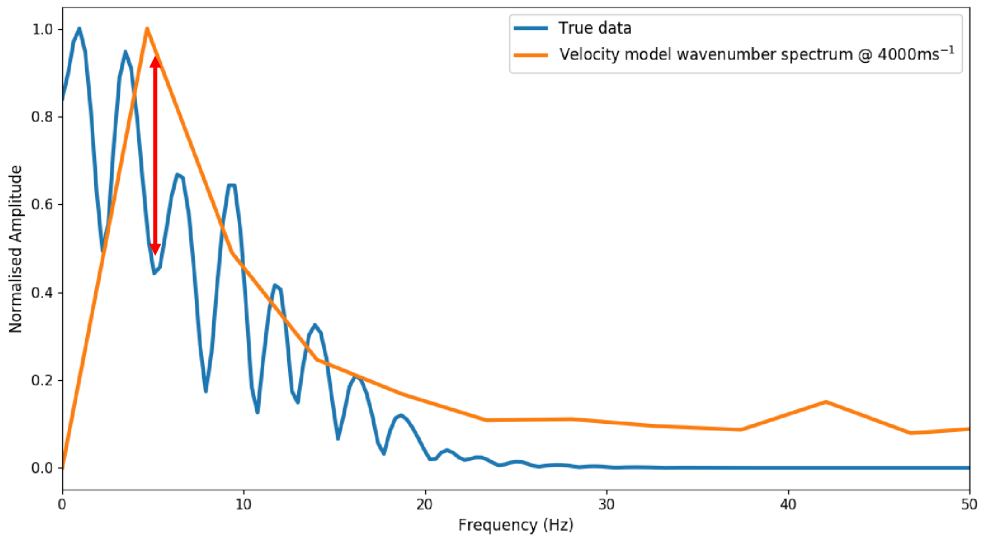


Figure 5.28: Comparison of the amplitude spectrum of the data from the true model to an amplitude spectrum derived from the wavenumber spectrum of the velocity model obtained after 1 iteration of Marchenko-Lippmann-Schwinger inversion with smoothing.

serve to further smooth the sharp interfaces of the underlying true velocity model. The regular periodicity of the noise in the velocity models suggests that there is a physical cause to the noise. To prove this, I transform the velocity model obtained after a single iteration of Marchenko-Lippmann-Schwinger inversion with smoothing from the space to the wavenumber domain by performing an FFT along the spatial dimension. The Marchenko-Lippmann-Schwinger inversion model is chosen because the noise amplitude is strongest in this model. The wavenumber amplitude spectrum can be converted to a temporal frequency by multiplication with an appropriate wavespeed, which allows the amplitude spectrum to be compared with that of the data. Figure 5.28 shows this comparison of the amplitude spectrum of the data from the true model to the amplitudes of the wavenumber spectrum evaluated at a velocity of 4000ms^{-1}, the velocity of layer 4 in the true model where the noise amplitude is strongest. The peak of the wavenumber spectrum overlaps with a notch in the data spectrum at 5.1Hz, shown with the red arrow, implying that the observed noise in the inversion may be due to attempting to invert the frequencies missing in the data, and that noise may be suppressed by design of a suitable notch filter a priori.

From a purely qualitative perspective, Tikhonov regularisation is the best performing method of regularisation. This is particularly true with respect to Lippmann-

Schwinger inversion, as can be seen in figure 5.19, with an excellent velocity match in layer 5, a greatly improved velocity in layer 6, and the identification of two sharp interfaces at the top and bottom of layer 4. Born and Marchenko-Lippmann-Schwinger inversions both make some improvements in squaring the profile of layer 4, but struggle elsewhere in the model. This assessment of the performance of Tikhonov-damped inversion is supported by figures 5.17, 5.21, and 5.25, which show that the Lippmann-Schwinger inversion with Tikhonov damping resulted in the biggest decrease in the data misfit of any combination of inversion method and regularisation.

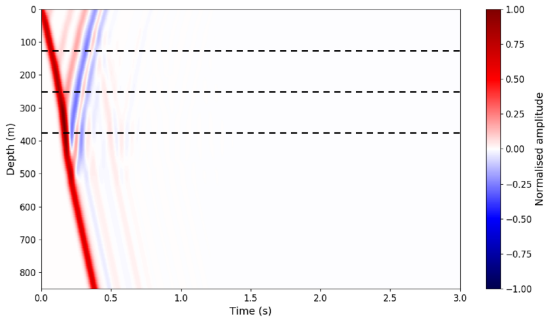
The impact of iteration when optimising the inversion with the data misfit appears to be greatly diminished when compared to optimisation with the model misfit. Figures 5.16 and 5.24 show that noise was amplified by iterating unregularised inversion and diminished by iterating smoothed inversion, in both cases the underlying velocity trends are unchanged from the results after a single iteration. This suggests that it is the repeated application of smoothing regularisation and NOT the iteration itself that is responsible for reduction of noise. Comparing the uniterated damped inversion (figure 5.19) to the iterated damped inversion (figure 5.20), there has been a small reduction in the overly large amplitudes at either side of layer 4 after iteration. In combination with figure 5.24, this suggests that repeated application of damping regularisation is able to correct amplitudes, where smoothing regularisation is not.

The similarity between the iterated and uniterated inversion models when regularisation is applied can be seen in figures 5.22 and 5.26, which show that the model misfit does not largely change after the first few iterations. This agrees with figures 5.21 and 5.25, which show that the data misfit follows the same pattern of behaviour.

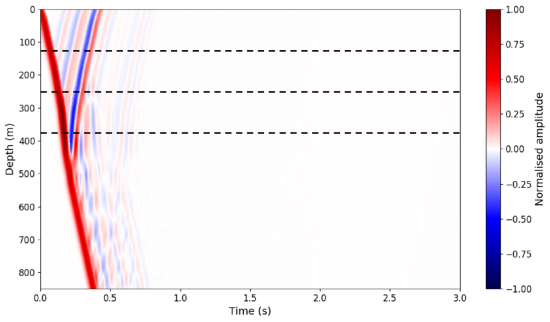
Comparing the wavefield in the true model and the Marchenko estimate of the true-model wavefield, figures 5.29a and 5.29b show that while the arrival times of the primary reflections are correctly estimated, the phase of the reflected wavefield is incorrect in the Marchenko estimates. The balance of amplitudes throughout the wavefield is also incorrect in the Marchenko-estimated wavefield. Figures 5.29c - 5.29e show comparisons between the true model Green's functions and the Marchenko estimates of the true model Green's functions at depths of 125m, 250m, and 375m. Displayed with normalised amplitudes, the imbalance of amplitudes in the Marchenko

estimates of the Green's functions can be seen across all times and depths. The sudden decrease in amplitude in the direct wave of the Marchenko-derived Green's function in figure 5.29d clearly demonstrates the end of the window function used in the Marchenko algorithm. With smoothing using a cosine taper, the Green's function match would likely have been improved at this location. The effects of applying a Marchenko window function that is too short can most clearly be seen in figure 5.29c, with the deformation of the expected Gaussian source signature particularly evident at 0.34s and 0.53s.

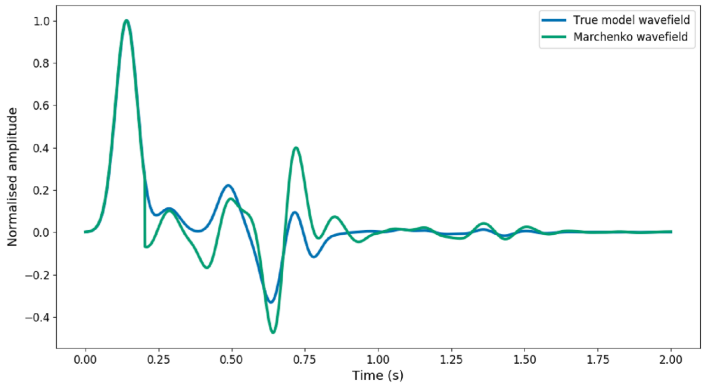
There are concerns about how the quantitative metrics of the data misfit and (in this specific experiment) the model misfit are used to determine what constitutes the "best" model. In the earlier model domain optimisation examples, it was seen that the models with the lowest model misfit did not correspond to the models with the lowest data misfit, and now in the opposite case we observe the same disconnect - the models which produce the lowest data misfit are not the models with the lowest model misfit. All this is to say that optimising the data misfit does not guarantee, nor is it necessarily likely, that the resulting model is a better fit to the (generally) unknown true model than the model you start with. In the absence of other forms of evidence, these metrics still offer an appealing method of discriminating between models, but should not be used blindly without looking at the model to consider if it makes sense to select that result. As an example of this, from figures 5.21 and 5.25, we can see that the Born inversion of the many-layered model with smoothing applied resulted in a very similar reduction of the data misfit to the best performing method, a Lippmann-Schwinger inversion with Tikhonov damping applied. If one were to look at the corresponding models, shown in figures 5.20 and 5.16 respectively, we see that the smoothed Born inversion has added almost no detail about the underlying structure compared to initial model, and done very little to correct the amplitudes of the initial model. In contrast, the damped Lippmann-Schwinger inversion has identified the locations of two high amplitude interfaces, found almost exactly the velocity of the layer below those interfaces, and then significantly improved the estimate of the next velocity layer. Clearly these two inversion models do not offer the same amount of insight into the underlying true velocity structure, but quantitatively in the data



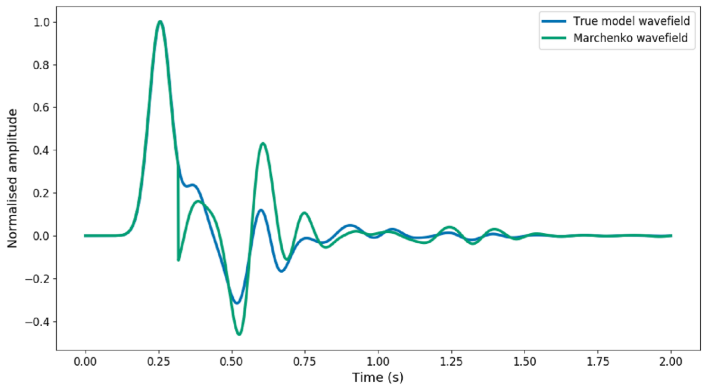
(a) True-model wavefield.



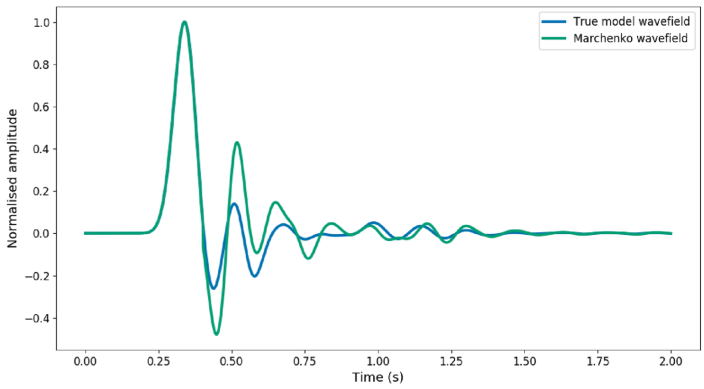
(b) Marchenko estimate of the true-model wavefield.



(c) Wavefields at z=175m



(d) Wavefields at z=250m



(e) Wavefields at z=375m

Figure 5.29: Comparison of the wavefield in the true model and its estimate using a Marchenko method before the first iteration of inversion. Both wavefields have been amplitude normalised for comparison. Depth sections (c)-(e) are marked by dashed black lines in panels (a) & (b).

domain, they essentially do.

To determine if there are any correlations in the model and data misfits, figure 5.30 shows cross-plots of the model and data misfits for every inversion at every iteration. The data points have been grouped by whether the inversions were optimised in the model domain (figures 5.30a, 5.30c, & 5.30e) or optimised in the data domain (5.30b, 5.30d, & 5.30f), and then clustered by inversion method (5.30c & 5.30d) or regularisation method (5.30e & 5.30f). When classed solely by the domain in which the inversions were optimised, the data points appear largely uncorrelated, especially the data domain optimisations. When clustered by inversion method, inversions optimised in the model domain form clear linear clusters (figure 5.30c), although the expected positive correlations in the data are not clear and weak at best. In the data domain (figure 5.30d), Born and Marchenko-Lippmann-Schwinger inversions appear uncorrelated, however a negative correlation is apparent in the Lippmann-Schwinger inversions. When clustered by the applied regularisation, the model-domain optimised inversions again group into linear clusters (figure 5.30e), whilst the data-domain optimised inversions (figure 5.30f) are widely scattered except for the unregularised inversions which are negatively correlated. To quantify the observable correlations in the cross-plots, figure 5.30g shows the Pearson's R-value for each of the clusters in panels (a) - (f), separated by optimisation in the model and data domains. Taking the inversions as a whole, model and data misfits correlate positively for model-domain optimisation and very weakly negative for data-domain optimisations. When categorised by inversion method, all three inversion methods show positive correlations for model-domain optimisation and negative correlations for data-domain optimisation. The negative correlations for the Lippmann-Schwinger and Marchenko-Lippmann-Schwinger are significantly greater than for the Born inversions. When categorised by regularisation method, figure 5.30g shows that unregularised inversion is negatively correlated regardless of how the inversion is optimised, whereas Tikhonov-damped and smoothed inversions correlate positively in both optimisation domains. This goes some way in supporting the claim that improving the misfit in one domain does not guarantee an improved misfit in another. One should always bear in mind that correlation does not imply causation, and that the observed trends

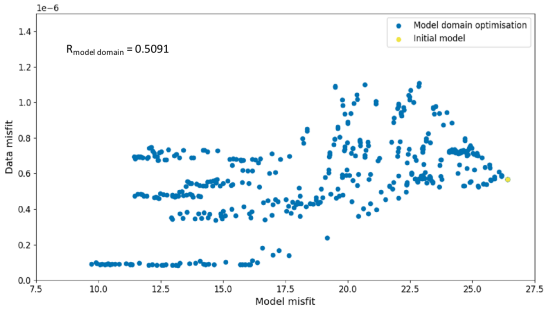
in the data and model misfits do not mean there is a fundamental physical connection between the two values.

Addressing the fact that not all inversions that optimised the model misfit also improved the data misfit and vice versa, it would be wrong to disregard the merits of the least-squares method based solely on this result. The deterministic least-squares method always leads to the solution that optimises some quality of the data i.e. our evidence that our model better represents the true model. In this case, that quality is the misfit between the data observed and the data modelled after inversion. It is only in a synthetic case where we already know the truth that we can even know this discrepancy between model and data misfits. In a real world example where the data is our only geophysical evidence, to disregard the data would be equivalent to choosing whatever model you can imagine as the underlying true model.

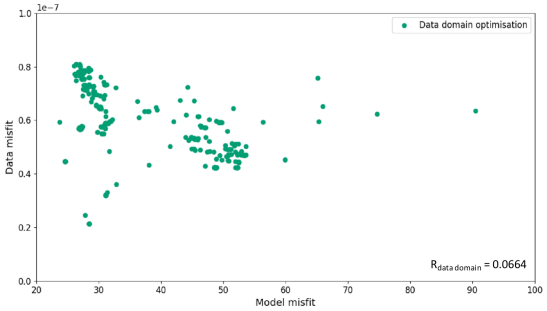
One possible explanation for the discrepancy between model and data misfits is that it is a consequence of the ill-posedness of the inverse scattering problem. Figures 5.5, 5.6, 5.9, 5.10, 5.13, and 5.14 show that for inversions optimised in the model domain, the data misfit only increases overall when the inversion is either unregularised or smoothing regularised. Given that Tikhonov regularisation is interpreted as showing superior performance compared to unregularised and smoothed inversions, I hypothesise that the mismatch between model and data misfit metrics is due to inadequate regularisation of the solution space.

Another possibility is that the model that optimises the model misfit is a statistically improbable but possible model in the space of all model solutions, one that happens to worsen the data misfit. It may be that by proper regularisation of the problem that the probability of the optimal model worsening the data misfit is minimised.

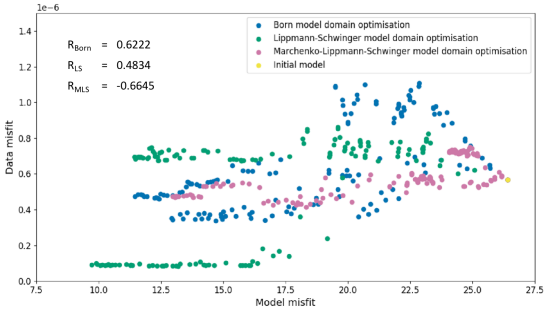
Evaluating the evidence assembled on 1D inversion as whole, it is clear that regularisation and available prior knowledge play an essential role in the one-dimensional inverse scattering problem. When prior knowledge is available in the form of a smooth background model, Lippmann-Schwinger inversion produces superior gradient directions to Born inversion, but that this improved gradient direction can only be expressed when the data misfit is not a factor - as the only piece of evidence we have in any real



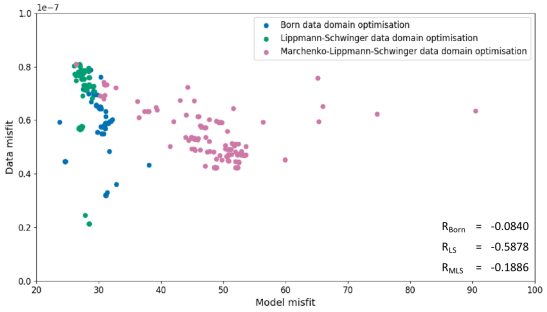
(a) All model domain optimised inversions.



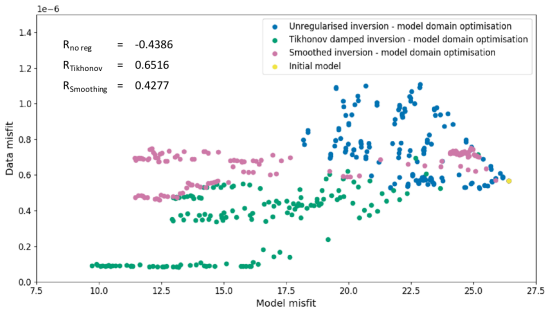
(b) All data domain optimised inversions.



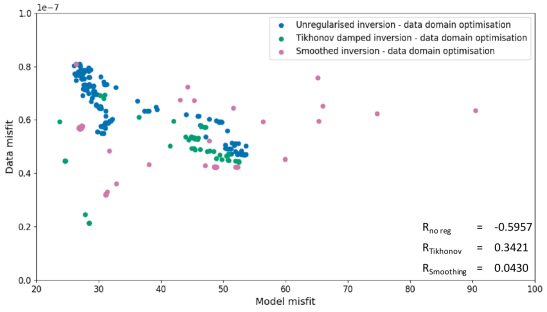
(c) Model domain optimised inversions clustered by inversion method.



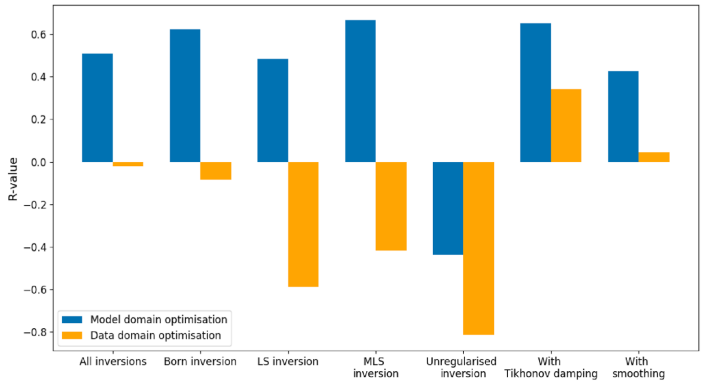
(d) Data domain optimised inversions clustered by inversion method.



(e) Model domain optimised inversions clustered by regularisation method.



(f) Data domain optimised inversions clustered by regularisation method.



(g) Pearson's R-value for each cluster in panels (a) - (f).

Figure 5.30: Variation of model and data misfits for inversions optimised in the model and data domains.

experiment, this is the same as saying to benefit from this better gradient direction, we require the results of the experiment we are performing before we have performed it. When the data misfit is the only information available to optimise the 1D inverse problem, one cannot accurately predict whether Lippmann-Schwinger inversion will perform better than the standard Born inversion with the appropriate regularisation applied.

When the performance of Marchenko-Lippmann-Schwinger inversion is compared to both standard Born inversion and Lippmann-Schwinger inversion, Marchenko-Lippmann-Schwinger inversion was the worst performing inversion in 4 out of 6 experiments as measured by the relevant metric, and in one of the cases where this was not the case (unregularised inversion optimised by the data misfit (figures 5.15 and 5.16)), not only does the magnitude of the added noise make accurate velocity estimation extremely difficult, but it is also fair to say that all inversion methods performed poorly in this experiment. In many regards, that inversion result is poorer than the initial model. The one instance where Marchenko-Lippmann-Schwinger inversion outperformed the other methods, the Tikhonov-damped inversion optimised using the model misfit, I find that the immediate improvement seen after a single iteration (notably in layer 6 in figure 5.7) is eventually achieved with iteration by both Born and Lippmann-Schwinger inversion (figure 5.8). Looking at the data misfit trend for that experiment, figure 4.24 shows the Born and Lippmann-Schwinger inversions both improved the data misfit as well as the model misfit within 3 iterations, a trend not generally seen when optimising with the model misfit. By contrast, it required 43 iterations of inversion of Marchenko-Lippmann-Schwinger to equal the data misfit of the initial model.

It is clear then that, even in one dimension, Marchenko-Lippmann-Schwinger inversion cannot perform comparably to Lippmann-Schwinger inversion. Since the only difference between the two methods is that Lippmann-Schwinger inversion uses full scattering Green's functions modelled using finite differences in the true model whereas Marchenko-Lippmann-Schwinger inversion uses estimates of the Green's functions retrieved using a Marchenko method, I must conclude that the factor prohibiting Marchenko-Lippmann-Schwinger inversion from performing equally to Lippmann-

Schwinger inversion is the difficulty in accurately reproducing the true model Green's functions using a Marchenko method. At present, the standard form of the Marchenko algorithm requires solving for the Green's functions to/from each model point to the surface sources and receivers individually. At each point, a number of different parameters must be optimised including the scaling factor between the amplitude of the (unknown) true model Green's function and the Marchenko estimate of that Green's function (van der Neut et al. (2015), Brackenhoff (2016), Dokter et al. (2019)), the number of iterations to perform, and the location of the windowing function that separates the Green's function domain from the domain of the focusing functions (Wapenaar et al., 2014b). The latter in particular can have a major impact on whether the retrieved Green's function is an accurate representation of the true model Green's function. It is clear that it is unreasonable in any synthetic experiment, where ground truth is knowable, to manually optimise the multiple parameters affecting the Marchenko algorithm at every individual point, and thus interpolation of parameters is required, resulting in a sub-optimal result at every point that has not been manually optimised. As a consequence, standard Marchenko Green's function retrieval results in Green's functions that may be inconsistent from point to point, lacking the natural "correlation" that real wavefields and wavefields modelled by finite-differences, for example, possess.

One possible solution to this last point would be a formulation of the Marchenko algorithm which, rather than solving for each point individually, solves for all locations simultaneously in a wavefield-based approach. This, in turn, will result in Marchenko-retrieved wavefields that are simultaneously optimised over all locations and thus should be self-consistent. Until such a time as this, or a similar, method becomes mainstream within the Marchenko research community, or until a novel superior methodology supersedes all Marchenko methods, I must conclude that Marchenko methods cannot reliably retrieve the Green's functions needed for model inversion, and thus that Marchenko-Lippmann-Schwinger inversion cannot improve on the standard Born inversion. For the remainder of this thesis, I will concentrate mainly on the comparison between Born inversion and Lippmann-Schwinger inversion, assuming that at some point in the future a suitable Green's function retrieval method will exist.

Chapter 6

Linear inversion of the Lippmann-Schwinger equation in two dimensions

In the previous chapter, I examined the 1D Born and Lippmann-Schwinger inverse scattering problem, and showed that in the absence of prior information, both problems are extremely ill-posed. When prior information in the form of a smoothly-varying inhomogeneous velocity model is available, the result can be improved with the application of regularisation, and of the regularisation methods tested, Tikhonov damping was found to be most effective. When the inverse problem is optimised using the misfit between the inversion model and the true model, Lippmann-Schwinger inversion was found to be superior to Born inversion regardless of the regularisation applied (or not). When the inverse problem was repeated optimising the data misfit between data from the true model and data from the inversion model, Lippmann-Schwinger inversion with Tikhonov damping was found to be marginally better than Born inversion with smoothing applied. The conclusion was that one cannot say Lippmann-Schwinger inversion will outperform Born inversion *a priori* for any given model.

In this chapter, I present results from the inversion of the Born and Lippmann-Schwinger equations for a 2D model. Despite the less-than-conclusive evidence that a two-dimensional inversion will perform better than a one-dimensional inversion, there

are reasons to be hopeful that it might. The first is that when the 1D inverse problem was optimised using the model misfit, the Lippmann-Schwinger inversion was found to be superior for all regularisation methods when applied to the many-layered model. That is to say, regardless of how the problem was regularised, Lippmann-Schwinger inversion found a more accurate gradient direction with which to update the initial model than the Born inversion.

The second is that the determinedness of the problem can be improved by moving from one dimension to two dimensions. For a surface-based source-receiver geometry, the one-dimensional inverse problem permits a single co-located source and receiver. Thus for each frequency, there is only a single data value available in the inversion process. When moving from 1D to 2D, the number of model parameters increases by a factor equal to the size of the x-dimension, but for a co-located source and receiver acquisition geometry, the number of data increases by a factor of the number of sources squared. So by choosing an appropriate number of sources, the problem can be changed from a formally under-determined to an over-determined one, and thus the 2D inverse problem may be more stable than the 1D problem.

The models I used for the 2D problem are shown in figures 6.1 and 6.2. The true and initial models for the 1D many-layered model respectively were taken from the z-axis of symmetry of these 2D models.

This chapter proceeds as follows: I will begin by verifying that the 2D Lippmann-Schwinger forward problem is valid for the combination of modelling software and model I have used. I will then show the results of Born and Lippmann-Schwinger inversion in two-dimensions.

For the results in this chapter, unless stated otherwise, 141 co-located sources and receivers were used with a horizontal spacing of 15m. Data were modelled at each source position for 3s, sampled every 0.5ms corresponding to 6001 samples, before subsampling 8x to give 751 samples at a final sampling rate of once every 4ms. This subsampling was done to save the amount of storage space required. Due to limitations in the amount of readily accessible active memory that is available on our computer servers for the inversion process, the number of (positive) frequencies was reduced from 376 to 5 whilst using all of the available sources and receivers.

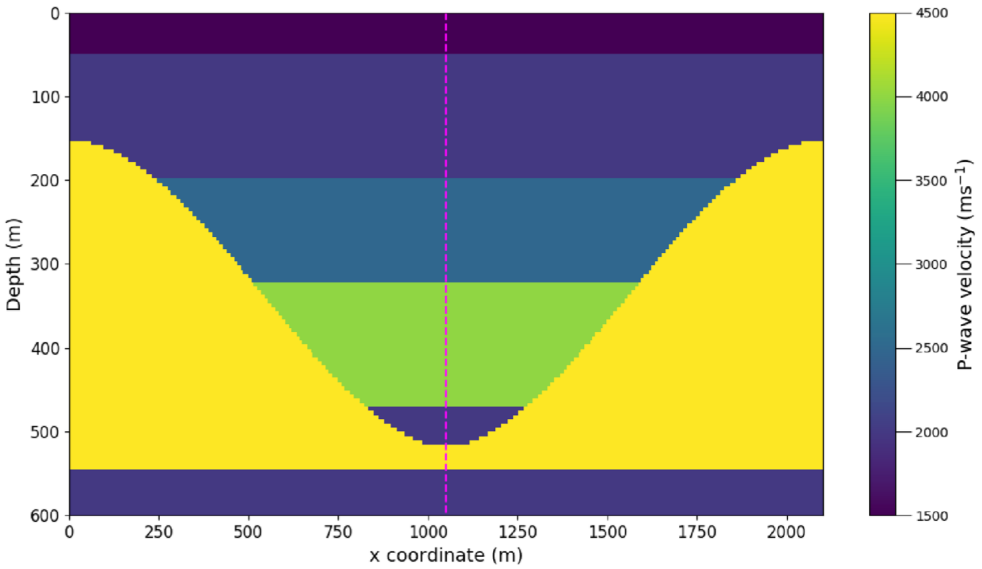


Figure 6.1: True model for the 2D inverse scattering problem. The true model for the 1D many-layered model in chapter 5 is taken from the profile shown by the dashed line.

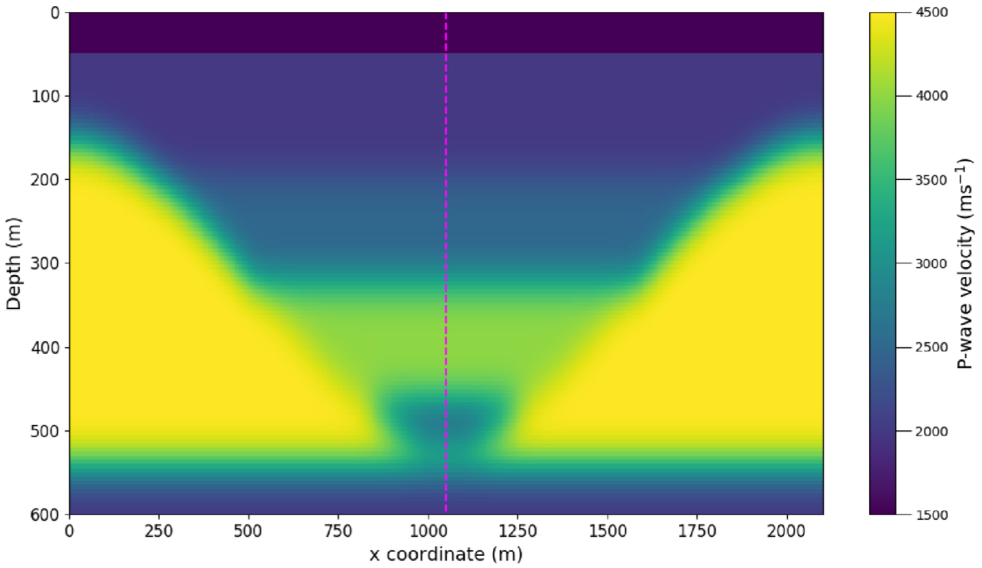


Figure 6.2: Initial model for the 2D inverse scattering problem. The initial model for the 1D many-layered model in chapter 5 is taken from the profile shown by the dashed line.

6.1 Verifying the Lippmann-Schwinger Green's function reconstruction in two dimensions

To test the accuracy of Lippmann-Schwinger forward modelling in two dimensions using my modelling software, I used a simplified version of the 2D syncline model which I used for testing the inversion methods. This model, shown in figure 6.3, is a horizontal layer cake model formed by taking the 1D profile shown by the magenta line in figure 6.1 (which is also the 1D many-layered model used in chapter 5), and extending it into the x-axis to the same dimensions as the 2D syncline model. With a single co-located source and receiver, this essentially reduces this test example to a 1D problem. If I had performed this test using the syncline model I used for the inversion tests, diagnosing errors would have been more difficult due to the introduction of acute-angle reflections from the sides of the syncline structure. The background model for this test is a homogeneous one, shown in figure 6.4, set to the velocity of the uppermost layer. Again, this was for reasons of simplifying the diagnosis of errors in the Lippmann-Schwinger forward modelling by eliminating errors due to incorrectly handled reflections in the smoothed syncline model. Source and receiver are placed at the surface in the centre of the x-axis (represented by a star and a triangle respectively in figures 6.3 and 6.4), thus the contribution to the Lippmann-Schwinger integral will be symmetric about the source-receiver location.

The result of Lippmann-Schwinger forward modelling in this setup is shown in figure 6.5. The data modelled in the true model by finite-differences has been convolved with the source function so that deconvolution of the source function from the Lippmann-Schwinger data is not required for the two signals to match. The match between modelled data and the data from the Lippmann-Schwinger reconstruction is exact for the initial 0.75s, after which there are two notable errors located in the black and red windows marked on figure 6.5.

Beginning with the black window, the arrival time of the artefact lies between the two-way traveltime from the source to the deepest interface and the two-way traveltime to the bottom of the model. This implies that this artefact is the result of constructive interference of the responses from the model points below the final

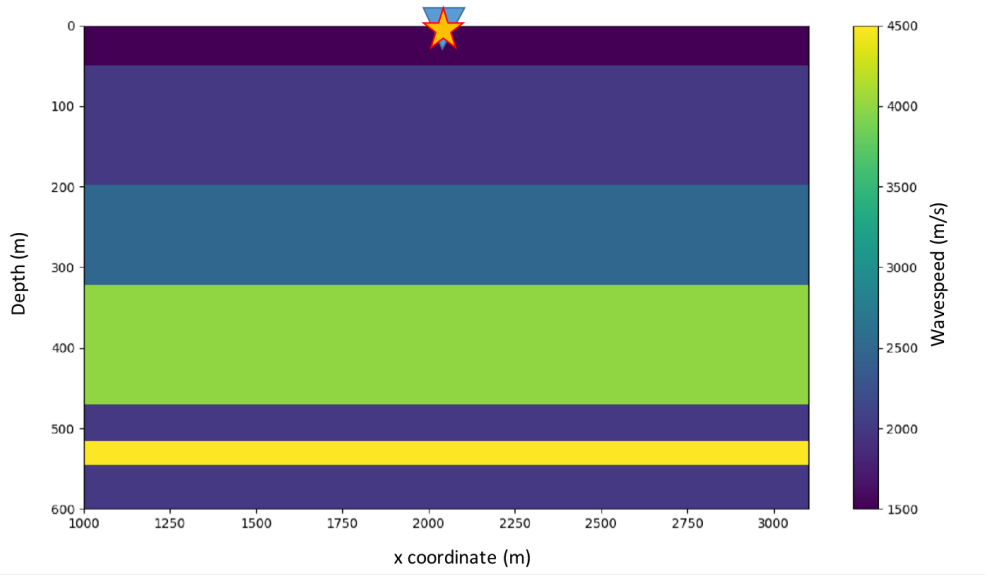


Figure 6.3: The true model for testing the accuracy of Lippmann-Schwinger data reconstruction in 2D, formed by extending the magenta depth profile in figure 6.1 in the x-direction. The single source and receiver location are indicated by the star and inverted triangle respectively.

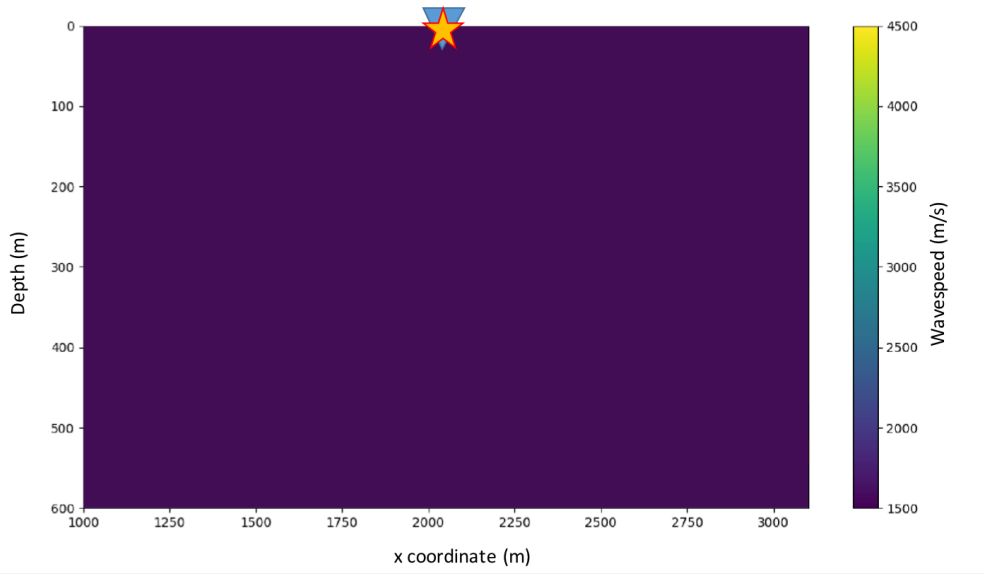


Figure 6.4: The background model for testing the accuracy of Lippmann-Schwinger data reconstruction in 2D. The single source and receiver location are indicated by the star and inverted triangle respectively.

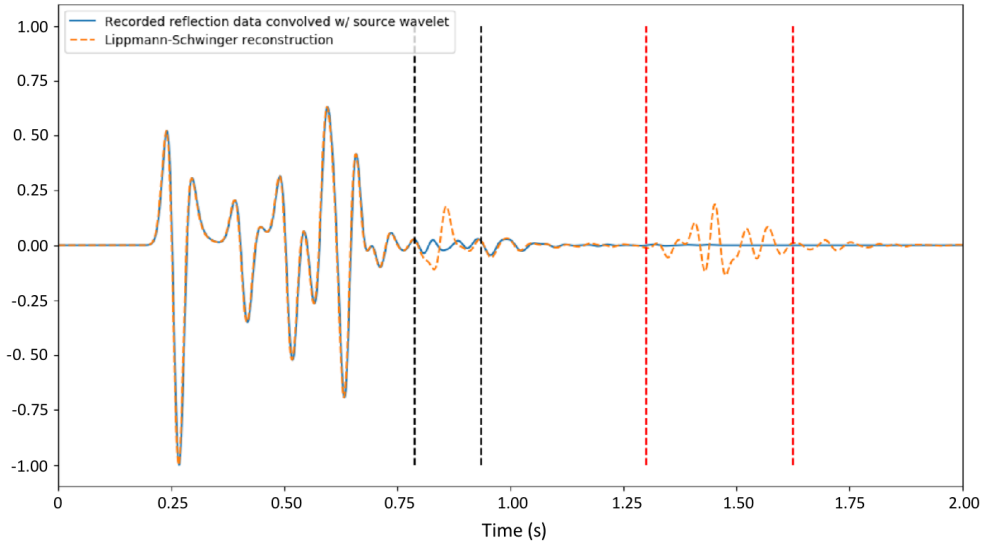


Figure 6.5: Comparison of the true model data modelled with finite-differences to the Lippmann-Schwinger reconstruction of the same data. Arrivals within the black and red dashed lines are artefacts from the Lippmann-Schwinger process.

reflector and above the bottom of the model. To test this, the velocity of the bottom layer of the model was changed from 2000ms^{-1} to 1500ms^{-1} so that the model perturbation below the final reflector is equal to zero. Solving the Lippmann-Schwinger forward problem again, the resultant data reconstruction is shown in figure 6.6. The reconstruction inside the black window is now exact, proving that errors are produced by the Lippmann-Schwinger equation when the model perturbation is not zero at the edge of the model.

Moving onto the artefact inside the red window, the relatively large time delay between the signal and the artefact implies the travel path of whatever is causing the artefact is significantly longer than that of the signal. Given the symmetry of the model and acquisition geometries, we would expect that the off-centre contributions to the integral would cancel each other out, leaving edge effects from the finite-difference modelling as the most likely candidate. With the delay on the artefact being so large, it is most likely not a result of a reflection from the bottom of the model, which lies on the short axis of the model. This means that edge effects from the sides or corners of the model are most probable.

To test this , I plotted the kernel of the Lippmann-Schwinger equation (equations 2.24 & 2.65) as a common receiver gather in the time domain, with the sources located

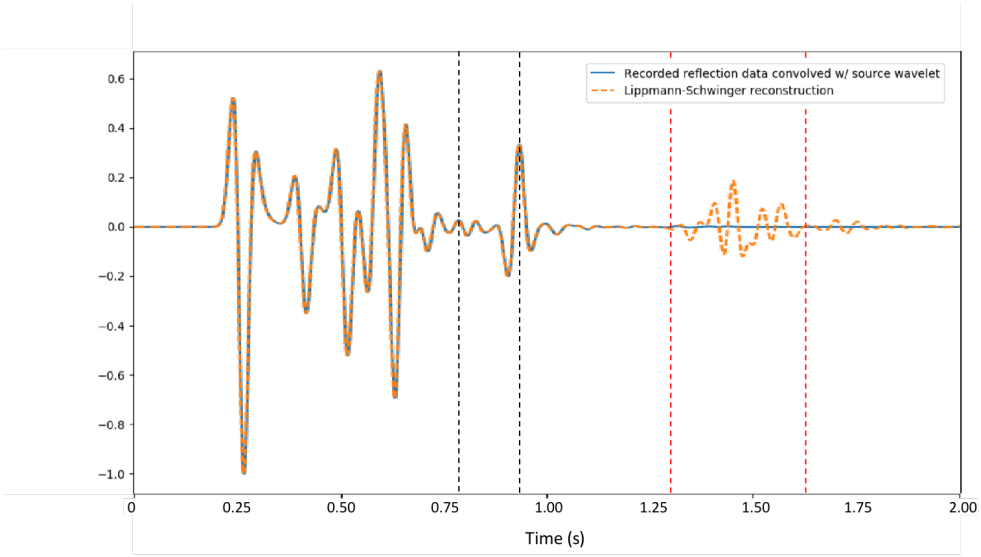


Figure 6.6: Comparison of the true model data modelled with finite-differences to the Lippmann-Schwinger reconstruction of the same data when the bottom layer velocity in the background model is exactly that in the true model. The black and red windows are the same as those in figure 6.5.

along the line path from the top left corner of the model, to the bottom left, then bottom right, and then the top right corner of the model. This plot, figure 6.7, shows that the amplitude of the kernel decreases with time for locations on the bottom edge of the model. For points on the left and right sides of the model, there is an increase in amplitude at times within the red time window in figure 6.5, which confirms that this is the origin of the observed noise. Given that the kernel of the Lippmann-Schwinger equation consists of the convolution of two Green's functions (scaled by the value of the model perturbation), and that the noise appears continuous across sources rather than random, we can conclude that this artefact is due to constructive interference of causal signals on non-physical raypaths. Thus approaches like stacking will not remove this artefact.

The approach I chose to remove this artefact is to apply a spatial cosine taper to the edge of the model. To preserve the model structures, considering the relatively small size of model I am using, this taper is applied by extending the model on the left, right, and bottom edges of the model in a homogeneous manner in the normal direction using the edge values, by the length in pixels of the taper to be applied. This approach, demonstrated in figure 6.8, minimises the number of additional pixels

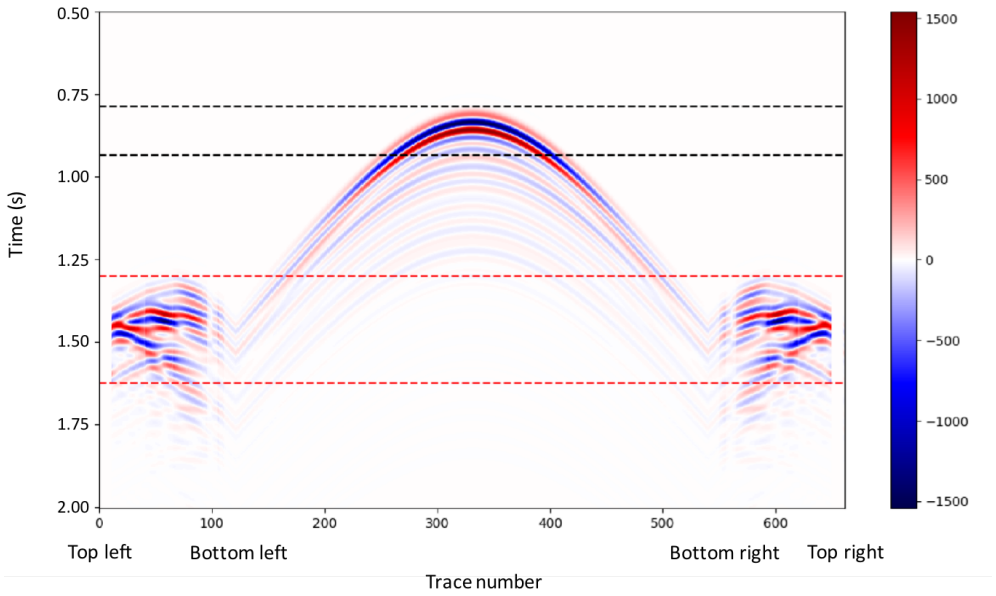


Figure 6.7: The kernel of the Lippmann-Schwinger modelling equation plotted in the time domain as a common receiver gather, taking only the contributions from model pixels on the left, bottom, and right edges of the model. Trace number increases from 0 in the top left corner of the model to 660 in the top right corner. The black and red dashed lines correspond to the same lines in figures 6.5 and 6.6.

in the kernel that will have to be inverted later. Figure 6.9 shows the impact on the Lippmann-Schwinger reconstruction of the data for varying amounts of tapering applied to the model edges. As the length of the taper increases, the amplitude of the artefact decreases and the phase moves closer to that of the true signal. Based on the results shown in figure 6.9, I chose to implement a 50 pixel length (250m in grid space) taper on the edges of the model. Since it is also clear that tapering the edges of the model actually mitigates both artefacts created by the Lippmann-Schwinger modelling process, it is unnecessary to change the background velocity model to remove the artefact in the black window associated with the velocity error.

The equal effect of model tapering and setting the model perturbation equal to zero at the model edges is a consequence of the Sommerfeld radiation condition (equation 2.5), which is needed to derive the Lippmann-Schwinger equation. This condition requires that the amplitude of the incoming wavefield decays to zero in the limit that the distance of the incoming wavefield approaches infinity. In the context of this example, the incoming wavefield is energy scattered back towards the surface acquisition array. Since all model perturbations act as scatterers in the Lippmann-

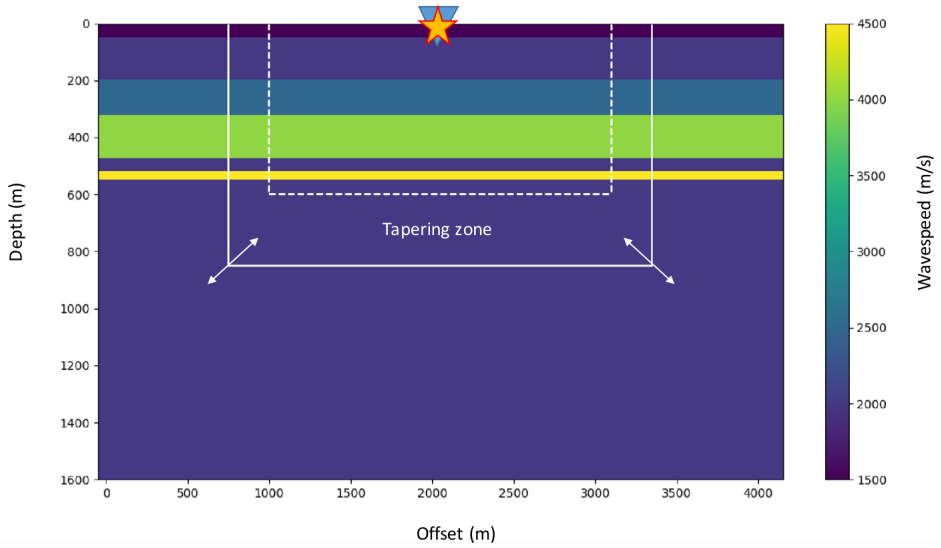
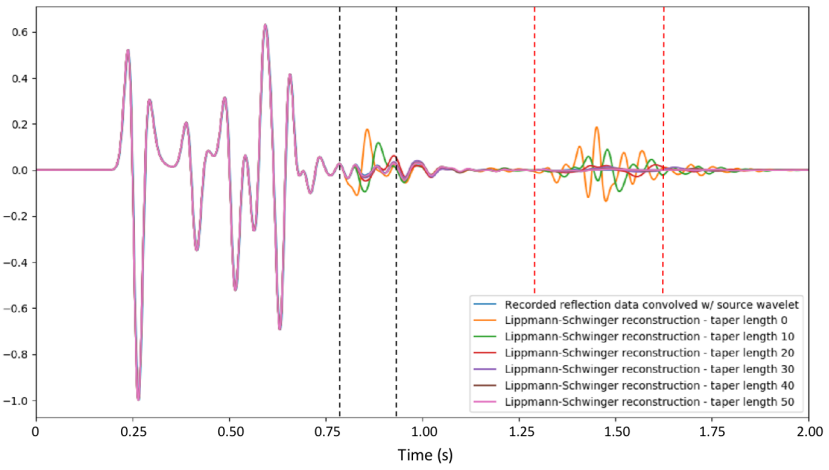


Figure 6.8: Schematic of the spatial tapering applied to the model. The original model area, enclosed by the dashed line, remains fixed. The model area is then extended homogeneously using the edge values of the model to the solid white line, with the taper running from 0 at the solid line to 1 at the dashed line. The taper length can be increased by increasing the amount the model is extended up to any taper length required (the effect on the solid boundary is indicated by the double-headed arrows).

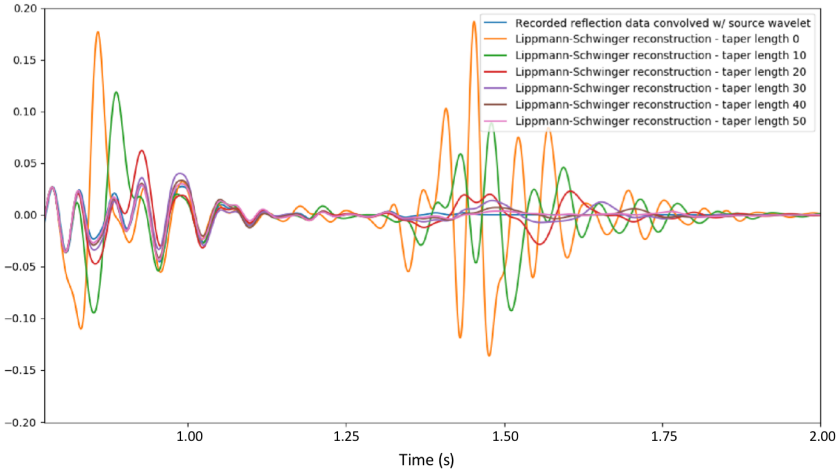
Schwinger description of scattering, the model perturbation below the final interface that caused the artefact in the black window and the edge effects causing the artefacts in the red window are both due to the violation of the Sommerfeld radiation condition. Setting the bottom layer velocity equal to the background model velocity works for the same reason that tapering does: it sets the amplitude of the incoming wavefield to zero at the edge of the model.

6.2 Born inversion versus Lippmann-Schwinger inversion in two dimensions

In the previous section, I showed that the modelled data can be reconstructed using the Lippmann-Schwinger equation when sufficient model padding and tapering are applied; in this case padding 50 pixels or 250m to the left, right, and bottom of the model. This results in an increase in the model size from 50,941 pixels to 89,091 pixels, or a 75% increase in the number of pixels. Any increase in the model size



(a) Data from the first two seconds of recording, the black and red windows are the same as those in figure 6.5.



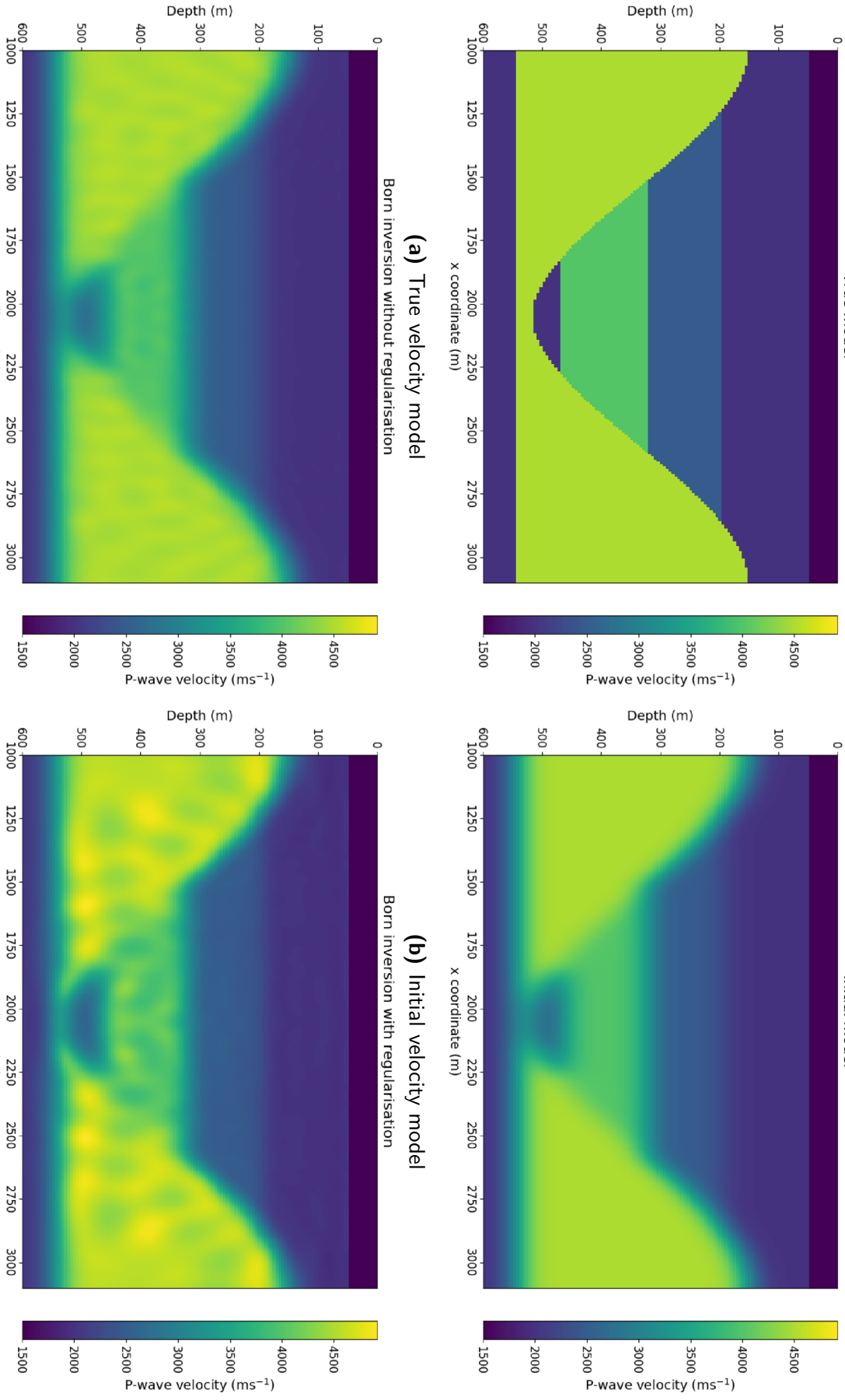
(b) Close up of the data in (a) affected by artefacts from Lippmann-Schwinger modelling.

Figure 6.9: Comparison of the true model data modelled with finite-differences to the Lippmann-Schwinger reconstruction of the same data for various lengths of taper applied at the model edges.

results in an equally sized increase in the size of the kernel matrix. Given the active memory limitations, this allows 5 frequencies to be inverted simultaneously. I now show the results of those inversions for a single iteration.

Figure 6.10 shows the updated models after Born inversion, both without regularisation and with optimised Tikhonov damping. A mute has been applied to the entire first layer of the model to remove bias from large amplitudes close to the source locations when optimising the model update. Since the velocity of the first layer is assumed to be known, the update in the layer should be zero, so applying a mute should not affect the outcome of the inversion. The optimal damping coefficient of the values tested is found to be $8x10^8$ after optimising the step-length of a gradient descent by minimising the L2-norm of the model misfit. As stated previously in chapter 5, this is equivalent to determining the optimal gradient direction in which to update the initial model. The unregularised Born inversion shows little change when compared to the initial model, except for low frequency but low amplitude noise throughout the high velocity syncline layer and the 4th horizontal velocity layer within the syncline itself. With an optimal amount of damping applied, the before mentioned noise is again observed, but now with much higher amplitude, however this is compensated by improved recovered velocities in layers 2, 3, and 4, resulting in better layer definition, even if the interfaces themselves are not significantly sharper.

The corresponding results for Lippmann-Schwinger inversion are shown in figure 6.11. In this instance, the unregularised inversion outperformed inversions with any amount of applied damping. In fact, the model misfit increased with increasing levels of damping. As a consequence, the optimal model with damping applied is one that would likely be deemed "under-regularised" had the regularisation been determined by a method such as the L-curve (Hansen, 1992), and is very similar to the unregularised inversion model. The Lippmann-Schwinger inversions show significantly sharper interfaces than the Born inversions, especially on the sides of the syncline layer, and between layers 2 and 3, and layers 3 and 4 of the horizontal layering. This is highlighted by the orange arrows in figure 6.11. As with the Born inversions, the Lippmann-Schwinger inversions show significant noise within the syncline layer itself, however at a higher frequency than the Born inversions.



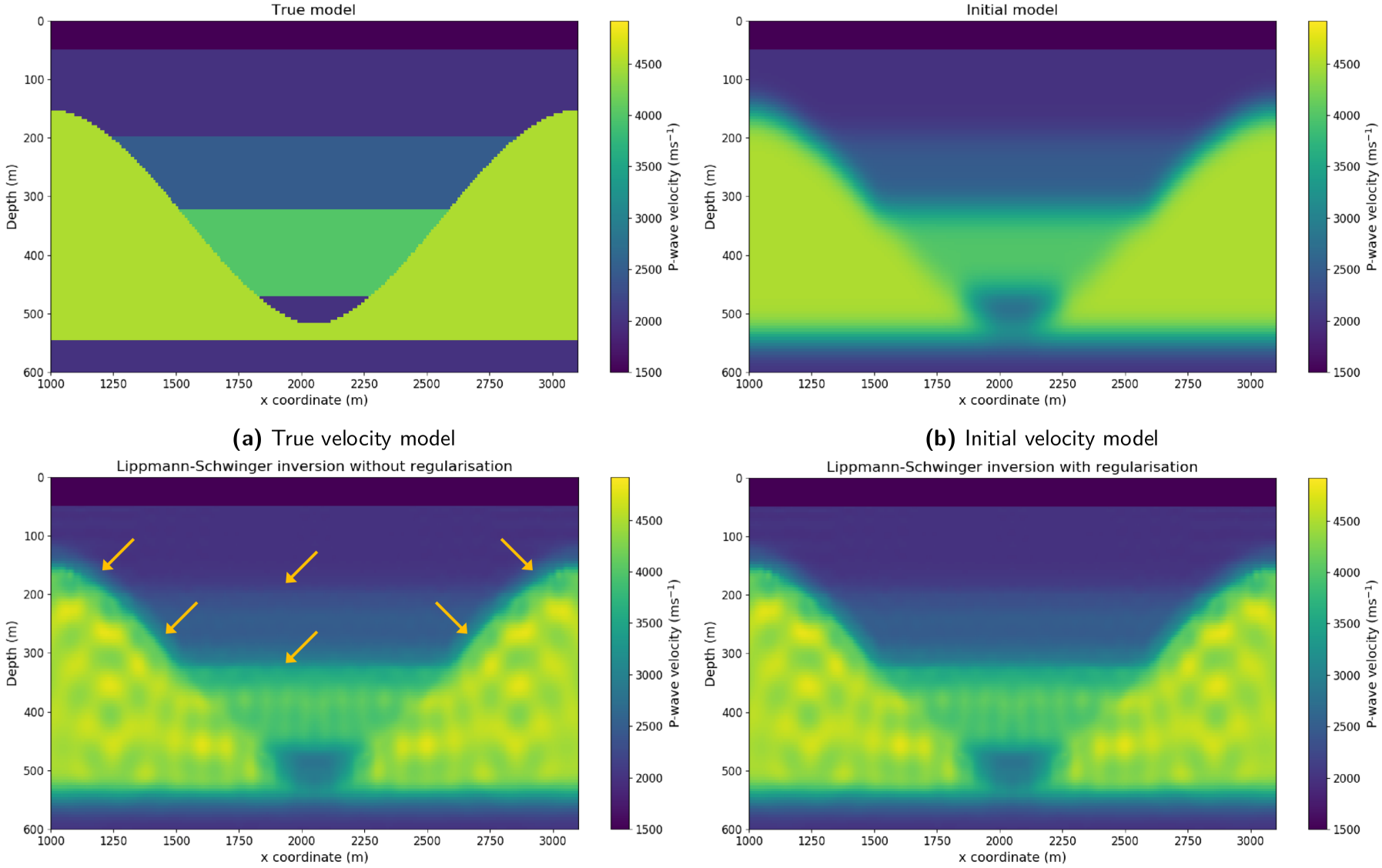
(a) True velocity model

(b) Initial velocity model

(c) After Born inversion

(d) After Born inversion with Tikhonov damping

Figure 6.10: Comparison of Born inversion velocity models to the true and initial models.



(a) True velocity model

(b) Initial velocity model

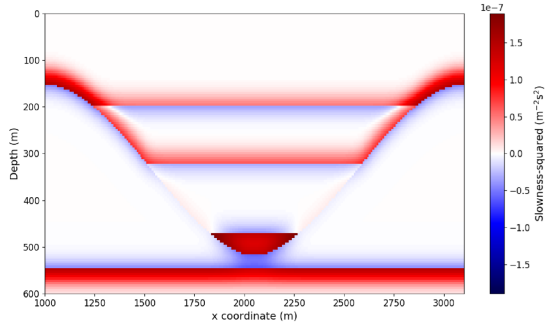
(c) After Lippmann-Schwinger inversion

(d) After Lippmann-Schwinger inversion with Tikhonov damping

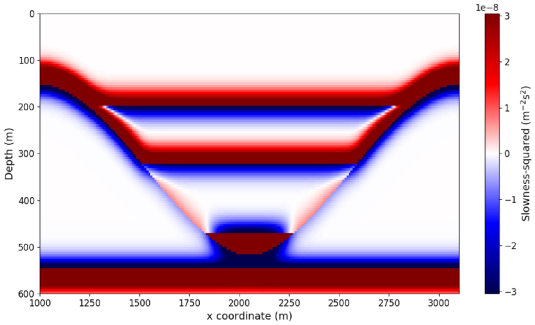
Figure 6.11: Comparison of Lippmann-Schwinger inversion velocity models to the true and initial models.

Examining the model updates i.e. the model perturbation estimates from each inversion method multiplied by the scalar that results in the best fit to the true model, we can see from figure 6.12 that all the inversion results are characterised by significant regular noise, as opposed to the smooth character of the actual model perturbation localised around the interfaces in the true model. The Born inversion without regularisation results in a washed-out model update, with little sense of amplitude differences in different velocity layers - the only sense that this represents a layered medium is the changing in the character of the noise from linear near the surface to patchy inside the syncline layer. With the application of damping, strong amplitude but low resolution interfaces emerge from this noise; especially below the bottom of the surface mute at the interface between layers 1 and 2 (as indicated by the black arrow in figure 6.12d), at the interfaces between layers 2 and 3, and between layers 3 and 4, and the shoulders of the syncline structure. The Lippmann-Schwinger inversions are also characterised by significant regular noise, however, unlike the unregularised Born inversion, amplitude highs can be clearly seen emerging from the noise at locations corresponding to the interfaces between layers 2 & 3 and layers 3 & 4, as well as along the side of the syncline down to layer 4, as indicated by the purple arrows in figures 6.12e and 6.12f. These amplitude highs are significantly higher frequency than those seen in the damped Born inversion update.

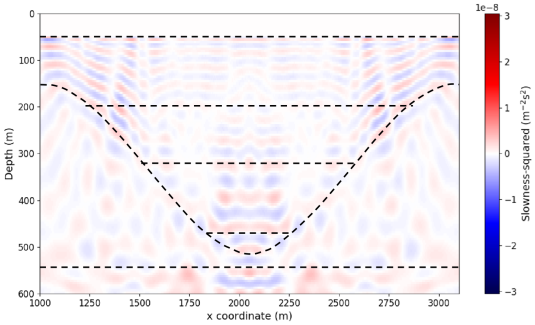
Because it is the slowness-squared model and not the velocity model itself that must be updated, the magnitude of the inversion noise can be better characterised by looking at the velocity model residuals instead of the inversion updates shown in figure 6.12, these velocity model residuals are shown in figure 6.13. We see that the noise added to the unregularised Born inversion is relatively low amplitude, on the order of $\pm 50 \text{ms}^{-1}$, whereas the damped Born inversion exhibits noise levels in excess of $\pm 350 \text{ms}^{-1}$. By comparison, the noise level in the Lippmann-Schwinger models reaches a maximum level on the order of $\pm 280 \text{ms}^{-1}$, significantly higher than the unregularised Born inversion but still less than the damped Born inversion. We also see that the damped Born inversion has changed the velocity at a depth of 50m where the other inversions have not, as marked by the black arrow on figure 6.13. This corresponds to the strong model update indicated by the black arrow in figure



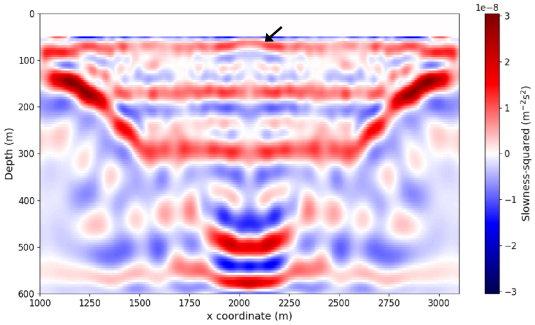
(a) True model perturbation from the initial model



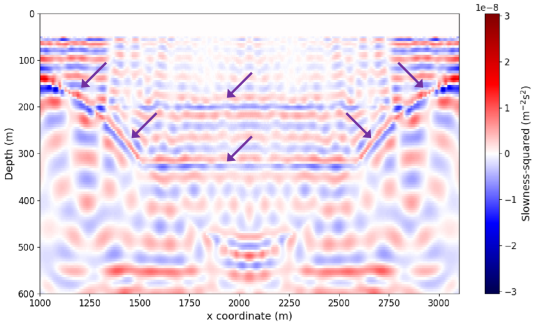
(b) True model perturbation from the initial model on a saturated scale



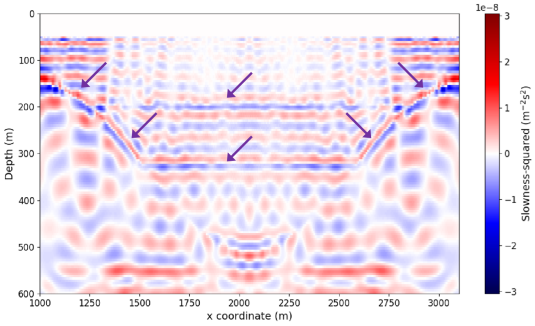
(c) After Born inversion



(d) After Born inversion with Tikhonov damping



(e) After Lippmann-Schwinger inversion



(f) After Lippmann-Schwinger inversion with Tikhonov damping

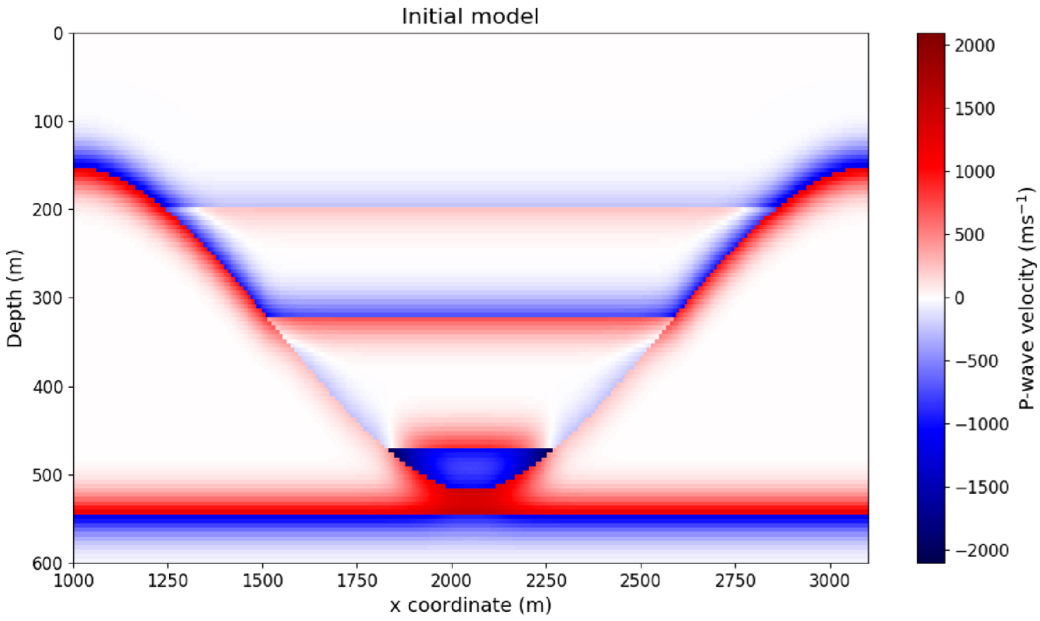
Figure 6.12: Comparison of slowness-squared model updates from Born and Lippmann-Schwinger inversions with the true model perturbation. Panels (b) - (f) are shown on a common scale, panel (a) is shown at true amplitude. The dashed lines in panel (c) indicate the location of interfaces in the true velocity model, whilst purple arrows in panels (e) & (f) match the orange arrows in figure 6.11.

6.12d.

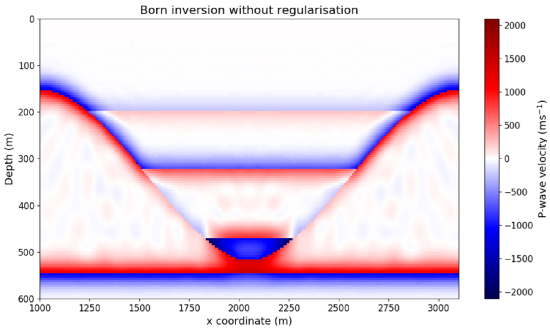
The data residuals obtained by finite-difference modelling a single source at the centre of the source-receiver array are shown figure 6.14. Because the direct wave through the unchanged first layer has significantly higher amplitude than the model reflections, I have presented the data as a residual against the true model data to maximally emphasise any changes to the reflections. Due to the large difference in amplitudes, if the direct wave is not completely removed the reflections will not be visible. One method to remove the direct wave is to apply a mute. However, an overly aggressive mute may remove some of the data as well as the direct wave. Thus to preserve the data I chose to subtract the direct wave instead. Since the model parameters are updated, the direct wave amplitude and kinematics may be changed from the initial model. I chose to present the data as a data residual to the true data rather than subtracting the direct wave modelled in the initial model because it better emphasised the reflection data.

As one might expect, there are very few differences between any of the inversion models and the initial smooth model except for subtle amplitude changes, since there was little change in the sharpness of the interfaces to generate the model contrasts needed to increase the reflectivity amplitude. The most noticeable change in the data (residual) is that from the Born inversion model with Tikhonov damping, wherein an entirely new and consistent reflection (marked by a black arrow) is observed across the receiver array. This is due to the change in velocity in the shallowest parts of the second velocity layer below the surface mute, also marked by black arrows in figures 6.12d and 6.13c.

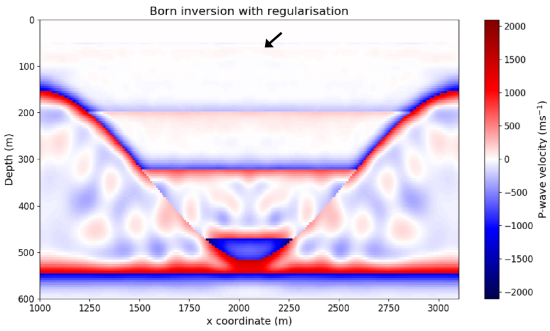
Evaluating the models using the same metrics as in the 1D examples, the L2-norms of the model residual and the data residual, we see in figure 6.15 that the Born inversion model with Tikhonov damping outperforms all other models when measured by both metrics. This is in spite of the fact that this model also showed the highest levels of inversion-related noise and introduced a new reflection to the data residual. The remaining inversion methods all performed comparatively by these metrics - the model misfit decreases by approximately 20%, whilst the data misfit changes negligibly; although it is worth noting that the changes for the Lippmann-Schwinger



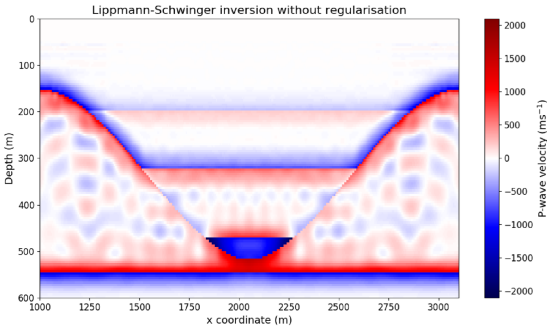
(a) Initial velocity model residual from the true model



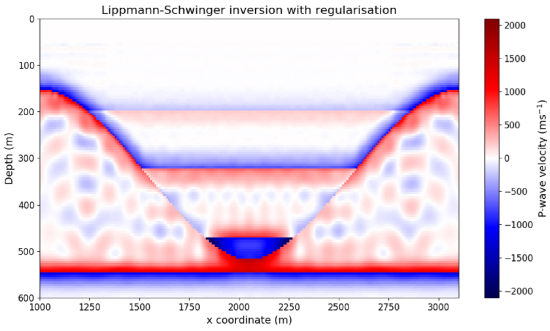
(b) After Born inversion



(c) After Born inversion with Tikhonov damping

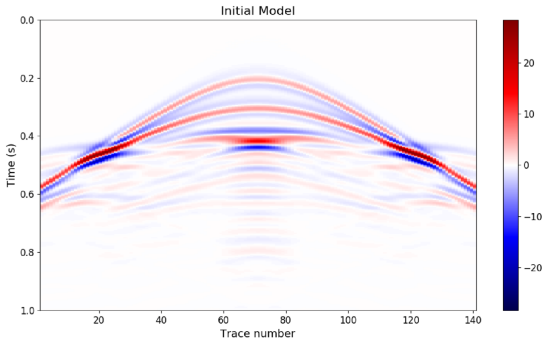


(d) After Lippmann-Schwinger inversion

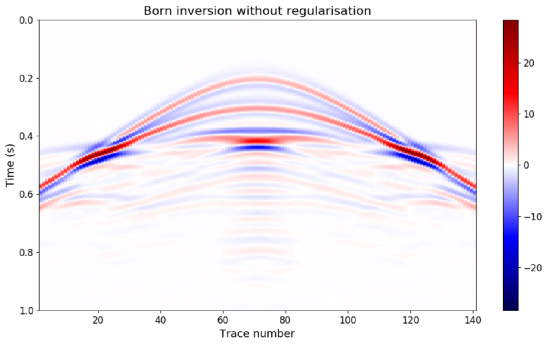


(e) After Lippmann-Schwinger inversion with Tikhonov damping

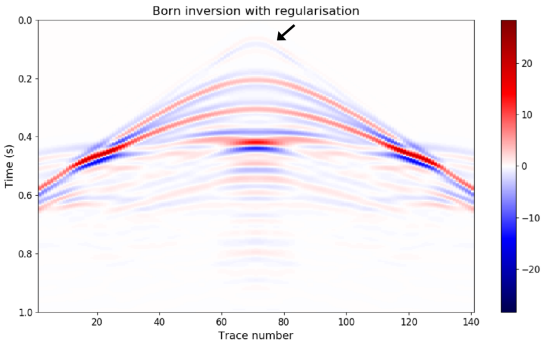
Figure 6.13: Comparison of velocity model residuals after Born and Lippmann-Schwinger inversions with the initial velocity model residual.



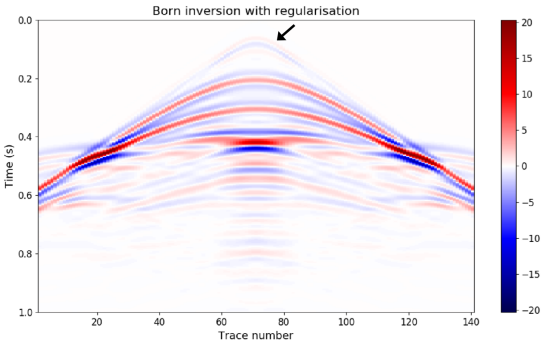
(a) Initial model



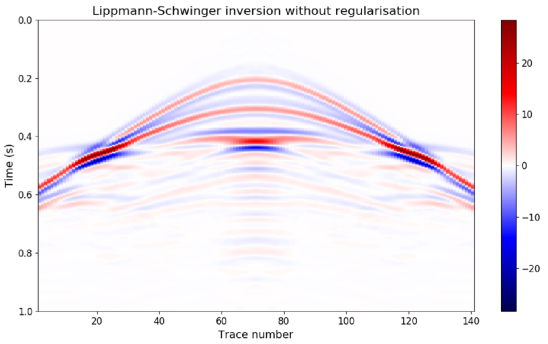
(b) After Born inversion



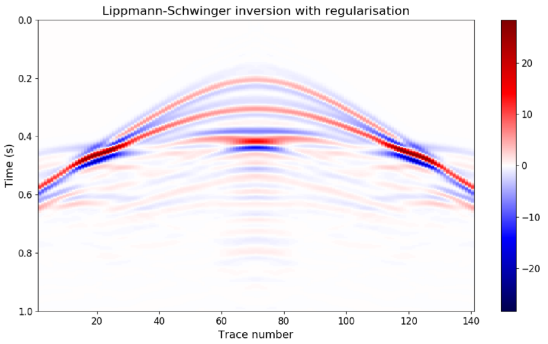
(c) After Born inversion with Tikhonov damping



(d) After Born inversion with Tikhonov damping (on true scale)



(e) After Lippmann-Schwinger inversion



(f) After Lippmann-Schwinger inversion with Tikhonov damping

Figure 6.14: Comparison of data residual gathers before and after inversion by the Born and Lippmann-Schwinger methods. Panel (d) is shown at true scale, all other panels are shown on a common scale.

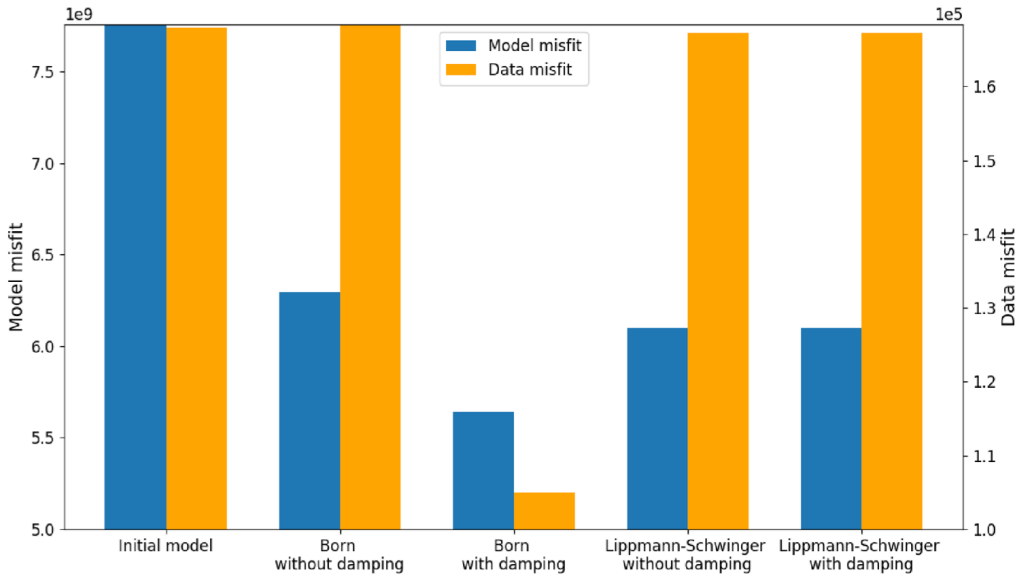


Figure 6.15: Model and data misfits for each inversion method.

models are marginally better than those for the unregularised Born inversion.

6.3 Discussion

Examining the model updates in figure 6.12, beginning with the unregularised inversions in panels (c) and (e) we see immediately that both are dominated by regular noise throughout the model, however while the Lippmann-Schwinger inversion is still able to emphasise the interfaces between layers 2 & 3, layers 3 & 4, and the upper syncline interface; the Born inversion shows little variation in amplitude above the noise - the black dashed lines in figure 6.12c are required to define where the interfaces in the true model are. This suggests that unregularised Lippmann-Schwinger inversion may be more resilient against noise than Born inversion.

When comparing the update from the most successful inversion method as determined by the model and data misfits, Born inversion with Tikhonov damping, to the true model perturbation, care should be taken to remember that while figure 6.12d is presented at true scale, when shown on the same scale figure 6.12b has a saturated colour palette and the perceived similarity in resolution is untrue. The actual model perturbation shown at true scale, figure 6.12a, is an order of magnitude greater in

amplitude and has much higher resolution structure at the interfaces between velocity layers. This is a reminder that even the most successful inversion in this experiment does not result in an accurate representation of the true model.

Comparing the inversion models which had Tikhonov damping applied (figures 6.12d & 6.12f), the optimally damped Lippmann-Schwinger inversion shows significantly higher resolution than the optimally damped Born inversion. Given that the unregularised Born and Lippmann-Schwinger inversions (figures 6.12c & 6.12e) exhibit similar resolution, the lower resolution of the damped Born inversion compared to the damped Lippmann-Schwinger inversion must be a consequence of applying an optimised amount of damping to each inversion individually. It is then coincidence that the optimally damped Born inversion model is smoother than the optimally damped Lippmann-Schwinger inversion.

Looking at all the inversion methods now, one sees that the 4th horizontal layer and the syncline layer are the most affected by noise from the inversion, and this is reflected in the resultant velocity models in figures 6.10 and 6.11. Compared to the true velocity model, we see that these layers are also the layers with the two highest velocities. It makes sense then that this is the case, since in the terms of the model parameter of the inversion, the (P-wave) slowness-squared of the medium, these layers have the smallest amplitudes, and are thus more susceptible to noise than those with slower velocities. From figure 6.13 we see that the noise pattern exhibits a squarish, almost checkerboard pattern within the syncline layer, whereas in the fourth layer the polarity of the noise amplitude appears to be aligned vertically. This is likely the result of a trade-off between model parameters in the inversion to optimise the result, with the alignment of the noise polarity in the fourth layer being a consequence of there being more reflections and reflectors in that region of the model providing more tie points to constrain the inversion.

From the same figures, it is also clear that no inversion method is able to well define the sides of the syncline structure inside the 4th horizontal layer. Figure 6.16 shows the results of ray tracing through the true velocity model from the location of source 21, 300m from the top left corner of the model area. Rays are modelled between 90°along the horizontal and 180°vertically downwards from the source. It

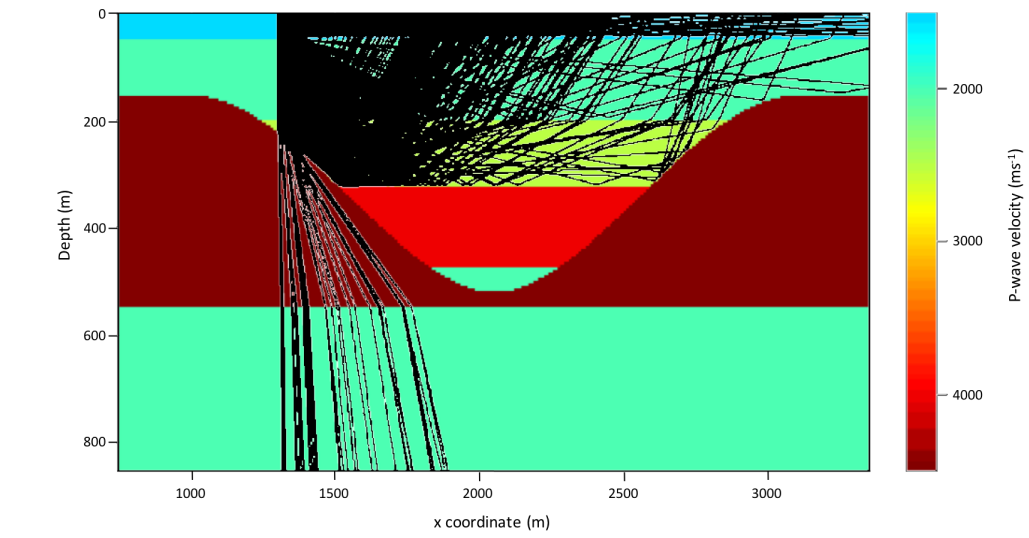
can be seen that with an initial 500 rays shot, no ray path enters the 4th horizontal velocity layer. When increased to 1000 rays, only one ray samples the layer. At 8000 rays, we see that while the upper layers of the model are extremely well sampled, only a handful of raypaths sample the 4th layer, and all enter from above the layer - no raypaths sample the layer having first reflected from the bottom of the syncline and entering through the syncline interface. In order to image this structure, it must be illuminated by the source along a raypath that leads back to a receiver. This demonstrates that it is extremely difficult for any method to image this structure.

The problem is, in fact, more severe than first stated. The source demonstrated, number 21, is the left-most source in the acquisition array that is able to sample the 4th velocity layer. For sources 1 to 20, no raypaths from these source positions are able to sample this layer due to critical refraction. Using the symmetry of the model and the acquisition array, this means that sources 121 to 141 are also insensitive to this layer - a total of 40 out of 141 sources or 28% of the available sources then are incapable of imaging the 4th velocity layer. This means the inversion is reliant on the zero- and near-zero offset sources to image this layer. Through a combination of sensitivity to noise and a lack of illumination, it makes sense that this layer is not well-handled by any of the inversion methods.

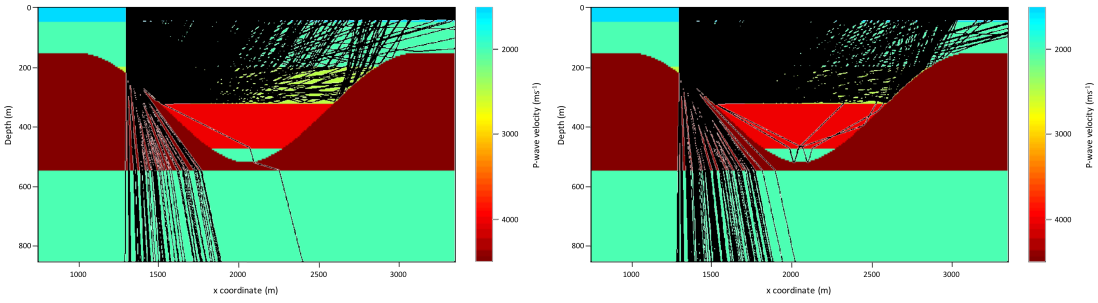
It may also be possible to conclude that since the Lippmann-Schwinger inversion was no more successful than Born inversion at imaging this layer, that allowing for multiply-scattered waves that may sample this layer does not (sufficiently) compensate for a lack of illumination from a significant part of the source array.

Based on the model and data misfits in figure 6.15, we can see that the trend seen in one dimension of the data misfit not improving with the model misfit when the inversion is optimised by the model misfit may also be true in two dimensions, at least for the first iteration. It is of course possible that the data misfit would improve if the inversion were iterated. The exception to this observation is the damped Born inversion which, when considering the different y-axis scales, shows improvement in the data misfit on par with the improvement in the model misfit, an approximately 40% reduction.

Looking at the data residual gathers, this seems counter-intuitive - whilst all the

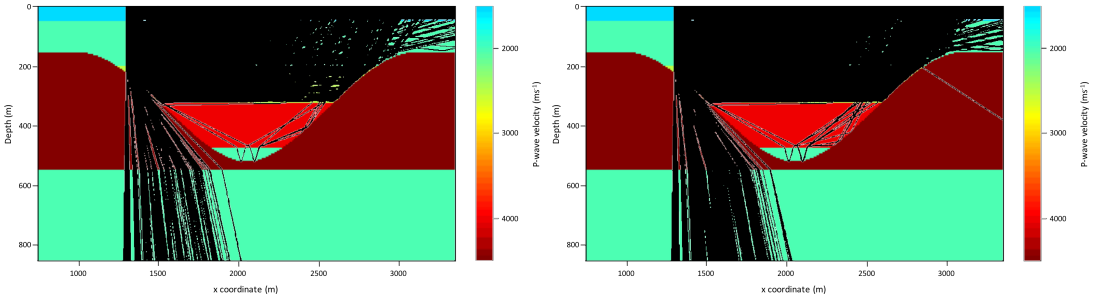


(a) Using 500 rays



(b) 1000 rays

(c) 2000 rays



(d) 4000 rays

(e) 8000 rays

Figure 6.16: Raypaths shot through the true velocity model from source position 21 (x=1300m), demonstrating the lack of illumination in deeper layers from wide offset sources.

other gathers are almost identical to the initial model's data residual, the damped Born inversion actually introduces a new reflection to the residual. By comparing the damped Born inversion data residual plotted on the common scale across figure 6.14 (panel c) to the same residual plotted at true scale (panel d), one observes that the maximum amplitude is approximately 20% less than the remaining gathers. When combined with the very low amplitude of the introduced reflection, it is no longer implausible. The fact that the data misfit can be so much improved compared to a model that introduced sharper interfaces but less accurate average velocities in each layer shows that the sum of the reflected energy over a smoother interface with more accurate velocities can more accurately match the true data than a higher resolution interface between less accurate velocities. This is a demonstration of the mechanism by which the Born and Lippmann-Schwinger equations describe the scattered wavefield, as a sum over the correlations of the Green's functions from source and receiver at all locations where there is a velocity perturbation regardless of the size of that perturbation, and not simply assuming that only sharp acoustic impedance contrasts can scatter waves.

The similarity of the data gathers from the initial model and the inversion models poses the question of whether or not the 2D inverse problem would rapidly converge when optimised with the data residual[1]. If many models can produce very similar data, then it would likely take many iterations of inversion for the problem to converge. On the other hand, the evidence from the 1D inversions says this may not be the case. In those instances, when the inversion was optimised by the model misfit, the data misfit generally did not improve with any amount of iteration, and when optimised with the data misfit, the inversion generally fell into a local minimum of the data misfit within a few iterations with a significantly lower misfit value, and then iterated around this local minimum i.e. the inversion rapidly converged on a single model.

In conclusion, despite Lippmann-Schwinger inversion showing evidence of producing higher resolution interfaces compared to Born inversion, the high levels of noise combined with the low resolution more generally and the minimal improvement in the model means that I cannot say that Lippmann-Schwinger inversion provides a

[1]Due to time limitations, testing was limited to optimisation with the velocity model residual.

better result than the standard Born inversion. When compared to other established methods of velocity model inversion such as least-squares migration, the quality of the results shown here is poorer.

Based on finding evidence of higher resolution at model interfaces with Lippmann-Schwinger inversion, it may yet be possible to show if Lippmann-Schwinger inversion leads to more accurate velocity models if more frequencies could be used in the inversion, a conclusion reached from the analytical 1D inversions in chapter 4. Not only do higher frequencies allow for greater vertical and horizontal resolution, but using a greater number of frequencies may also reduce the inversion noise by reducing the impact of the Gibbs phenomenon. As shown in figure 6.17 (reproduced from Gubbins (2004)), the Fourier representation of a square wave becomes more accurate with lower amplitude, higher frequency amplitude oscillations as the length of the Fourier series increases i.e. more frequencies are used. In this context, the discretised velocity model can be thought of as a 2D square filter which is multiplied by the kernel matrix to produce the data; ergo, in the inverse problem, when more frequencies are used in the data and the kernel matrix, the resulting model will be more accurate with lower amplitude, higher frequency noise oscillations.

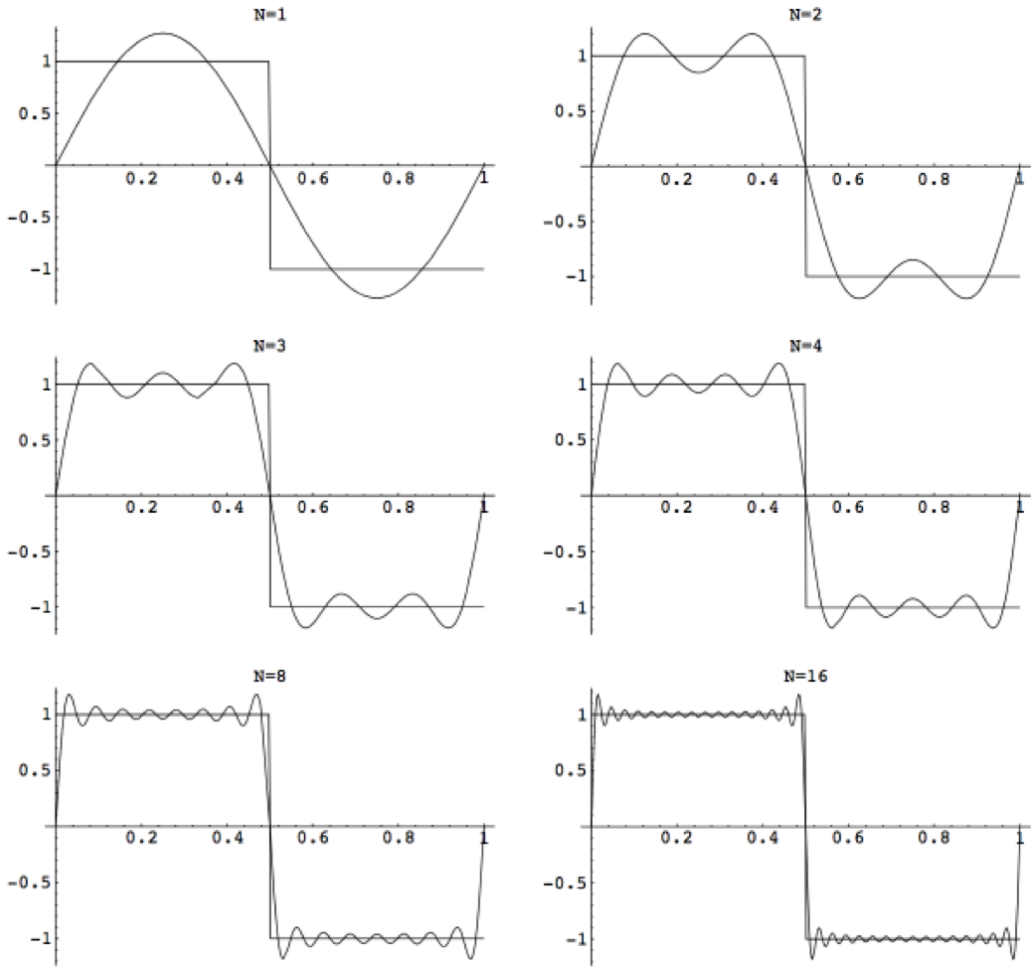


Fig. A1.1. Fourier series approximation to a square wave. The Fourier series has been truncated at 1,2,4,8, and 16, terms. Note the overshoot and compression of the oscillations around the discontinuity.

Figure 6.17: A demonstration of the Gibbs phenomenon in the Fourier representation of a square wave (reproduced from Gubbins (2004)).

Chapter 7

Reconsideration of the Lippmann-Schwinger inverse problem in two dimensions

In the previous chapter, I demonstrated the linear inversion of the Born and Lippmann-Schwinger modelling equations by explicit calculation of the inverse kernel matrix for a two dimensional velocity model and acquisition geometry. Data were acquired at 141 sources and receivers co-located at the surface of the model, and all sources and receivers were used in the inversion as described in chapter 2. Due to computational limitations, the number of frequencies used in the inversion was reduced from 376 to 5, which amounts to a 98.7% reduction or approximately 75-fold decrease. The resulting velocity models were disappointing, with low spatial resolution and contaminated with high amplitude periodic noise that limited the amount the model could be updated (even in a low resolution direction). As a consequence, I was unable to conclude with any confidence whether there was any improvement from using the Lippmann-Schwinger inversion method over the Born inversion method, or whether any observed benefits were the result of chance. If there is any possibility of making such a determination with the limited computational resources available, it is clear that another approach is necessary. Moreover, such an approach needs to be applicable in a computationally efficient manner.

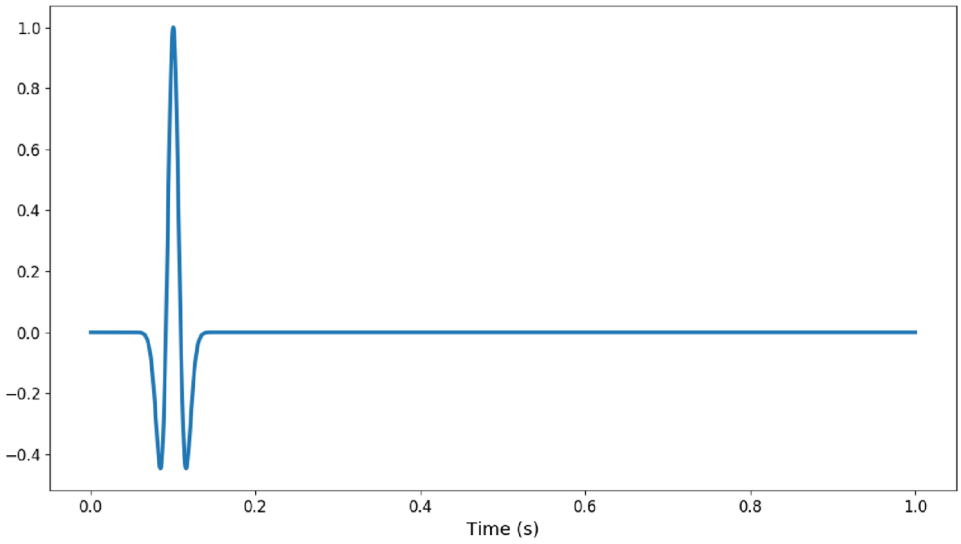
Looking again at the matrix form of the Born forward problem, equation 2.35 (with the equivalent form of the Lippmann-Schwinger equation replacing one of the G_0 terms with G), for a single frequency, the kernel of the Born equation is an mxpq matrix where m is the number of model parameters, p the number of sources, and q the number of receivers. The scattered Green's function between each source-receiver pair is then formed by multiplying the kernel element for that source-receiver pair at a given location with the model perturbation at that location, and summing over all locations. The inverse problem then consists of multiplying the pqxm inverse kernel with the 1xpq scattered wavefield vector for each source-receiver pair. The model perturbation at each location is thus formed by multiplying the inverse kernel element for the given model location and source-receiver pair with the scattered wavefield between that source and receiver, and then summing over all sources and receivers. Since all the operations described are linear, the sum of the inverted-for model perturbations as determined from each source individually should be equal to the model perturbation determined using all sources simultaneously. The equivalence of the solutions is only approximate because building the inverse kernel when all sources and receivers are used involves mixing of information from different sources and receivers. This information is lost when the inverse kernels are built for each source individually. The mixing of information from different sources and receivers in the inverse kernel is not what would normally be described as "cross-talk" since there is no ambiguity in the forward problem, the forward kernel creates the data for each source-receiver pair without influence from any other source. Separating the forward kernel for each source individually can be expected to increase the size of the model null space, since the source and receiver mixing in the inverse kernel increases the amount of information that can constrain the inversion.

The solution I present in this chapter is to reduce the number of sources from 141 to 1, inverting for the slowness-squared perturbation for each source separately and then stacking over sources to produce a single update direction that can then be optimised by a line-search for instance. This approach can thus be considered akin to the typical approach used in migration-type imaging methods such as Kirchoff migration or reverse-time migration. The advantage of this approach is that because

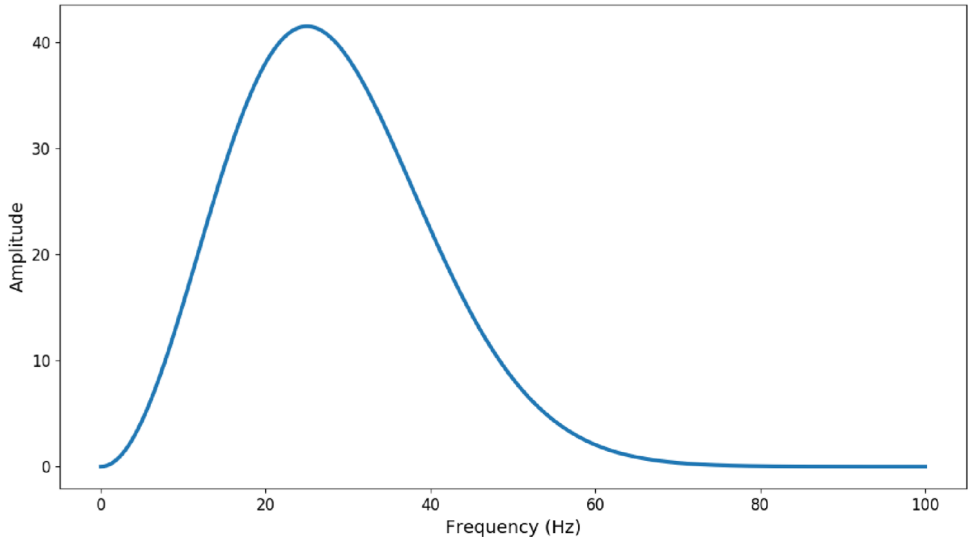
the size of the kernel matrix is linearly dependent only on the numbers of sources, receivers, and frequency bins, the reduction in the kernel size by using a stacked single source approach more than compensates for the reduction in the number of frequency bins previously required. So I am now able to restore all the modelled frequencies to the kernel matrix and simultaneously make savings on the amount of memory required for the inversion.

The downside of this approach is that it increases the amount of time required overall to obtain a usable result. Once all of the finite-difference modelling of the Green's functions has been completed, the most time intensive process is calculating the individual elements of the kernel matrix. The simplest, and most logical approach, would be to build the elements of the kernel as needed and immediately perform the inversion; however, since the process is sufficiently computationally intensive, it is more efficient to compute the kernel in advance for the sources and frequencies required and then read in the elements immediately prior to inversion: this allows for the process to be parallelized, which in practice ranges between 2x parallelization and up to 30x parallelization, resulting in the time needed for completion ranging between less than one hour and in excess of 12 hours for a single source. The kernel is constructed model point by model point. Parallelization is implemented by grouping the model volume into columns and submitting each column separately to a supercomputer. The degree of parallelization achievable is limited by resources available at the time of submission, with the maximum possible parallelization equal to the number of columns of model points. In the model I used, this would be a maximum of 521x parallelization. The increase in time to construct the kernel due to processing multiple sources is significantly less than the time required to construct the kernels separately for each source.

Based on the low resolution of the previous inversion models, I decided to change the wavelet from the 12Hz Gaussian derivative wavelet to a 25Hz Ricker wavelet (shown figure 7.1) in an attempt to increase the maximum vertical resolution possible. Given that the widest-offset sources do not illuminate the deepest velocity layers inside the syncline, as seen in figure 6.16, it is especially important to maximise the information that can be retrieved from the sources that do illuminate these layers.



(a) In time domain, truncated at 1 second.



(b) Amplitude spectrum truncated at 100Hz.

Figure 7.1: 25Hz Ricker wavelet in both time and frequency domains.

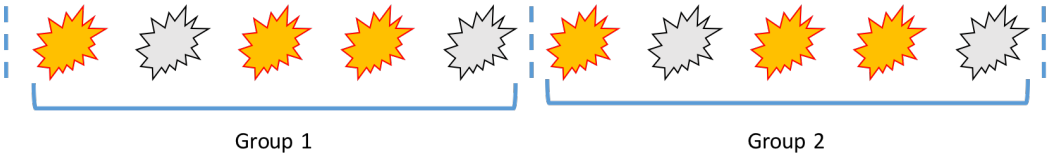


Figure 7.2: A pictorial demonstration of the source subsampling scheme used. Coloured explosions represent sources used for inversion and grey explosions sources omitted from the inversion.

In order to save time, I decided to subsample the source array by 60%. Sources were sampled at regular intervals along the array by breaking the source array into groups of 5 sources, from which the 1st, 3rd, and 4th sources of each group were selected. This subsampling method is demonstrated graphically in figure 7.2. Since there are 141 sources in total, this results in 28 groups of 5 sources with one remaining source. Using this source results in a source array that is symmetric about the centre of the acquisition array.

To make any comparison to the results of chapter 6 a fair one, I decided to still use the wide-offset sources that do not illuminate the deeper layers, despite having the prior knowledge that these source positions do not contribute any information to the inversion in these parts of the model.

7.1 Born inversion versus Lippmann-Schwinger inversion in two dimensions

I now present the results of the inversions of the Born and Lippmann-Schwinger equations using the stacked single source inversion approach described above.

Figure 7.3 shows the velocity models that result from regularised and unregularised Born inversion using the shot-stacking approach. After a single iteration of Born inversion without regularisation (figure 7.3c), the resulting model is very similar to the initial model with the addition of low amplitude noise in the highest velocity layers, as well as spurious velocity lows in layers 3 and 5. These appear to coincide with the increases in velocity in the smooth initial model in regions above sharp velocity increases in the true model. When regularisation is applied (figure 7.3d), clearly

defined interfaces appear in the updated velocity model surrounded above and below by smooth velocity errors. The smooth transition from the bottom of the syncline layer to the bottom layer has also been uplifted from its true depth (figure 7.3a) and extended to form an effectual erroneous velocity layer in the updated model. The low amplitude periodic noise pattern in the unregularised inversion has been eliminated and replaced by linear artefacts indicative of the incomplete destructive interference of the single source images. This is likely due in part to the subsampling of the source array, since the sources are not evenly distributed, the interference pattern where there are two active sources next to each other is different to the interference pattern of the isolated active sources and energy is left in the image after the stack as an artefact.

The velocity models resulting from regularised and unregularised Lippmann-Schwinger inversion using the shot-stacking approach are shown in figure 7.4. The models are extremely similar to those from the Born inversions - unregularised inversion results in a largely smooth model with low amplitude periodic noise in the highest velocity layers, while regularised inversion results in defined interfaces surrounded by smooth velocity errors, the uplift of the bottom interface, and the replacement of periodic noise with migration stacking artefacts.

Examining the optimised model update directions (figure 7.5), we observe that without regularisation, both Born and Lippmann-Schwinger inversions are highly contaminated by periodic noise, with the spatial frequency of said noise increasing as the velocity decreases. By and large, neither model update direction is representative of the true model perturbation (figures 7.5a & 7.5b); in fact, the only interface that is correctly identified, the bottom interface, has the incorrect sign thus worsening the model fit in the only part of the model that is clearly identifiable. In the opposite fashion, the optimally damped Born and Lippmann-Schwinger model updates agree significantly across the entire model. The amplitudes and resolutions of interfaces are extremely similar with both inversion methods, and except for some very low amplitude noise in the high velocity syncline layer, with both methods damping has essentially eliminated the periodic noise in the unregularised inversions. Compared

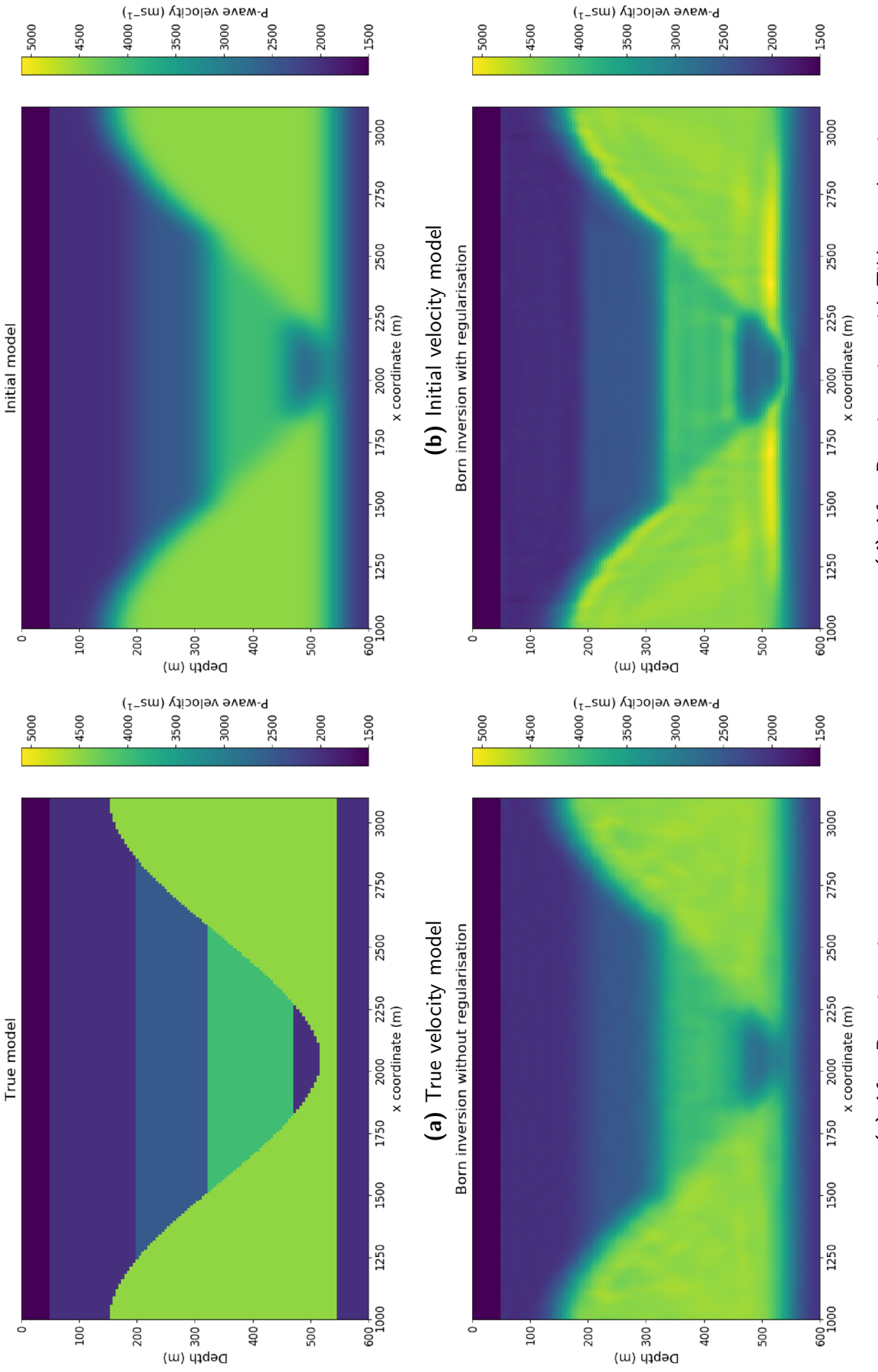
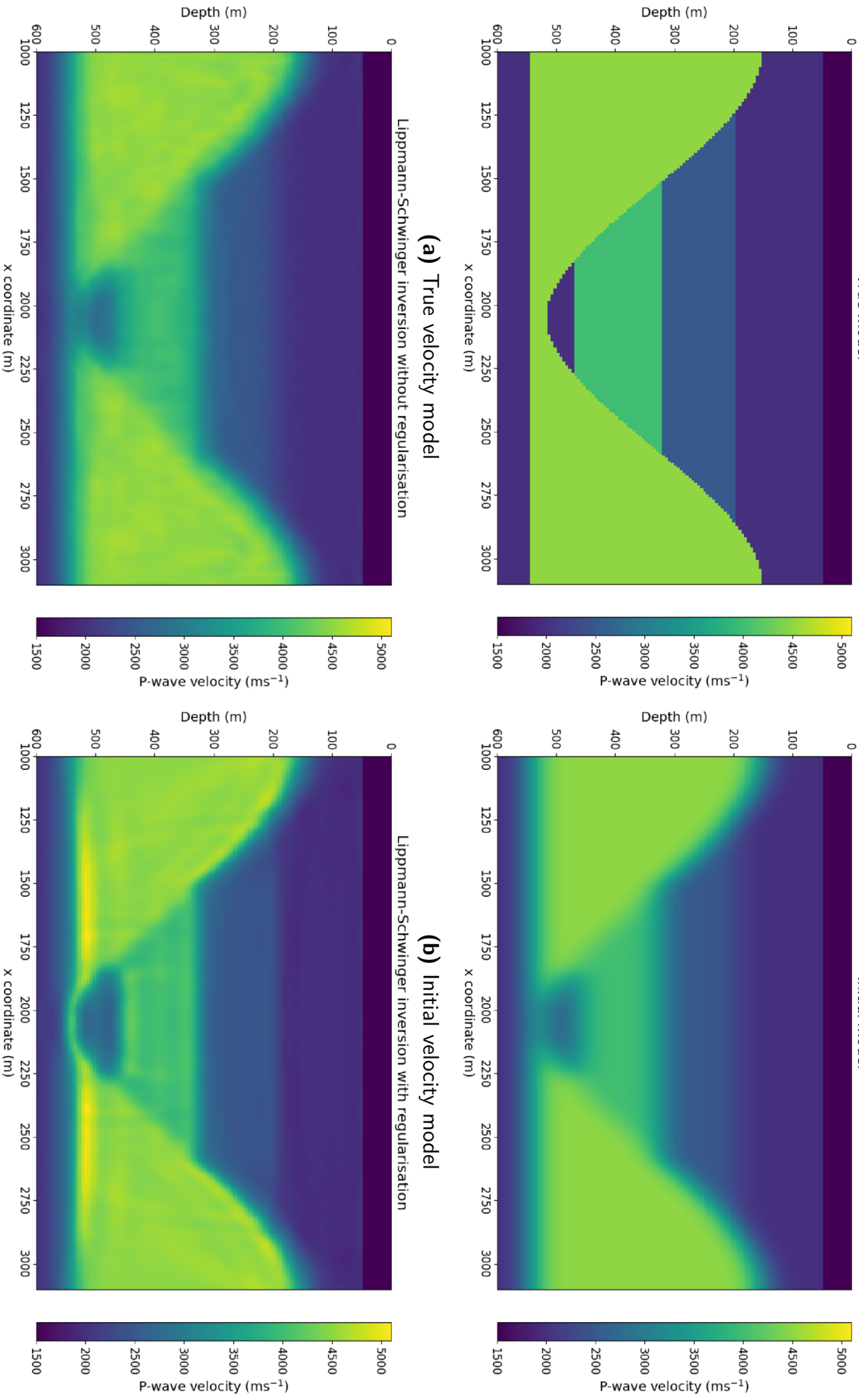


Figure 7.3: Comparison of Born inversion velocity models to the true and initial models.



(a) True velocity model

(b) Initial velocity model

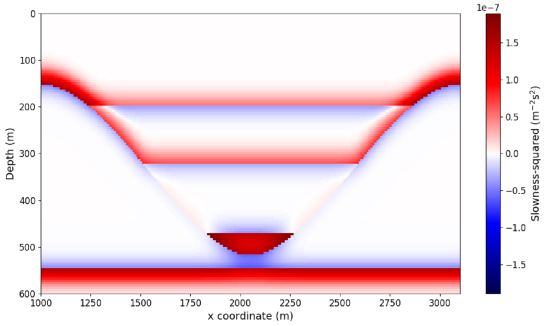
(c) After Lippmann-Schwinger inversion

(d) After Lippmann-Schwinger inversion with Tikhonov damping

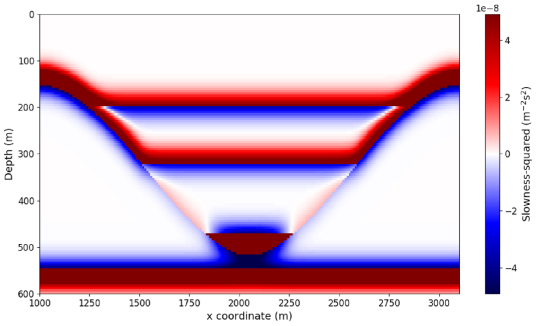
Figure 7.4: Comparison of Lippmann-Schwinger inversion velocity models to the true and initial models.

to the unregularised inversions, both Tikhonov-damped inversions (figures 7.5d & 7.5f) display regularly-spaced vertical artefacts in the second velocity layer. The only immediately noticeable difference between the damped inversions is a slight reduction in the amplitudes of the stacking artefacts in the second velocity layer with the Lippmann-Schwinger inversion.

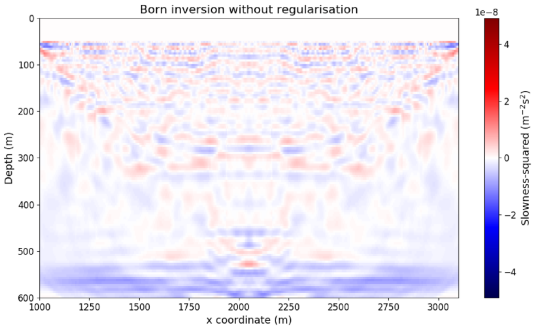
Taking the residual of the true data and the data modelled before and after inversion (to remove the obscuring effect of the much higher amplitude direct wave), figure 7.6 shows that the data modelled after unregularised Born and Lippmann-Schwinger inversions is almost identical to the data from the initial model in terms of both amplitudes and observed reflections. As with the models and model updates, the data after regularised Born and Lippmann-Schwinger inversion are extremely similar to each other in the same ways as the data from the unregularised inversions are to each other. The differences between the regularised and unregularised models data are two fold: firstly, the amplitude of arrivals in the regularised model data gathers is significantly lower than in the unregularised model data gathers, and secondly, an entirely new reflection has been introduced to both the regularised model data gathers before the reflection arrivals present in the data from the initial model.



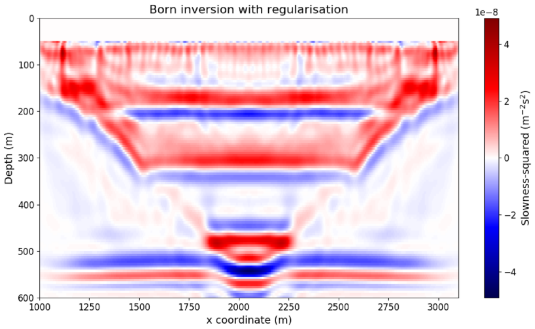
(a) True model perturbation from the initial model



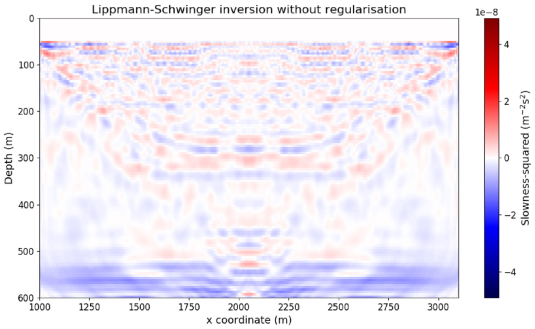
(b) True model perturbation from the initial model on a saturated scale



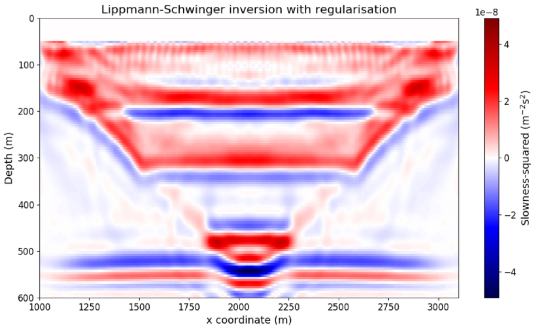
(c) After Born inversion



(d) After Born inversion with Tikhonov damping

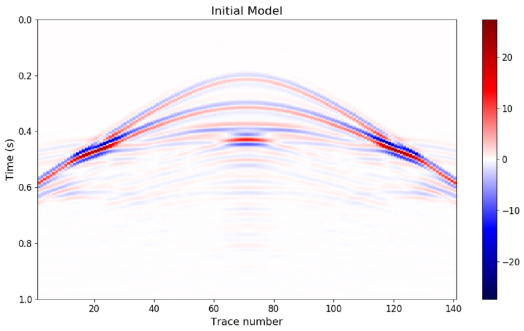


(e) After Lippmann-Schwinger inversion

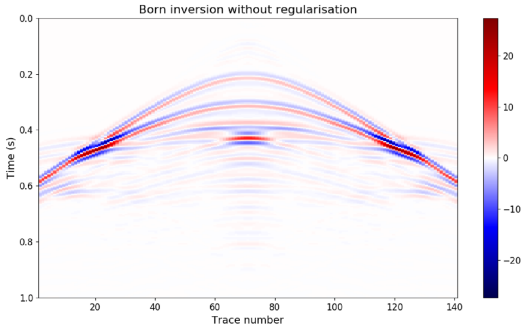


(f) After Lippmann-Schwinger inversion with Tikhonov damping

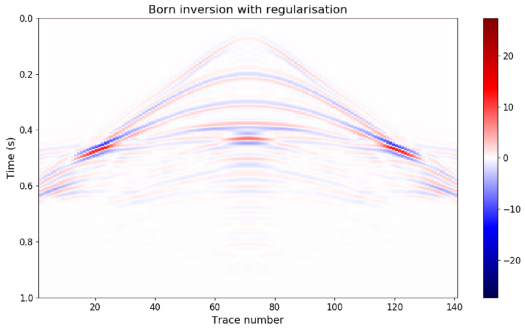
Figure 7.5: Comparison of slowness-squared model updates from Born and Lippmann-Schwinger inversions with the true model perturbation. Panels (b) - (f) are shown on a common scale, panel (a) is shown at true amplitude.



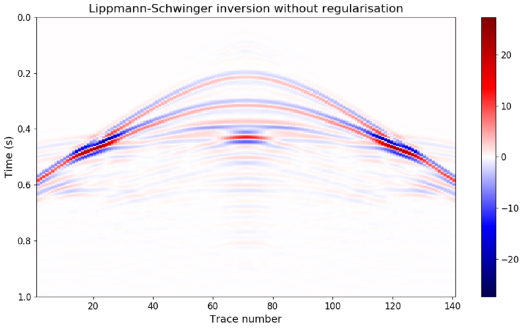
(a) Initial model



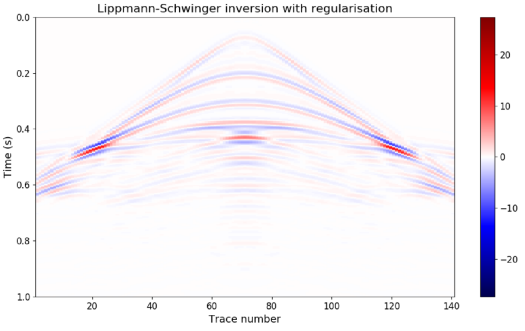
(b) After Born inversion



(c) After Born inversion with Tikhonov damping



(d) After Lippmann-Schwinger inversion



(e) After Lippmann-Schwinger inversion with Tikhonov damping

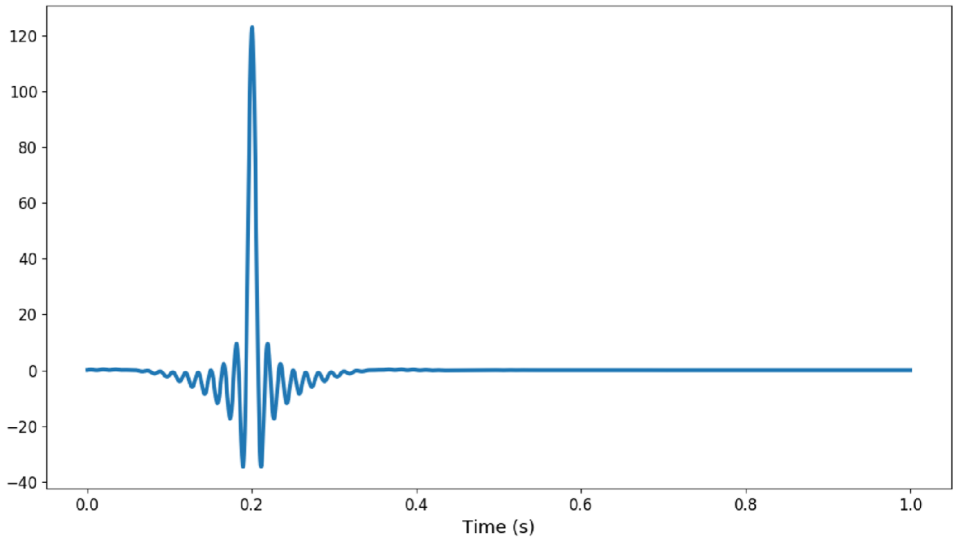
Figure 7.6: Comparison of data residual gathers before and after inversion by Born and Lippmann-Schwinger methods.

7.2 Marchenko-Lippmann-Schwinger inversion in two dimensions

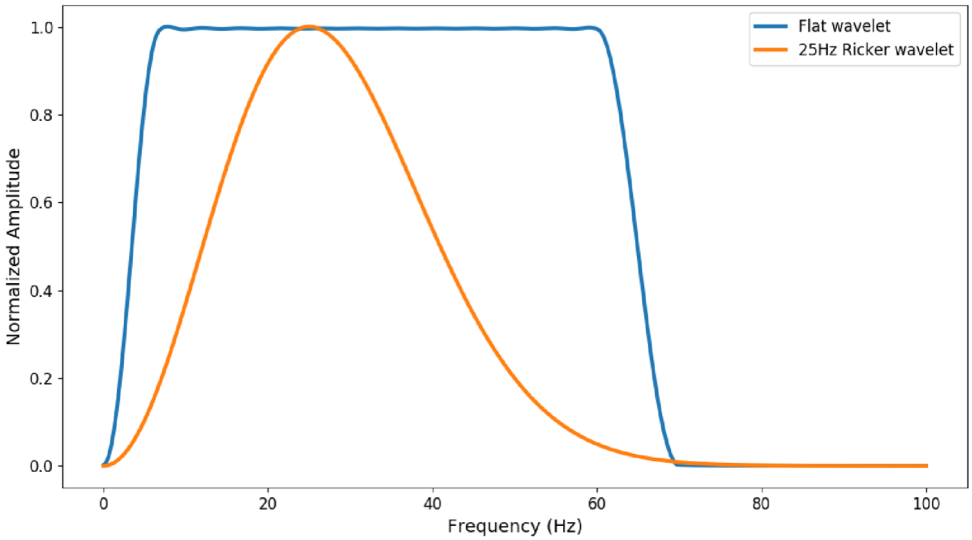
Based on the noticeable noise suppression seen in the velocity models obtained using a stacked single source inversion approach in the previous section as compared to those seen in chapter 6, I decided to attempt inverting the Lippmann-Schwinger equation using Green's functions derived using a Marchenko method to determine if Marchenko methods can perform comparably to the exact full-scattering Green's functions in the context of inversion for a velocity model. Because the 25Hz Ricker wavelet used in the previous section contains higher frequencies than the 12Hz Gaussian derivative wavelet used in chapter 6, the flat spectrum wavelet needed for Marchenko must also be updated so that the amplitude spectrum is (approximately) flat within the bandwidth of the 25Hz Ricker wavelet. Figure 7.7 shows the new flat wavelet employed in the Marchenko process and its amplitude spectrum, alongside the amplitude spectrum of the Ricker wavelet.

To generate the direct wave arrivals required for Marchenko methods from each source and receiver position located at the surface to every location in the subsurface where Marchenko is applied, I used the Green's functions previously modelled in the smooth initial model. This is in contrast to the approach originally intended for use with the 12Hz Gaussian derivative wavelet, had the noise levels been lower in the results shown in chapter 6, whereby the true Green's functions modelled in the true Earth velocity model would be windowed to exclude all arrivals after the direct wave. The reasoning behind the originally intended approach was to attempt to obtain the most accurate Marchenko Green's functions possible, and this is only possible if the window calculated from the traveltime of the direct arrival is exact.

The major flaw with this approach, besides the fact that it is only possible in synthetic experiments (otherwise we would already have the answer we seek!) is that the source wavelet convolved with the direct arrival is used to provide the Marchenko Green's function with the correct source signature. If the shape of the wavelet is altered in any way, the Marchenko-derived Green's functions will not have the correct source signature. This scenario is especially likely close to reflectors, whereby the



(a) In time domain, truncated at 1 second.



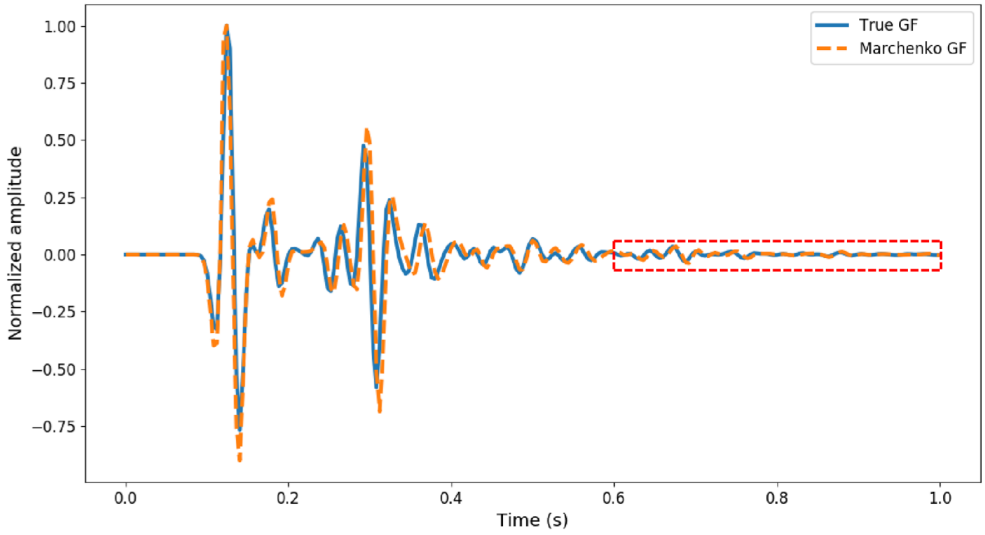
(b) Amplitude spectrum truncated at 100Hz.

Figure 7.7: Flat spectrum wavelet for use in a Marchenko method in both time and frequency domains.

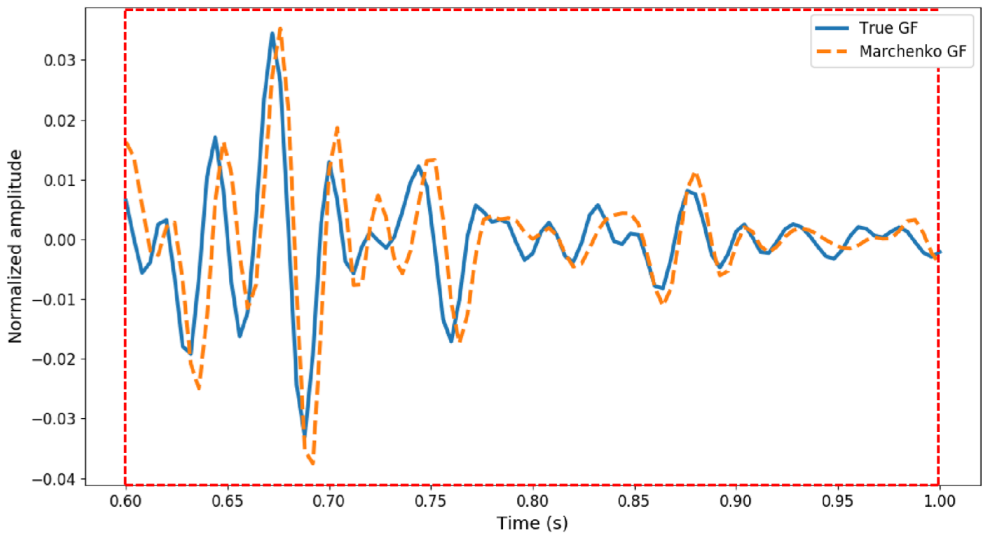
direct wave and the reflection from the interface can overlap in time. I originally intended to combine these Marchenko Green's functions with Marchenko Green's functions using a direct arrival with a flat spectrum source signature in another experiment (specifically, to test the Marchenko-Lippmann-Schwinger FWI method detailed in chapter 2), but it quickly became clear that the long wave train of the flat spectrum wavelet in figure 7.7 would make the original approach impossible. An example of a Green's function obtained using Marchenko methods for my model is shown in figure 7.8 in comparison to the true Earth model Green's function at the same subsurface point. The match between the two time series is reasonable but not an exact match. Although the peak amplitudes of the direct arrivals occur at exactly the same time, the amplitudes of the direct arrivals do not match exactly after amplitude normalisation has been applied. The resulting waveform fit, both in terms of amplitude and phase, is inexact and inconsistent down the entire trace. Focusing on the multiply-scattered coda in the red box in figure 7.8, we see that the fit deteriorates as the wave train becomes more complicated. Even though the fit is not exact, the arrival times of each reflection event are still approximately correct, and while we thus expect some artefacts to be introduced by erroneous phases and amplitudes, it is also possible that by chance these errors lead to constructive or destructive interference that aids the retrieval of a more accurate velocity model.

Following Green's function retrieval for all points using Marchenko methods, I substituted these Green's functions into the kernel of the Lippmann-Schwinger equation and solved for the least-squares solution, which I call Marchenko-Lippmann-Schwinger inversion. Figure 7.9 shows the results after a single iteration of Marchenko-Lippmann-Schwinger inversion with and without regularisation. Without regularisation, the updated velocity model is visibly unchanged from the initial model. With regularisation, although there is little change in the macro-scale velocities within each layer, there is a noticeable improvement in the resolution of the horizontal interfaces, with particular improvement in the interfaces between horizontal layers 2 & 3, and layers 4 & 5. The horizontal bottom interface, however, appears unchanged.

From the slowness-squared updates shown in figure 7.10, it is clear why the updated velocity model after inversion is so similar to the initial model: the optimally

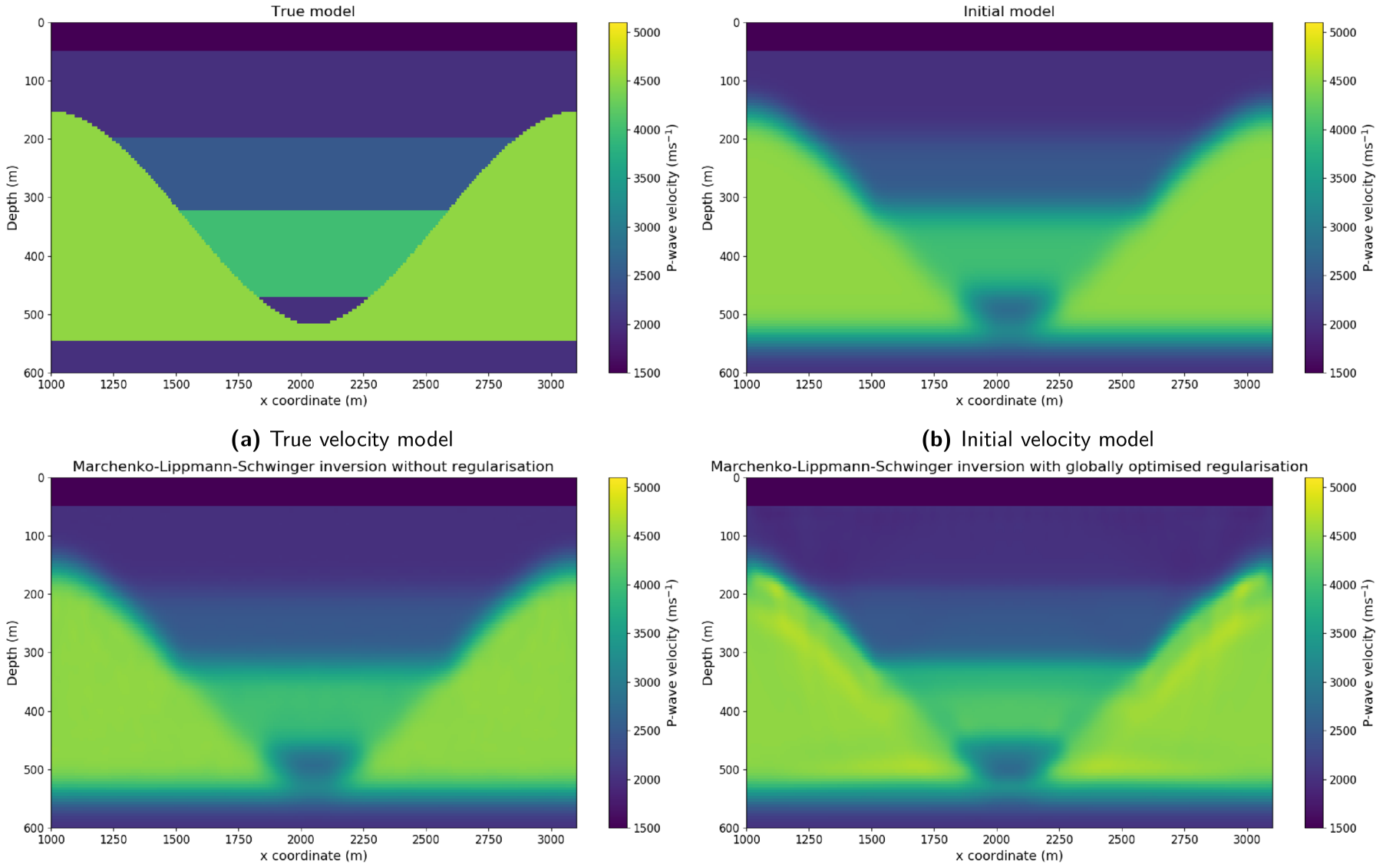


(a) A scattering Green's function, truncated at 1 second.



(b) Close up of the multiply-scattered wave train highlighted by the red box in window (a).

Figure 7.8: Comparison of Marchenko-derived Green's function to the true Earth Green's function between the same two points.



(a) True velocity model

(b) Initial velocity model

(c) After Marchenko-Lippmann-Schwinger inversion

(d) After Marchenko-Lippmann-Schwinger inversion with Tikhonov damping

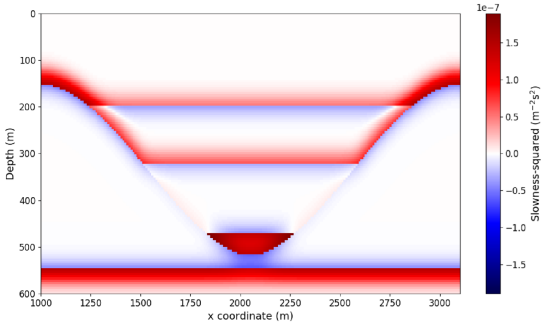
Figure 7.9: Comparison of Marchenko-Lippmann-Schwinger inversion velocity models to the true and initial models.

scaled update (figure 7.10c) is 2 orders of magnitude smaller than the true model perturbation (7.10a) , and 1 order of magnitude smaller than the common scale used to present all the slowness-squared updates from Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversions (7.10b). Shown at true amplitude (figure 7.10d), the unregularised update is extremely noisy, and the structures in 7.10a are impossible to see. With sufficient regularisation, figure 7.10e shows that all the interfaces in the true model perturbation are interpretable from the regularised inversion result. The correct relative amplitudes, particularly the positive slowness-squared increases above the second and third interfaces, are missing from the update. There are also artefacts in layer 2 due to incomplete destructive interference in the source stacking process.

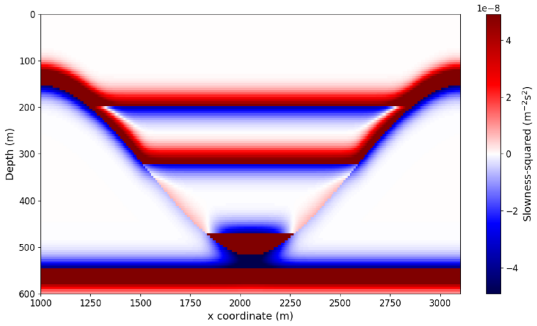
Figure 7.11 shows the velocity model and data misfits for the initial model and the models resulting from Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversions. All inversion methods resulted in an improved velocity model misfit, with damped Born and Lippmann-Schwinger inversions outperforming their unregularised counterparts. All Marchenko-Lippmann-Schwinger inversions performed comparably to unregularised Born and Lippmann-Schwinger inversions in terms of the model misfit. In the data domain, regularised Born and Lippmann-Schwinger inversions show significant improvement in the data misfit. Unregularised Born and Lippmann-Schwinger inversions show negligible improvement in the data domain, while Marchenko-Lippmann-Schwinger inversions worsen the data misfit.

7.3 Discussion

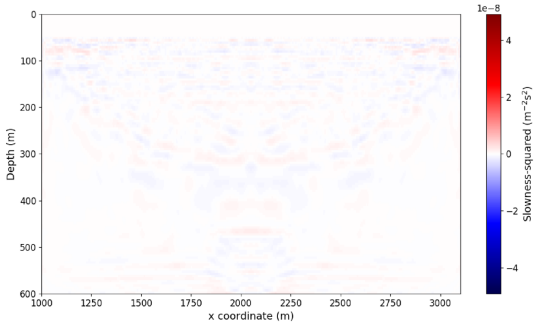
Comparing figures 7.6 and 7.11, we see that despite introducing a new, coherent reflection arrival to the data residual gathers, regularised Born and Lippmann-Schwinger inversions both significantly outperformed all other methods in terms of the data misfit. By careful examination of the data residual gathers, we can see that the amplitudes of the regularised model data gathers are significantly lower than those from the initial model and the unregularised inversions, with the maximum absolute amplitude approximately 40% lower than the data residual from the initial model. By reconsid-



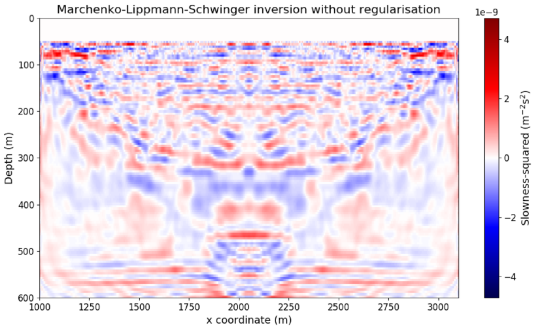
(a) True model perturbation from the initial model



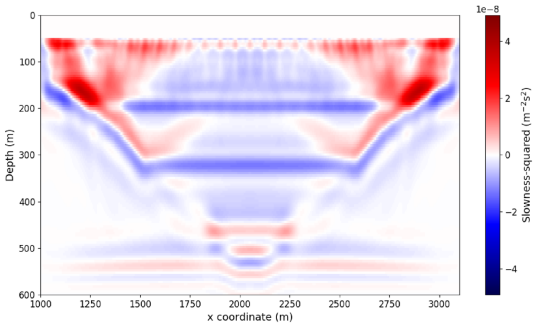
(b) True model perturbation from the initial model on a saturated scale



(c) After Marchenko-Lippmann-Schwinger inversion



(d) After Marchenko-Lippmann-Schwinger inversion (shown at true scale)



(e) After Marchenko-Lippmann-Schwinger inversion with Tikhonov damping

Figure 7.10: Comparison of slowness-squared model updates from Marchenko-Lippmann-Schwinger inversions with the true model perturbation. Panels (b), (c), and (e) are shown on a common scale, panels (a) and (d) are shown at true amplitude.

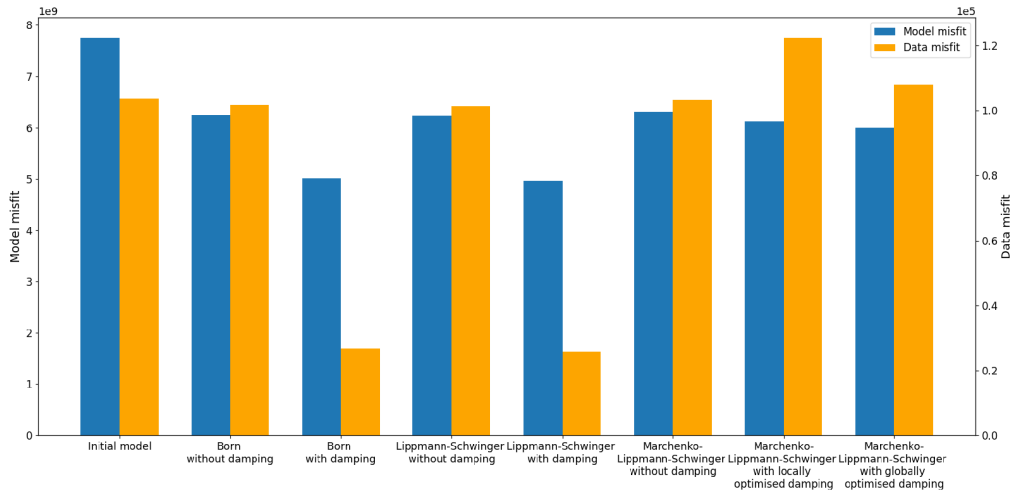


Figure 7.11: Comparison of velocity model and data misfits before and after inversion.

ering the updated velocity models in figures 7.3 and 7.4, the reduction in reflection residual amplitudes appears to correlate to the perceived increase in resolution at layer interfaces in the regularised inversion models.

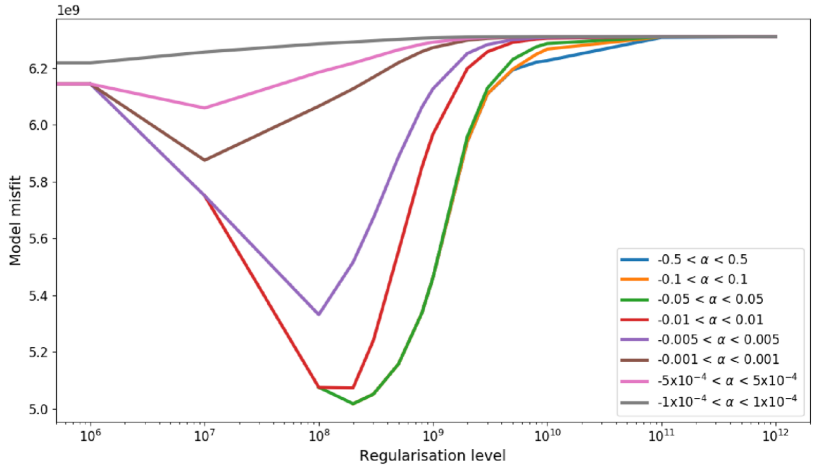
Based on the regular spacing, the vertical artefacts observed in figure 7.5 are likely the result of the source-receiver geometry, although the true cause is currently unknown. I hypothesise that these artefacts are the result of incomplete destructive interference of the wavefields from different sources, as might be observed when using an adjoint-state imaging method, and that with a higher spatial density of sources and receivers that the artefacts would be suppressed.

Further considering the introduced reflections in the regularised model data residual gathers, since the new arrival is also the first arrival, one can conclude that the cause must be located above the first reflector that is not present in the initial model. Since my initial model includes the first interface, the interface in question is between the second and third interface. Looking at the model updates in figure 7.5, we see that both regularised inversions result in a strong positive update immediately below the mute that finishes at the first interface, and above the clearly identifiable second interface. I conclude therefore that this perturbation to the velocity immediately below the first interface is likely responsible for this arrival in the data, even though such a perturbation is not obvious in the updated models.

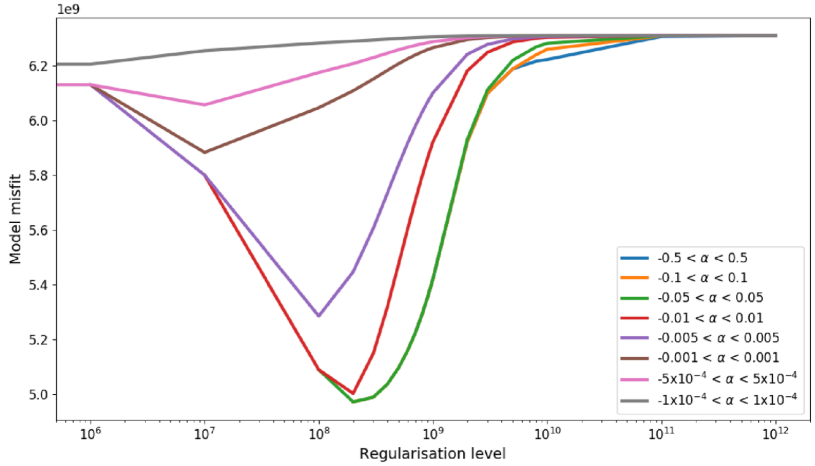
Observant readers will note figure 7.11 details two regularised Marchenko-Lippmann-

Schwinger inversions, one optimised globally and one optimised locally. This is because, unlike the Born and Lippmann-Schwinger inversions, the line-search for the step length i.e. the scalar multiplication factor that optimises the best fit between the true and updated velocity models, has two minima. The model misfits determined by the line-searches for the Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversions are shown figure 7.12. Because standard line-search algorithms such as Brent's method (Brent, 1971) and the golden-section search (Kiefer, 1953) both failed to either converge within 100,000 evaluations or find an optimal step length resulting in a physical velocity model (one with non-negative wavespeeds) when unbounded, the line-search was performed by iterating a bounded Brent search using narrowing bounds until a model misfit minimum was found. As figure 7.12 shows, the optimised misfit is highly dependent on the choice of bounding values. Ideally, one needs to look within bounds at most one order of magnitude greater than the optimal step length for the search to converge to the optimal value. Interestingly, the line searches for the Born and Lippmann-Schwinger inversions result in nearly identical trends, having the same optimal level of regularisation and almost equal model misfits at that regularisation level. By comparison, the locally and globally optimised Marchenko-Lippmann-Schwinger inversions are separated by 4 orders of magnitude in the level of regularisation required, which is 2 and 6 orders of magnitude respectively greater than the regularisation required for the Born and Lippmann-Schwinger inversions. This suggests that the Marchenko-Lippmann-Schwinger solution is far less well conditioned than the Born and Lippmann-Schwinger inversions.

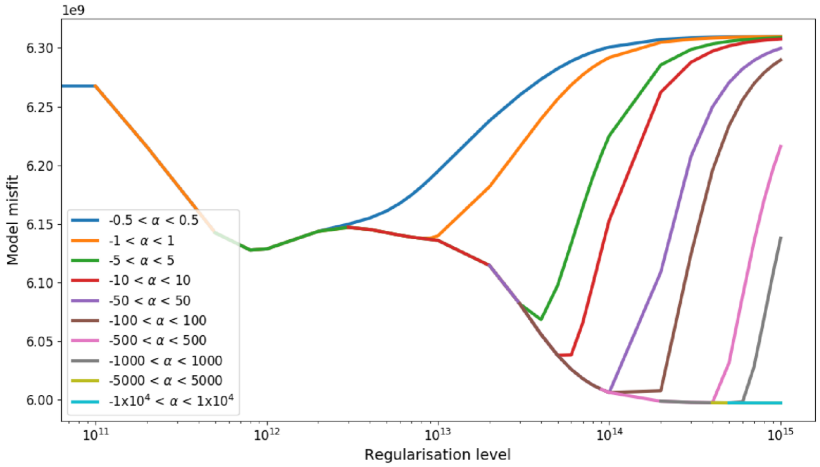
If we compare the updated model from the globally optimised Marchenko-Lippmann-Schwinger inversion ($\lambda = 1\text{x}10^{15}$) with the updated model obtained from the locally optimised inversion ($\lambda = 8\text{x}10^{11}$), shown figure 7.13, we see that the lesser regularised inversion results in an opposite polarity on the horizontal reflectors compared to the more regularised inversion, but that there are also arcing amplitude anomalies that appear to diverge from horizontal reflectors. These could be interpreted as wavefronts imprinted on the image that have not been completely removed in the stacking process by destructive interference. If this interpretation is correct (shown figures 7.13d and 7.13e), this suggests that regularisation may play a role in suppressing scattered



(a) Born inversion



(b) Lippmann-Schwinger inversion



(c) Marchenko-Lippmann-Schwinger inversion

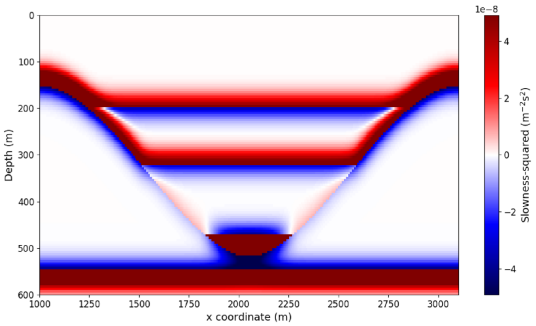
Figure 7.12: Comparison of the impact of applying different bounds to a Brent search for the optimal scaling for model updates from Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversions.

waves in the inversion. Given that a high level of regularisation is required simply to stabilise the solution, it may be that unless the scattered wavefield is of comparable amplitude to the direct arrival, scattered energy plays little or no role in the inversion.

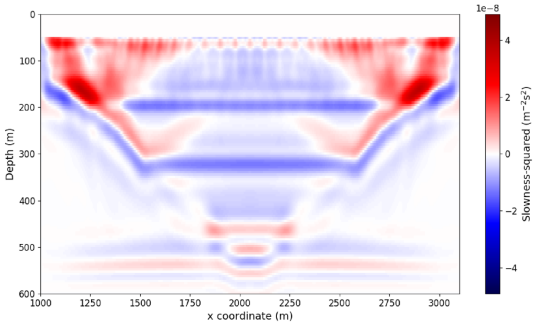
Considering all the evidence together, I must conclude that Lippmann-Schwinger inversion, with all orders of scattering in the wavefield contained in the kernel matrix, does not result in a meaningful improvement over the standard, single-scattering Born inversion approach. When Marchenko-derived Green's functions are used in place of the true Earth Green's functions, the result is worse than either Lippmann-Schwinger or Born inversions.

7.4 Summary

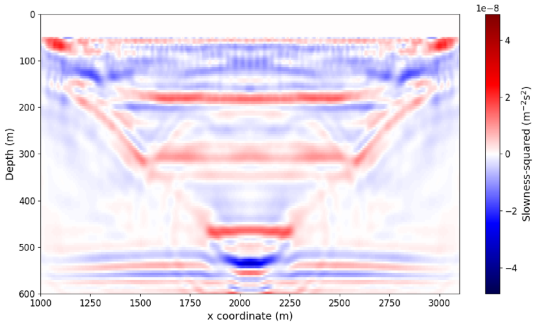
In this chapter, I have rewritten the inverse Born and Lippmann-Schwinger problems in 2D in an approximately equivalent form that allows the inverse problem to be solved using limited computational resources. By replacing a single inversion using all sources and receivers simultaneously with a stacked single source inversion approach, I have increased the number of frequencies usable in the inversion from 5 to 376, as 75-fold increase. Using only 60% of the total sources, periodic noise from the inversion process has been almost entirely eliminated from the inversion results. The inversion models and subsequent modelled data show that regularised Born and Lippmann-Schwinger inversions perform almost identically in both model and data domains. Based on this, I conclude that the higher orders of scattering contained in the Lippmann-Schwinger kernel matrix do not lead to improvements in inversion for the underlying velocity structure. By using a Marchenko method to retrieve the Green's functions from every source and receiver position to every location in the subsurface and substituting these Green's functions into the kernel of the Lippmann-Schwinger equation, the resulting inversions perform worse in both data and model domains as compared to regularised Born and Lippmann-Schwinger inversions, and comparably to unregularised Born and Lippmann-Schwinger inversions.



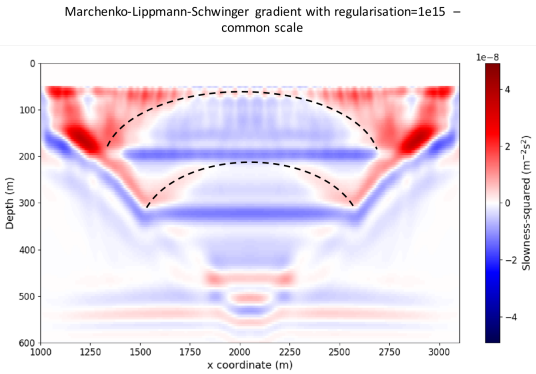
(a) True model perturbation from the initial model on saturated scale



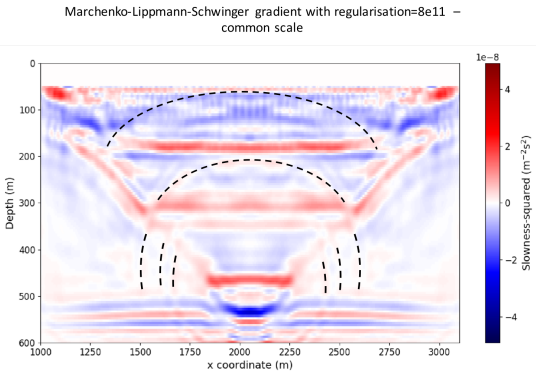
(b) After Marchenko-Lippmann-Schwinger inversion with globally optimised damping



(c) After Marchenko-Lippmann-Schwinger inversion with locally optimised damping



(d) 7.13b with a wavefront artefact interpretation



(e) 7.13c with a wavefront artefact interpretation

Figure 7.13: Comparison of slowness-squared model updates from Marchenko-Lippmann-Schwinger inversions with the true model perturbation. Panels (b), (c), and (e) are shown on a common scale, panels (a) and (d) are shown at true amplitude.

Chapter 8

Discussion

8.1 A logical order of work

A recurring theme in this thesis is that the lessons learned studying the 1D problem, particularly the analytical inversion in chapter 4.1, hold true when applied to more complex models and to higher dimensional problems. Reading this thesis, one could conclude that I had recognised these lessons and ignored them for confused reasons. The primary reason for this is that the order in which this body of work has been presented is not the order in which it was carried out. To address this issue, I will briefly discuss this thesis in the order work was carried out.

This study began by investigating the inversion of a 1D boxcar perturbation by inversion of the matrix kernel, as explored in chapter 4.2. As the simplest possible model structure besides a step function, it was assumed that if the method could not work for this example then it would not work in more complex models. From the results shown in that chapter, it is clear that the problem is extremely ill-posed and requires additional conditioning to obtain a physically reasonable velocity structure.

This lead to the study of the 1D inverse problem for the many-layered model presented in chapter 5. The decision to move on from the boxcar model to the many-layered model was motivated by the fact that a boxcar is not a good representation of the Earth's subsurface. It is difficult to draw meaningful conclusions of the validity of a method if it has only been tested in an unreasonable context. There were also good reasons to believe that moving to the many-layered model would be more

successful. First, even on the seismic scale, a single geological unit homogeneous over 1km is unrealistic. It is also difficult to suppress the oscillatory noise over such a large scale, particularly when the velocity change is so large either side of the interval. By replacing a single perturbation with many thinner perturbations, the wavelength of the oscillations becomes comparable to the depth of the layers, and in doing so it becomes difficult or impossible to distinguish between the model signal and the oscillatory noise. So rather than suppressing the oscillations, they become disguised as part of the inversion solution, and indeed, positive and negative components of the oscillations may cancel to some extent. Second, a smooth representation of the true model is used as the starting model for the many-layered inversion process, injecting accurate prior information into the problem. This is in contrast to the boxcar example, in which the starting model is homogeneous, with no prior information about the layer to guide the inversion.

As the results show, the first reason proved to be false. Without regularisation, the inversion solution was dominated by high spatial frequency oscillatory noise. Arguably, the second reason to be hopeful also proved false. In most cases, the inversion results retained the form of the smooth initial model.

At this point, the importance of having a wide range of frequencies sampled densely was also not clear. All of the 1D inversions had made use of the full frequency bandwidth up to the Nyquist frequency, and yet the inversion was still proving to be ill-posed. The clearest source of difficulty encountered in solving the 1D problem is the limitation to a single source and receiver for a surface-based acquisition geometry; from this perspective, the 1D problem is inherently under-determined. By increasing the number of dimensions from one to two, and using sufficient sources and receivers, the increase in the number of model parameters can be countered by the increase in the size of the data vector, and the problem can be changed from under-determined to formally over-determined.

The down-side is that while the number of model parameters increases linearly, and the number of data super-linearly, the size of the data kernel increases by a factor equal to the product of these increases. The computational requirements needed to build the kernel matrix and solve its inverse thus also grow super-linearly. Further

complications were added by the results of chapter 6.1, showing the need to pad and taper the edges of the model to minimise artefacts in Lippmann-Schwinger forward modelling. This increased the number of model parameters, and thus the size of the kernel, by an additional 75%, adding to the computational requirements. As a result, it was necessary to decimate the frequencies used in the inversion. From chapter 6.2, it is clear that this was a mistake. The resulting velocity model updates had low resolution and still exhibited oscillatory artefacts, with sufficient amplitude that the model updates were poorly scaled and the updated velocity models not significantly better than the initial model.

The failure of the first attempt at 2D inversion was the motivating factor to return to 1D and resulted in finding the 1D analytical inversion by Beydoun and Tarantola (1988). In attempting to reproduce their result, I came to the key conclusion that high frequency-sampling density was necessary to correctly set the amplitude of the inversion update. This conclusion was the motivating factor to reintroduce the full frequency bandwidth at the expense of using all sources in the inversion, resulting in the conclusion of chapter 7 that Lippmann-Schwinger inversion, and thus full-scattering physics, does not significantly improve the inverse scattering problem compared to the single-scattering Born method.

8.2 The dominance of single-scattering

Since the difference between the Born modelling equation and the Lippmann-Schwinger equation is the inclusion of all orders of scattering (through the Green's function G in the true model), the similarity between the inversion results of the respective methods in chapter 7 indicates that the direct wave plays a dominant role in the inversion. In chapters 6 and 7, I presented the modelled data in terms of the residual between the true-model data and the inversion-model data because the amplitude of the direct wave is approximately 10x greater than the scattered waves. This is despite the 2D velocity model having been designed to have the strongest velocity contrasts that were reasonably possible so that the reflections, primaries and multiples, had the highest possible amplitudes when recorded at the surface.

To test the theory that the inverse problem is dominated by the direct wave due to weak scattered wave amplitudes, a gain as a function of time such as a t^2 gain could be applied to the data and to the kernel. The most efficient way to do this is to apply the gain to one of the wavefields G or G_0 before building the kernel matrix, for Born inversion this means having one copy of the background model wavefield with the gain and one without. A t^2 gain biases the amplitudes of later-arriving scattered waves (especially highly scattered multiples) relative to earlier arriving waves such as the direct wave, thus the scattered wavefield can be made comparable, or even greater in amplitude than the direct wavefield. I did not test this due to having limited time, although it would be an obvious starting point for any future work.

8.3 Relation to Diekmann et al. (2021)

Concurrently with this work, a study was published by Diekmann et al. (2021)[1] on the same topic as this thesis. In their work, a single source was used to invert for the velocity inside an area of a 2D laterally inhomogeneous velocity structure in a single iteration of inversion. This study appears to be consistent and inconsistent with my work in various ways. In terms of similarities, both studies agree that large amounts of regularisation are required to stabilise the inversion solution, and both studies observe oscillatory artefacts. They also both find that having incorrect direct arrivals reduces the quality of the inversion using Marchenko Green's functions compared to having Green's functions with exact travel times. The differences are that Diekmann et al. find that Lippmann-Schwinger inversion (LSI in Diekmann et al. (2021)) and Marchenko-Lippmann-Schwinger inversion (MLSI in Diekmann et al. (2021)) are less ill-posed and return better inversion results than Born inversion, and that multiple scattering helps to stabilise the inversion.

There are several important differences between that study and my work in this thesis that may explain the differences in results. The first is that Diekmann et al. (2021) only used a single source for their study. Since it is unreasonable to expect a single source (placed in the corner of the area of interest) to sufficiently illuminate

[1]My supervisor Andrew Curtis and I are co-authors of this study.

the entire model area (of interest), the authors' interpretations of their results are based on 1D vertical profiles through the study area, so there may have been other vertical profiles where the comparison between inversion methods was less preferable. It may also turn out that when multiple sources are used and either stacked together as demonstrated in chapter 7 or inverted simultaneously using greater computational resources, that the artefacts limiting Born inversion are cancelled out resulting in far more comparable results from Born inversion and Lippmann-Schwinger inversion, as demonstrated in this thesis.

The second major difference is that Diekmann et al. (2021) apply their inversions to the problem of target-oriented imaging. That is, they do not seek to solve for their entire 2D velocity model, but rather a subsection of it. As the authors acknowledge, this means that scattered waves from outside of the area of interest can be scattered into the area of interest. As we saw from chapter 6.1, the Lippmann-Schwinger equation only correctly reconstructs the scattered wavefield when there is no incoming energy from outside the volume contributing to the Lippmann-Schwinger integral. This is a consequence of the Sommerfeld condition needed in the derivation of the Lippmann-Schwinger equation. Any such back-scattered energy will therefore result in errors in the kernel of the Lippmann-Schwinger forward problem, and thus the inverse kernel also. When their example is extended to more sources, to better illuminate the model area of interest, they may experience artefacts that more greatly affect the fully scattered Green's functions than the smooth model Green's functions due to the greater amount of back-scattering.

A third major difference is that the authors had time to optimise the Marchenko redatuming process twice, once for Marchenko redatuming using Green's functions with exact traveltimes (i.e. using windowed arrivals from the true model Green's functions), and again using direct arrivals from the smooth initial model. In contrast, the time available to me to optimise the Marchenko Green's functions for the examples in chapter 7 was limited by the time remaining to submit this thesis. To produce the Marchenko Green's functions, I manually optimised 8 relevant parameters at 52 points in the model (out of a total of 89091 model cells) using the least-squares data misfit. I then interpolated the parameter values found at these points throughout the model.

It is possible that with significantly more time the Marchenko redatuming could be improved, but without using direct arrivals taken from the true model, it cannot be as accurate as the exact Green's functions from finite-difference modelling in the true model.

A final difference is in the interpretation of the results, and specifically what constitutes an improvement significant enough to warrant using one method over another. Any interpretation is subjective, and the results found by Diekmann et al. (2021) show that all the updated models closely resembled the initial model used in the inversion. Given that Marchenko is an expensive process, both computationally and in time, and the result cannot in practice be perfect, I question whether the observed improvement is sufficiently significant to warrant replacing Born inversion with Lippmann-Schwinger or Marchenko-Lippmann-Schwinger inversion without further study - particularly since my further study in this thesis appears to show no real improvement from Marchenko-Lippmann-Schwinger at all.

8.4 The Marchenko method

An important limitation of this study and that of Diekmann et al. is the viability of Marchenko redatuming to obtain the Green's functions needed to solve the Lippmann-Schwinger equation as a linear system. Although Marchenko methods are relatively young in development (starting with Broggini et al. (2012)), Marchenko redatuming, and a type of deconvolutional imaging called Marchenko imaging (Wapenaar et al., 2014b), have been tested in numerous studies in the 10 years following. This includes both synthetic and real-data studies. Being a young field of study, there are still issues that need to be resolved.

A well-known problem is the issue of the so-called "scaling factor". Marchenko methods require that the source signature for the seismic reflection acquisition is known exactly. If the source signature is not known exactly, then the wavelet will be estimated from the data, and the resulting deconvolution of the source wavelet from the reflection response will have the wrong amplitude. Without the inclusion of this scaling factor to correct the amplitude of the reflection response, the Marchenko

algorithm may not converge, and even if it does it may converge to a Green's function estimate that is significantly in error. Therefore, it is essential to be able to accurately calculate the correct scaling factor from the data. The first method to do this was the minimum energy criterion of van der Neut et al. (2015) and Brackenhoff (2016), also known as the "j-curve". The criterion says that the correct scaling factor a for the reflection response at a given location in the subsurface is the value of a that minimises the ratio of the l2-norm of the upgoing Green's function G^+ after Marchenko redatuming and the l2-norm of the upgoing Green's function G_0^+ in the reference velocity model:

$$j(a) = \frac{|G^+|}{|G_0^+|} \tag{8.1}$$

The minimum energy criterion has been shown to work on synthetic and field data examples by Brackenhoff (2016), Mildner et al. (2017), and Brackenhoff et al. (2019); however, an experiment testing Marchenko imaging on ultrasonic data (da Costa Filho et al., 2018) was unsuccessful in determining the correct scaling factor using this method. An alternative method to the minimum energy criterion for determining the scaling factor was devised by Dokter et al. (2019), whilst a Marchenko method that does not require a scaling factor was demonstrated on real data by Staring et al. (2018). To date, no method has achieved ubiquity in the field, and issues related to the scaling factor are likely to persist. Unfortunately, the need to optimise the scaling factor at every image point (likely every model point) adds to the number of parameters that need to be optimised in some manner at every point of interest, or at least at sufficient locations to interpolate to all other points.

The other major problem with all but one Marchenko methods, at least with respect to inversion with Marchenko-derived Green's functions, is that the Marchenko method is applied at each model point, with the Green's function retrieved at each location obtained independent of that from any other point. This means that the natural correlations in amplitude and phase between neighbouring model points may not be preserved in the Marchenko-derived Green's functions. When the aim is simply to calculate the optimal image at each point separately, without requirement for accurate amplitudes, it is less important that amplitudes and phases be consistent across neighbouring pixels. In Marchenko-Lippmann-Schwinger inversion, we are using

the wavefield at all model points simultaneously. If the amplitudes and phases are not correct, or at least consistent across model points, the kernel of the forward problem will not reproduce the observed data, and thus we cannot expect the kernel inverse to encode the correct relationship between the data and the model. Without further advances that allow for redatuming at multiple physical locations simultaneously, Marchenko methods may prove to be inadequate for use in inverse problems that require the solution for the entire medium to be solved simultaneously i.e. explicit matrix inversions.

8.5 Future research

In chapter 2, I derived the cost function gradient for a new full-waveform inversion scheme that employs full scattering Green's functions derived from a Marchenko method. This is called Marchenko-Lippmann-Schwinger full-waveform inversion (MLS-FWI). In light of the conclusion of chapter 7, a valid question would be whether there is value in researching other inversion methods employing Marchenko Green's functions?. The answer is yes, for several reasons. First, and perhaps most importantly, the fundamental difference between the work presented in this thesis and MLS-FWI is that the latter does not require the explicit inversion of the kernel matrix, whereas the former does. We have seen in this thesis that the presence of strong oscillations in the inversion solution is a serious problem in both one and two dimensional inversions. In principle, this can be avoided in MLS-FWI because the explicit matrix inversion is replaced by a series of forward modelling and convolution operations. As shown in chapter 6.1, the Lippmann-Schwinger forward problem is far more stable than the inverse problem. Thus constructing the inverse problem solution using only forward modelling may resolve much of this instability.

A further advantage of the full-waveform inversion formulation of Marchenko-Lippmann-Schwinger inversion is that there is now a plethora of methods based on the most basic full-waveform inversion formalism I assume in my methodology. As long as the model update direction can be defined in terms of the Fréchet derivative of the wavefield, the full Green's function derivative can be integrated with the method.

Research into full-waveform inversion has focused on two different directions (Métivier et al., 2016): improved and more accurate physics allowing inversion for physical parameters other than the P- or S-wavespeed, and improving the applicability and robustness of full-waveform inversion by adjustment of the cost function of FWI.

Applicability of the Marchenko-Lippmann-Schwinger method to FWI schemes based on more realistic physics is the simplest to envisage. The extension to acoustic physics with variable density and velocity is the most obvious, since the only change to the Lippmann-Schwinger equation is to the form of the scattering potential to include a density term, and there is no change to the Marchenko methodology. In principle, there is no limitation to the physics with which the Marchenko-Lippmann-Schwinger method can be used, since there exists a Lippmann-Schwinger equation for every type of physics and an appropriate Marchenko scheme said assumed physics. The obvious downside to using increasingly complex physics is the associated increase in computational cost that comes with modelling the wavefield for an increasing number of physical parameters.

Research into more robust formulations of full-waveform inversion has resulted in a diverse range of methods, many of which aim to overcome the problem of cycle-skipping. The problem of cycle-skipping can be overcome by use of a hierarchical approach (Bunks et al., 1995), beginning with very low frequencies where the period of a half-cycle is greatest and slowly introducing higher and higher frequencies as the accuracy and resolution of the model increases. Unfortunately, by itself the hierarchical approach is insufficient to overcome the required accuracy needed of the initial model for the inversion to converge. Since the cause of the cycle-skipping problem is the least-squares data misfit, there is a need for alternative methods utilising alternate cost functions. One such method replaces the least-squares distance between the data with the so-called "optimal transport" distance (Métivier et al., 2016). For the example of two Ricker wavelets the local minima in the least-squares data misfit cost function space are replaced by a single global minimum in the optimal transport cost function space.

Another method that aims to solve the problem of cycle-skipping is adaptive waveform inversion (AWI) (Warner and Guasch, 2016). Here, the problem of opti-

mising the distance between waveforms is replaced by the problem of designing an optimal Wiener filter that matches the modelled data to the observed data. This filter is constructed by deconvolving the auto-correlation of the modelled data from the cross-correlation of the observed and modelled data, whilst the cost function to be optimised is the matching filter normalised by the filter weights. The issue with incorporating Marchenko-Lippmann-Schwinger derivatives into such methods is that the Marchenko method itself requires an initial velocity model that does not cause cycle-skipping, thus the main appeal of such methods is diminished. Additionally, the Wiener matching filters used in AWI have limited ability to predict arrivals not visible in the data, meaning that the method already partially incorporates the novel feature of Marchenko-Lippmann-Schwinger methods.

A third approach to tackling cycle-skipping is seismic envelope inversion (Luo and Wu, 2015). Rather than changing the metric used to optimise the inversion, or optimising something other than the data, seismic envelope inversion instead transforms the data to its envelope function and optimising the least-squares misift of the observed and modelled data envelopes. Using the envelope enhances the very low frequencies in the data, which when applied in a hierarchical manner reduces the sensitivity to the starting model, and can accelerate the convergence of standard FWI when envelope inversion is used to obtain the starting model for FWI. An appealing aspect of envelope inversion is that the Marchenko-Lippmann-Schwinger derivative could be easily substituted into the formalism. Alternatively, since the major advantage of Marchenko-Lippmann-Schwinger methods is that the high wavenumbers can be correctly handled from the first iteration where Born methods should only correctly handle the low wavenumbers, envelope inversion could be used to accelerate MLS-FWI in the same manner that it can for FWI.

A final modification of full-waveform is wavefield reconstruction inversion (WRI). In this scheme, instead of only inverting for the model parameters, one also inverts for the wavefield inside the true medium as a joint inversion (Van Leeuwen and Herrmann, 2013). An interesting development in this field is the use of an approach similar to WRI to invert for the focusing functions for Marchenko (Hajjaj et al., 2022), which in turn can be used to obtain the Green's functions. Thus there appears to already

be a natural opening for Marchenko Green's functions to be introduced to wavefield reconstruction inversion.

The second reason why continued research in this field is worthwhile is that while all these methods are named after Marchenko methods, there is no requirement for a Marchenko method to be used. It is only because there currently exists no alternative method that can retrieve full scattering Green's functions with one-sided illumination of the Earth i.e. using only data acquired at the surface of the Earth. If a new method is developed that can also do this, then it would be equally valid to substitute those Green's functions into the inversion formulae.

The third is the potential of the MLS-FWI gradient to be applied to other currently existing methods. Fundamentally, the gradient derivation for MLS-FWI is a derivation for the derivative of the least-squares norm of the data residual using full scattering Green's functions. Least-squares inversion is perhaps the most ubiquitous method of inversion by optimisation in geophysics today. If it could be shown that full scattering improves full-waveform inversion, then it may also improve any least-squares inverse scattering problem. For example, I have derived a form of least-squares migration that I call Marchenko-Lippmann-Schwinger Least-Squares Migration (MLS-LSM). This is only one possible application in geophysics.

There is also further research to be done on the Lippmann-Schwinger inverse problem by explicit matrix inversion. A question that this thesis poses is "Why is the 1D inverse scattering problem so ill-posed when the 2D inverse problem is stable?". Since the number of model parameters in the 1D problem is almost always significantly less than a 2D problem[2], the ratio of the number of data when using all available frequencies in the data to the number of model parameters is likely to be massive when compared to the 2D problem. There are also no issues of limited illumination in 1D that can reduce the effectiveness of 2D inversions. The reason for the instability of the 1D problem is therefore unclear, and its solution may have relevance to other 1D inverse problems.

There is also promising research by Diekmann and Vasconcelos (2020) using

[2]It is of course possible to design a very deep, very fine resolution 1D model with more model parameters than a coarse resolution, shallow 2D model. But the need for such a model is unknown, and it is unlikely someone would create such a model.

Marchenko Green's functions to create virtual receivers everywhere in the subsurface that can then be used in the data vector to solve the 2D inverse Lippmann-Schwinger problem. When the exact full-scattering Green's functions modelled in the true velocity model are used, the resulting inversion velocity models are an exact match to the true model. With computational resources increasing all the time, improvements in the Marchenko methodology may soon lead to high resolution velocity model estimation.

It is clear then that despite the results of this thesis showing no benefit of Marchenko-Lippmann-Schwinger imaging over Born imaging, there remains significant research to be done to explore the potential of the inverse Lippmann-Schwinger problem and the use of Marchenko-derived Green's functions in inverse problems.

Chapter 9

Conclusion

Accurate seismic velocity models are essential for producing accurate images in depth of the Earth's subsurface from seismic reflection data. The reflections observed at the surface are themselves due, in part, to changes in the velocity structure of the Earth, and the images we make are a representation of this velocity structure. This is a symbiotic relationship - the reflection data is used to infer the velocity structure, which is used to image the Earth with the data, that in turn can improve the estimated velocity structure *ad nauseum*. Typically, only singly-scattered waves are used in this process because we cannot model higher scattering waves accurately, and great effort must be made to remove highly scattered waves from our reflection data as a result. The single scattering assumption is known as the Born approximation, and the process of obtaining the Earth's velocity structure from singly scattered data is called Born inversion.

With the development of "Marchenko" methods, that solve the Marchenko equation for the fully scattering waves using only the reflection data and a smooth velocity model, we can now solve for the Earth's velocity structure using all orders of scattering in the data. The relationship between the fully scattering data and the velocity structure of the Earth is called the Lippmann-Schwinger equation, and the process of solving the equation for velocities is called Lippmann-Schwinger inversion. When the scattered waves from Marchenko methods are used, I call it Marchenko-Lippmann-Schwinger inversion. The aim of this thesis was to determine if using all orders of scattering in Lippmann-Schwinger inversion results in more accurate velocity models

than the single scattering Born inversion. Current state-of-the-art inversion methods are largely based on the Born approximation, requiring that the data be processed (inexactly) to remove the higher-order scattered waves before the inverse problem can be solved. By determining if the inclusion of higher-order scattering in the physics of the inverse problem improves the accuracy of the solution, we are also determining the impact that assuming the Born approximation has on the solution to the inverse problem. This was previously not well understood because we lacked the methods to construct the full-scattering kernel of the Lippmann-Schwinger equation based solely on data acquired at the surface of the Earth and a smooth representation of the Earth, which is the typical starting point for the inverse scattering problem.

In chapter 2 of this thesis, I define the acoustic inverse scattering problem for constant density, and showed that the Born approximation is a linearization of the Lippmann-Schwinger equation. A brief review of Marchenko methods is given. I also define the derivative of the Lippmann-Schwinger equation with respect to the scattering potential, for use in a proposed full-waveform inversion that conditions the gradient direction using full scattering Green's functions. I call this Marchenko-Lippmann-Schwinger full-waveform inversion.

In chapter 3, I review my experimental design choices and the modelling workflow employed, with consideration given to limitations on available computational resources.

In chapter 4.1, I repeat the experiment of Beydoun and Tarantola (1988), using analytical descriptions of the singly-scattered wavefield and the unknown perturbation to a homogeneous velocity model to solve the 1D Born inverse scattering problem for the case of a boxcar perturbation. This shows that unless deconvolution is applied to the wavefields, the source wavelet used will be imprinted on the velocity model update. Increasing the frequency-sampling density is shown to improve the recovered amplitudes of the velocity model update, and to reduce any DC shifts in the inversion result. Most importantly, it shows that even when one has analytic formulae providing exact single scattering wavefields with no errors, that Born inversion cannot reproduce the true medium exactly. There will always be an error when the single scattering assumption is used.

In chapter 4.2, I compare the results from the analytical inversion of the boxcar perturbation to Born, Lippmann-Schwinger, and Marchenko-Lippmann-Schwinger inversions for the same model, using wavefields modelled by finite-differences and explicitly calculating the inverse kernel matrix. The results show that for any method of inversion, the 1D inverse scattering problem is extremely ill-posed. By applying Tikhonov damping or smoothing regularisation, the locations of the boxcar interfaces can be identified. These results can be further improved by iterating the inversion process. Even with regularisation and iteration, the solution contains significant oscillatory artefacts that deteriorate the solutions.

In chapter 5, I replace the boxcar model with a many-layered model, using a smooth representation of the true model as the starting model in place of a homogeneous model. It is found that prior information may have an important role in suppressing oscillatory noise. Optimising the inversion with the model misfit does not guarantee an improvement in the data misfit, with the reverse true for optimisation using the data misfit. When the inverse problem is iterated, iteration is less effective for data optimisation than model optimisation, with the inverse problem generally converging within fewer than 10 iterations. The under-performance of Marchenko-Lippmann-Schwinger compared to Lippmann-Schwinger inversion across all regularisation and optimisation methods implies that flaws in Marchenko Green's function retrieval inhibit the performance of MLS inversion.

In chapter 6, I solve the 2D Born and Lippmann-Schwinger inverse models for a syncline model with horizontal layers. Due to computational limitations, using all 141 sources and receivers requires the number of frequencies be reduced from 376 to 5. Without regularisation, the model updates from both inversion methods are contaminated by the same oscillatory noise observed in the 1D inverse problem. With strong Tikhonov damping applied, the oscillations are reduced and the perturbations at some interfaces are revealed. The updated models, however, are low resolution and not significantly different to the initial model.

In chapter 7, I write the inverse problem so that a single simultaneous inversion of all sources and receivers is replaced by single source, many receiver inversions that can be stacked like a migration image to form a single model update. By reducing

the number of sources, I am able to reinstate the full frequency bandwidth. When regularisation is applied, the method of stacked images results in model updates that are almost entirely free of oscillatory artefacts. The models obtained from Born and Lippmann-Schwinger inversion are almost identical. Marchenko-Lippmann-Schwinger inversion results in a model worse than either Born or Lippmann-Schwinger inversions, as measured with the model misfit and the data misfit of data modelled in the updated velocity model. On this basis that Born and Lippmann-Schwinger perform equally well, I conclude that including higher order scattering in inversion does not result in a better velocity model.

In chapter 8, I discuss the consequences and limitations of my work, as well as future research that might advance the field further. The immediate questions posed by this research can be grouped into three areas: the dominance of single scattering, the instability of the 1D inverse problem, and the potential of full-scattering adjoint methods.

Despite efforts to increase the scattering strength by introducing strong velocity contrasts, the scattering wavefield exhibits very low amplitudes compared to the direct wave, especially in the coda where higher-order scattering events are found. This means higher-order scattering events contribute relatively little energy to the inversion, and consequently the problem is more similar to the Born single-scattering problem than expected. The natural method to increase the scattering amplitudes is to introduce density perturbations in addition to velocity perturbations, as is observed in the real Earth. A simpler and faster method would be to apply a t^2 gain to the wavefield to bias the highly-scattering wavefield over the initial arrivals in the data and the kernel, although this approach will also amplify errors in the Marchenko Green's function retrieval.

A further important question is to determine why the one dimensional inverse scattering problem is so ill-posed. This is especially concerning given the fact that the ill-posedness of the two dimensional inverse problem was able to be overcome with the inclusion of a wide band of frequencies in the data and strong Tikhonov regularisation. Since the 1D problem is nominally simpler, with effectively infinite offset data and no possibility of artefacts present in two dimensional problems such as

cross-talk between sources and energy lost to evanescent waves, the fact that the 1D problem could not be properly constrained even with strong regularisation poses the potential for currently unknown characteristics of the inverse scattering problem to be discovered. Such characteristics may prove invaluable to improving the solutions to the inverse scattering problem in higher dimensions, a lesson frequently learned in this thesis.

The most important question that is left unanswered by this work is whether Marchenko-Lippmann-Schwinger full-waveform inversion, a new full-waveform inversion scheme derived in this thesis that uses all orders of scattering in the derivatives used to define the update direction, is more effective than full-waveform inversion schemes that use the Born approximation. Even more promising is that since the definition of the derivative is not specific to full-waveform inversion, the Marchenko-Lippmann-Schwinger method may be extendable to other least-squares optimisation problems that presently require the Born approximation. The ultimate promise of this thesis may be in proving the effectiveness of any one of these possible new methods that may open up entirely new fields of inverse methods in geophysics.

Bibliography

Abramowitz, M. and I. A. Stegun (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Vol. 55. US Government printing office.

Abubakar, A., W. Hu, T. M. Habashy, and P. M. van den Berg (2009). "Application of the finite-difference contrast-source inversion algorithm to seismic full-waveform data". In: *Geophysics* 74.6, WCC47–WCC58.

Agranovich, Z. and V. Marchenko (1964). *The Inverse Problem of Scattering Theory*. Russian Monographs and Texts on the Physical Sciences. Gordon and Breach. ISBN: 9780677200101.

Agudo, O. C., L. Guasch, P. Huthwaite, and M. Warner (2018). "3D imaging of the breast using full-waveform inversion". In: *Proc. Int. Workshop Med. Ultrasound Tomogr.* Pp. 99–110.

Aki, K. and W. Lee (1976). "Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes: 1. A homogeneous initial model". In: *Journal of Geophysical research* 81.23, pp. 4381–4399.

Aktosun, T. and J. H. Rose (2002). "Wave focusing on the line". In: *Journal of Mathematical Physics* 43.7, pp. 3717–3745.

Alam, A. and J. Austin (1981). "Multiple attenuation using slant stacks: Tech". In: *Rep., Western Geophysical Company*.

Alkhalifah, T. (2016). *Full Waveform Inversion in an Anisotropic World: Where are the Parameters Hiding?* Education tour series. EAGE Publications. ISBN: 9789073834835.

Araújo, F. V., A. B. Weglein, P. M. Carvalho, and R. H. Stolt (1994). "Inverse scattering series for multiple attenuation: An example with surface and internal multiples". In: *SEG Technical Program Expanded Abstracts 1994*. Society of Exploration Geophysicists, pp. 1039–1041.

Berkhout, A. (1987). *Applied Seismic Wave Theory*. Advances in Seismic Wave Theory Series v. 1. Elsevier. ISBN: 9780444428998.

Berkhout, A. J. and D. Verschuur (1997). "Estimation of multiple scattering by iterative inversion, Part I: Theoretical considerations". In: *Geophysics* 62.5, pp. 1586–1595.

Berkhout, A. J. (1980). *Seismic Migration: Imaging of Acoustic Energy by Wave Field Extrapolation..: Imaging of Acoustic Energy by Wave Field Extrapolation*. Elsevier.

Beydoun, W. B. and A. Tarantola (1988). "First Born and Rytov approximations: Modeling and inversion conditions in a canonical example". In: *The Journal of the Acoustical Society of America* 83.3, pp. 1045–1055.

Biondi, B. and P. Sava (1999). "Wave-equation migration velocity analysis". In: *SEG Technical Program Expanded Abstracts 1999*. Society of Exploration Geophysicists, pp. 1723–1726.

Bishop, T., K. Bube, R. Cutler, R. Langan, P. Love, J. Resnick, et al. (1985). "Tomographic determination of velocity and depth in laterally varying media". In: *Geophysics* 50.6, pp. 903–923.

Bleistein, N., J. Cohen, W. John Jr, et al. (2001). *Mathematics of Multidimensional Seismic Imaging, Migration, and Inversion*. Vol. 13. Springer Science & Business Media.

Born, M. (1926). "Quantenmechanik der stoßvorgänge". In: *Zeitschrift für physik* 38.11, pp. 803–827.

Brackenhoff, J. (2016). "Rescaling of incorrect source strength using Marchenko redatuming". MA thesis.

Brackenhoff, J., J. Thorbecke, and K. Wapenaar (2019). "Virtual sources and receivers in the real Earth: Considerations for practical applications". In: *Journal of Geophysical Research: Solid Earth* 124.11, pp. 11802–11821.

Brent, R. P. (1971). "An algorithm with guaranteed convergence for finding a zero of a function". In: *The computer journal* 14.4, pp. 422–425.

Broggini, F., R. Snieder, and K. Wapenaar (2012). "Focusing the wavefield inside an unknown 1D medium: Beyond seismic interferometry". In: *Geophysics* 77.5, A25–A28.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent (1995). "Multiscale seismic waveform inversion". In: *Geophysics* 60.5, pp. 1457–1473.

Červený, V. (1985). "Ray synthetic seismograms for complex two-dimensional and three-dimensional structures". In: *J. geophys* 58.2, p. 26.

Claerbout, J. (2014). *Geophysical image estimation by example*. Self published. ISBN: 978-1-312-47467-3. URL: `https://sepwww.stanford.edu/sep/prof/gee/book-sep.pdf`.

Cohen, J. and J. Stockwell (2017). *CWP/SU: Seismic Unix Release No. 44: a free package for seismic research and processing*. Center for Wave Phenomena, Colorado School of Mines.

Cohen, J. K. and N. Bleistein (1979). "Velocity inversion procedure for acoustic waves". In: *Geophysics* 44.6, pp. 1077–1087.

Courant, R., K. Friedrichs, and H. Lewy (1967). "On the partial difference equations of mathematical physics". In: *IBM journal of Research and Development* 11.2, pp. 215–234.

Curtis, A., H. Nicolson, D. Halliday, J. Trampert, and B. Baptie (2009). "Virtual seismometers in the subsurface of the Earth from seismic interferometry". In: *Nature Geoscience* 2.10, pp. 700–704.

Curtis, A. and D. Halliday (2010). "Source-receiver wave field interferometry". In: *Physical Review E* 81.4, p. 046601.

da Costa Filho, C. A., M. Ravasi, A. Curtis, and G. A. Meles (2014). "Elastodynamic Green's function retrieval through single-sided Marchenko inverse scattering". In: *Physical Review E* 90.6, p. 063201.

da Costa Filho, C. A., G. A. Meles, and A. Curtis (2017). "Elastic internal multiple analysis and attenuation using Marchenko and interferometric methods". In: *Geophysics* 82.2, Q1–Q12.

da Costa Filho, C. A., K. Tant, A. Curtis, A. Mulholland, and C. M. Moran (2018). "Using laboratory experiments to develop and test new Marchenko and imaging methods". In: *SEG Technical Program Expanded Abstracts 2018*. Society of Exploration Geophysicists, pp. 4352–4356.

DeSanto, J. A. (1992). *Scalar wave theory*. Springer.

Diekmann, L. and I. Vasconcelos (2020). "Imaging with the exact linearised Lippmann-Schwinger integral by means of redatumed in-volume wavefields". In: *SEG International Exposition and Annual Meeting*. OnePetro.

Diekmann, L., I. Vasconcelos, D. Cummings, and A. Curtis (2021). "Towards exact linearized full-waveform inversion via Marchenko redatuming". In: *SEG/AAPG/SEPM First International Meeting for Applied Geoscience & Energy*. OnePetro.

Dokter, E., A. Curtis, and J. Rickett (2019). "Marchenko redatuming: A novel approach to scaling factor estimation". In: *EGU General Assembly Conference Abstracts*, p. 19215.

Dorren, H., E. Muyzert, and R. Snieder (1994). "The stability of one-dimensional inverse scattering". In: *Inverse Problems* 10.4, p. 865.

Dziewonski, A. M., B. H. Hager, and R. J. O'Connell (1977). "Large-scale heterogeneities in the lower mantle". In: *Journal of Geophysical Research* 82.2, pp. 239–255.

Embree, P., J. P. Burg, and M. M. Backus (1963). "Wide-band velocity filtering—The pie-slice process". In: *Geophysics* 28.6, pp. 948–974.

Fail, J. and G. Grau (1963). "Les filtres en eventail". In: *Geophysical prospecting* 11.2, pp. 131–163.

Fichtner, A. (2010). *Full seismic waveform modelling and inversion*. Springer Science & Business Media.

Fomel, S., P. Sava, I. Vlad, Y. Liu, and V. Bashkardin (2013). "Madagascar: Open-source software project for multidimensional data analysis and reproducible computational experiments". In: *Journal of Open Research Software* 1.1.

Freter, H. (1992). "An Integral Equation Method for Seismic Modelling and Inversion". In: *Theory and Practice of Geophysical Data Inversion: Proceedings of the 8th International Mathematical Geophysics Seminar on Model Optimization in Exploration Geophysics 1990*. Ed. by A. Vogel, A. K. M. Sarwar, R. Gorenflo, and O. I. Kounchev. Wiesbaden: Vieweg+Teubner Verlag, pp. 243–264. ISBN: 978-3-322-89417-5. DOI: 10.1007/978-3-322-89417-5_16.

Gelfand, I. and B. Levitan (1951). "Amer. Math. Transl." In: *On the determination of a differential equation from its spectral function* 1, pp. 239–253.

Guasch, L., M. Warner, and C. Ravaut (2019). "Adaptive waveform inversion: Practice". In: *Geophysics* 84.3, R447–R461.

Guasch, L., O. Calderón Agudo, M.-X. Tang, P. Nachev, and M. Warner (2020). "Full-waveform inversion imaging of the human brain". In: *NPJ digital medicine* 3.1, pp. 1–12.

Gubbins, D. (2004). *Time series analysis and inverse theory for geophysicists*. Cambridge University Press.

Hadamard, J. (1902). "Sur les dérivées des fonctions de lignes". In: *Bulletin de la Société Mathématique de France* 30, pp. 40–43.

Hajjaj, R. F., S. A. L. de Ridder, P. W. Livermore, and M. Ravasi (2022). *Wavefield reconstruction inversion modelling of Marchenko focusing functions*. arXiv: 2210.14570 [physics.geo-ph].

Hampson, D. (1986). "Inverse velocity stacking for multiple elimination". In: *SEG Technical Program Expanded Abstracts 1986*. Society of Exploration Geophysicists, pp. 422–424.

Hansen, P. C. (1992). "Analysis of Discrete Ill-Posed Problems by Means of the L-Curve". In: *SIAM Review* 34.4, pp. 561–580. DOI: 10.1137/1034115. eprint: https://doi.org/10.1137/1034115.

Jost, R. and W. Kohn (1952). "Construction of a potential from a phase shift". In: *Physical Review* 87.6, p. 977.

Kiefer, J. (1953). "Sequential minimax search for a maximum". In: *Proceedings of the American mathematical society* 4.3, pp. 502–506.

Kosloff, D., A. Q. FILHO, E. TESSMER, and A. BEHLE (1989). "Numerical solution of the acoustic and elastic wave equations by a new rapid expansion method 1". In: *Geophysical prospecting* 37.4, pp. 383–394.

Kosloff, D. D. and E. Baysal (1982). "Forward modeling by a Fourier method". In: *Geophysics* 47.10, pp. 1402–1412.

Kouri, D. J. and A. Vijay (2003). "Inverse scattering theory: Renormalization of the Lippmann-Schwinger equation for acoustic scattering in one dimension". In: *Physical Review E* 67.4, p. 046614.

Lailly, P. (1983). "The seismic inverse problem as a sequence of before stack migration: Conference on inverse scattering: Theory and applications: Society for Industrial and Applied Mathematics". In: *Expanded Abstracts*.

Lamb, G. L. (L. (1980). *Elements of soliton theory*. eng. Pure and applied mathematics. New York: Wiley. ISBN: 0471045594.

Lesage, A.-C., R. Eftekhar, D. J. Kouri, J. Yao, and F. Hussain (2013). "Inverse acoustic scattering series using the Volterra renormalization of the Lippmann-Schwinger equation". In: *SEG Technical Program Expanded Abstracts 2013*, pp. 4645–4649. DOI: 10.1190/segam2013-1147.1. eprint: https://library.seg.org/doi/pdf/10.1190/segam2013-1147.1.

Lippmann, B. A. and J. Schwinger (1950). "Variational principles for scattering processes. I". In: *Physical Review* 79.3, p. 469.

Liu, W., A. M. Papovici, D. Bevc, and B. Biondi (2001). "3D migration velocity analysis for common image gathers in the reflection angle domain". In: *SEG Technical Program Expanded Abstracts 2001*. Society of Exploration Geophysicists, pp. 885–888.

Liu, Z. and N. Bleistein (1995). "Migration velocity analysis: Theory and an iterative algorithm". In: *Geophysics* 60.1, pp. 142–153.

Liu, Z. (1997). "An analytical approach to migration velocity analysis". In: *Geophysics* 62.4, pp. 1238–1249.

Lobkis, O. I. and R. L. Weaver (2001). "On the emergence of the Green's function in the correlations of a diffuse field". In: *The Journal of the Acoustical Society of America* 110.6, pp. 3011–3017.

Lomas, A. and A. Curtis (2019). "An introduction to Marchenko methods for imaging". In: *Geophysics* 84.2, F35–F45.

Luo, J. and R.-S. Wu (2015). "Seismic envelope inversion: reduction of local minima and noise resistance". In: *Geophysical Prospecting* 63, pp. 597–614.

Marchenko, V. A. (1955). "On reconstruction of the potential energy from phases of the scattered waves". In: *Dokl. Akad. Nauk SSSR*. Vol. 104. 5, pp. 695–698.

Meles, G. A., K. Löer, M. Ravasi, A. Curtis, and C. A. da Costa Filho (2015). "Internal multiple prediction and removal using Marchenko autofocusing and seismic interferometry". In: *Geophysics* 80.1, A7–A11.

Meles, G. A., K. Wapenaar, and A. Curtis (2016). "Reconstructing the primary reflections in seismic data by Marchenko redatuming and convolutional interferometry". In: *Geophysics* 81.2, Q15–Q26.

Meles, G. A., K. Wapenaar, and J. Thorbecke (2018). "Virtual plane-wave imaging via Marchenko redatuming". In: *Geophysical Journal International* 214.1, pp. 508–519.

Meles, G. A., L. Zhang, J. Thorbecke, K. Wapenaar, and E. Slob (2020). "Data-driven retrieval of primary plane-wave responses". In: *Geophysical Prospecting* 68.6, pp. 1834–1846.

Menke, W. (1984). *Geophysical data analysis: Discrete inverse theory*. Academic press.

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux (2016). "Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion". In: *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society* 205.1, pp. 345–377.

Mildner, C., T. S. Becker, K. de Vos, F. Broggini, and J. O. Robertsson (2017). "Source wavelet estimation using Marchenko focusing functions: Theory and laboratory data example". In: *2017 SEG International Exposition and Annual Meeting*. OnePetro.

Moses, H. (1956). "Calculation of the scattering potential from reflection coefficients". In: *Physical Review* 102.2, p. 559.

Mosher, C. C., S. Jin, and D. J. Foster (2001). "Migration velocity analysis using common angle image gathers". In: *SEG Technical Program Expanded Abstracts 2001*. Society of Exploration Geophysicists, pp. 889–892.

Parker, R. L. (1994). *Geophysical inverse theory*. Vol. 1. Princeton university press.

Pestana, R. C. and P. L. Stoffa (2010). "Time evolution of the wave equation using rapid expansion method". In: *Geophysics* 75.4, T121–T131.

Pratt, R. G., C. Shin, and G. Hick (1998). "Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion". In: *Geophysical journal international* 133.2, pp. 341–362.

Pratt, R. G. (1999). "Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model". In: *Geophysics* 64.3, pp. 888–901.

Pratt, R. G., L. Huang, N. Duric, and P. Littrup (2007). "Sound-speed and attenuation imaging of breast tissue using waveform tomography of transmission ultrasound data". In: *Medical Imaging 2007: Physics of Medical Imaging*. Vol. 6510. International Society for Optics and Photonics, 65104S.

Prosser, R. T. (1969). "Formal solutions of inverse scattering problems". In: *Journal of Mathematical Physics* 10.10, pp. 1819–1822.

Prosser, R. T. (1976). "Formal solutions of inverse scattering problems. II". In: *Journal of Mathematical Physics* 17.10, pp. 1775–1779.

Rajan, S. D. and G. V. Frisk (1989). "A comparison between the Born and Rytov approximations for the inverse backscattering problem". In: *Geophysics* 54.7, pp. 864–871.

Ravasi, M., I. Vasconcelos, A. Kritski, A. Curtis, C. A. da Costa Filho, and G. A. Meles (2016). "Target-oriented Marchenko imaging of a North Sea field". In: *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society* 205.1, pp. 99–104.

Ravasi, M. (2017). "Rayleigh-Marchenko redatuming for target-oriented, true-amplitude imaging". In: *Geophysics* 82.6, S439–S452.

Rawlinson, N. and M. Sambridge (2003). "Seismic traveltime tomography of the crust and lithosphere". In: *Advances in geophysics* 46, pp. 81–199.

Rawlinson, N., S. Pozgay, and S. Fishwick (2010). "Seismic tomography: a window into deep Earth". In: *Physics of the Earth and Planetary Interiors* 178.3-4, pp. 101–135.

Rose, J. H. (2001). ""Single-sided" focusing of the time-dependent Schrödinger equation". In: *Physical Review A* 65.1, p. 012707.

Rose, J. H. (Nov. 2002). " Single-sided autofocusing of sound in layered materials". In: *Inverse Problems* 18.6, pp. 1923–1934. DOI: 10.1088/0266-5611/18/6/329.

Rudin, L. I., S. Osher, and E. Fatemi (1992). "Nonlinear total variation based noise removal algorithms". In: *Physical Review D: nonlinear phenomena* 60.1-4, pp. 259–268.

Sava, P. and B. Biondi (2004). "Wave-equation migration velocity analysis. I. Theory". In: *Geophysical Prospecting* 52.6, pp. 593–606. DOI: https://doi.org/10.1111/j.1365-2478.2004.00447.x.

Scales, J. A., M. L. Smith, and S. Treitel (1994). *Introductory geophysical inverse theory*. Samizdat Press Golden.

Schot, S. H. (1992). "Eighty years of Sommerfeld's radiation condition". In: *Historia mathematica* 19.4, pp. 385–401.

Schuster, G. (2009). *Seismic interferometry*. Cambridge University Press.

Schuster, G. T. (Jan. 2017). *Seismic Inversion*. Society of Exploration Geophysicists. ISBN: 9781560803416. DOI: 10.1190/1.9781560803423.

Sei, A. (1995). "A family of numerical schemes for the computation of elastic waves". In: *SIAM Journal on Scientific Computing* 16.4, pp. 898–916.

Sethian, J. A. (1996). "A fast marching level set method for monotonically advancing fronts". In: *Proceedings of the National Academy of Sciences* 93.4, pp. 1591–1595.

Sethian, J. A. and A. M. Popovici (1999). "3-D traveltime computation using the fast marching method". In: *Geophysics* 64.2, pp. 516–523.

Shin, C. and D.-J. Min (May 2006). "Waveform inversion using a logarithmic wavefield". In: *Geophysics* 71.3, R31–R42. DOI: 10.1190/1.2194523. eprint: https://pubs.geoscienceworld.org/geophysics/article-pdf/71/3/R31/3212710/gsgpy_71_3_R31.pdf.

Singh, S., R. Snieder, J. van der Neut, J. Thorbecke, E. Slob, and K. Wapenaar (2017). "Accounting for free-surface multiples in Marchenko imaging". In: *GEOPHYSICS* 82.1, R19–R30. DOI: 10.1190/geo2015-0646.1. eprint: https://doi.org/10.1190/geo2015-0646.1.

Snieder, R. (1990a). "A perturbative analysis of non-linear inversion". In: *Geophysical Journal International* 101.3, pp. 545–556. DOI: 10.1111/j.1365-246X.1990.tb05569.x. eprint: https://academic.oup.com/gji/article-pdf/101/3/545/1637813/101-3-545.pdf.

Snieder, R. (1990b). "The role of the Born approximation in nonlinear inversion". In: *Inverse Problems* 6.2, pp. 247–266. DOI: 10.1088/0266-5611/6/2/008.

Snieder, R. and J. Trampert (1999). "Inverse problems in geophysics". In: *Wavefield inversion*. Springer, pp. 119–190.

Snieder, R. (2004). "Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase". In: *Phys. Rev. E* 69 (4), p. 046610. DOI: 10.1103/PhysRevE.69.046610.

Snieder, R. and K. Van Wijk (2015). *A guided tour of mathematical methods for the physical sciences*. Cambridge University Press.

Sommerfeld, A. (1912). "Die Greensche funktion der schwingungsgleichung". In: *J.-Ber. Deutsch Math.-Verein* 21, pp. 309–353.

Staring, M., R. Pereira, H. Douma, J. van der Neut, and K. Wapenaar (Nov. 2018). "Source-receiver Marchenko redatuming on field data using an adaptive double-focusing method". In: *Geophysics* 83.6, S579–S590. DOI: 10.1190/geo2017-0796.1. eprint: https://pubs.geoscienceworld.org/geophysics/article-pdf/83/6/S579/4600465/geo-2017-0796.1.pdf.

Staring, M. and K. Wapenaar (2020). "Three-dimensional Marchenko internal multiple attenuation on narrow azimuth streamer data of the Santos Basin, Brazil". In: *Geophysical Prospecting* 68.6, pp. 1864–1877. DOI: https://doi.org/10.1111/1365-2478.12964.

Sun, Y. and G. T. Schuster (1993). "Time-domain phase inversion". In: *SEG Technical Program Expanded Abstracts 1993*, pp. 684–687. DOI: 10.1190/1.1822588. eprint: https://library.seg.org/doi/pdf/10.1190/1.1822588.

Symes, W. W. (2008). "Migration velocity analysis and waveform inversion". In: *Geophysical Prospecting* 56.6, pp. 765–790. DOI: https://doi.org/10.1111/j.1365-2478.2008.00698.x.

Taner, M. T. and F. Koehler (1969). "Velocity spectra—Digital computer derivation applications of velocity functions". In: *Geophysics* 34.6, pp. 859–881.

Tarantola, A. (Aug. 1984). "Inversion of seismic reflection data in the acoustic approximation". In: *Geophysics* 49.8, pp. 1259–1266. DOI: 10.1190/1.1441754. eprint: https://pubs.geoscienceworld.org/geophysics/article-pdf/49/8/1259/3163308/1259.pdf.

Thorbecke, J., E. Slob, J. Brackenhoff, J. van der Neut, and K. Wapenaar (2017). "Implementation of the Marchenko method". In: *Geophysics* 82.6, WB29–WB45.

Thorbecke, J. W. and D. Draganov (2011). "Finite-difference modeling experiments for seismic interferometry". In: *Geophysics* 76.6, H1–H18.

Tikhonov, A. and V. Arsenin (1977). *Solutions of Ill-posed Problems*. Halsted Press book. Winston. ISBN: 9780470991244.

Treitel, S., P. R. Gutowski, and D. E. Wagner (Oct. 1982). "Plane-wave decomposition of seismograms". In: *Geophysics* 47.10, pp. 1375–1401. DOI: 10.1190/1.1441287. eprint: https://pubs.geoscienceworld.org/geophysics/article-pdf/47/10/1375/3115193/1375.pdf.

van den Berg, P. and A. Abubakar (2001). "Contrast source inversion method: State of art". In: *Progress in Electromagnetics research* 34, pp. 189–218.

van der Neut, J., K. Wapenaar, J. Thorbecke, and E. Slob (2015). "Practical challenges in adaptive Marchenko imaging". In: *SEG Technical Program Expanded Abstracts 2015*. Society of Exploration Geophysicists, pp. 4505–4509.

van der Neut, J. and K. Wapenaar (2016). "Adaptive overburden elimination with the multidimensional Marchenko equation". In: *Geophysics* 81.5, T265–T284.

Van Leeuwen, T. and F. J. Herrmann (2013). "Mitigating local minima in full-waveform inversion by expanding the search space". In: *Geophysical Journal International* 195.1, pp. 661–667.

Verschuur, D. J., A. J. Berkhout, and C. P. A. Wapenaar (Sept. 1992). "Adaptive surface-related multiple elimination". In: *Geophysics* 57.9, pp. 1166–1177. DOI: 10.1190/1.1443330. eprint: https://pubs.geoscienceworld.org/geophysics/article-pdf/57/9/1166/3162508/1166.pdf.

Virieux, J. and S. Operto (Dec. 2009). "An overview of full-waveform inversion in exploration geophysics". In: *Geophysics* 74.6, WCC1–WCC26. DOI: 10.1190/1.3238367. eprint: https://pubs.geoscienceworld.org/geophysics/article-pdf/74/6/WCC1/3138505/gsgpy_74_6_WCC1.pdf.

Vogel, C. R. and M. E. Oman (1996). "Iterative methods for total variation denoising". In: *SIAM Journal on Scientific Computing* 17.1, pp. 227–238.

Vogel, C. R. (2002). *Computational methods for inverse problems*. SIAM.

Wapenaar, K., F. Broggini, and R. Snieder (2012). "Creating a virtual source inside a medium from reflection data: heuristic derivation and stationary-phase analysis". In: *Geophysical Journal International* 190.2, pp. 1020–1024.

Wapenaar, K., F. Broggini, E. Slob, and R. Snieder (Feb. 2013). "Three-Dimensional Single-Sided Marchenko Inverse Scattering, Data-Driven Focusing, Green's Function Retrieval, and their Mutual Relations". In: *Phys. Rev. Lett.* 110 (8), p. 084301. DOI: 10.1103/PhysRevLett.110.084301.

Wapenaar, K., J. Thorbecke, J. van der Neut, F. Broggini, E. Slob, and R. Snieder (2014a). "Green's function retrieval from reflection data, in absence of a receiver at the virtual source position". In: *The Journal of the Acoustical Society of America* 135.5, pp. 2847–2861.

Wapenaar, K., J. Thorbecke, J. van der Neut, F. Broggini, E. Slob, and R. Snieder (2014b). "Marchenko imaging". In: *Geophysics* 79.3, WA39–WA57.

Wapenaar, K. (Dec. 2014). "Single-sided Marchenko focusing of compressional and shear waves". In: *Phys. Rev. E* 90 (6), p. 063202. DOI: 10.1103/PhysRevE.90.063202.

Warner, M. and L. Guasch (Sept. 2016). "Adaptive waveform inversion: Theory". In: *Geophysics* 81.6, R429–R445. DOI: 10.1190/geo2015-0387.1. eprint: https://pubs.geoscienceworld.org/geophysics/article-pdf/81/6/R429/3683674/geo2015-0387.pdf.

Weglein, A. B., F. A. Gasparotto, P. M. Carvalho, and R. H. Stolt (1997). "An inverse-scattering series method for attenuating multiples in seismic reflection data". In: *Geophysics* 62.6, pp. 1975–1989.

Woodward, M. J. (Jan. 1992). "Wave-equation tomography". In: *Geophysics* 57.1, pp. 15–26. DOI: 10.1190/1.1443179. eprint: https://pubs.geoscienceworld.org/geophysics/article-pdf/57/1/15/3161351/15.pdf.

Yilmaz, Ö. (Jan. 2001). *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data*. Society of Exploration Geophysicists. ISBN: 9781560800941. DOI: 10.1190/1.9781560801580.

Appendix A

Addendum to Chapter 4

In response to comments from the examiners in the viva with respect to chapter 4 of this thesis, additional work was carried out to answer those questions. This additional work is presented here. Specifically, it relates to two questions:

1. Is there any change in the inversion result when the wavelet is removed from the data by deconvolution in place of using flat spectrum Green's functions?

2. The solution presented by Beydoun and Tarantola (1988) for the scattering potential in the frequency domain cannot make use of the 0Hz component of the Gabor wavelet. By solving the problem in the time domain, this component can be used in the inversion. Does this change the solution obtained?

A.1 The effect of source function deconvolution on the inverse problem solution

In section 4.3, I explain the decision to use flat spectrum Green's functions in place of using deconvolution to remove the source function from the data and kernel matrix. In the viva, the examiners noted this and felt it was necessary to demonstrate the effect this decision made on the final inversion model. The purpose of this section is to address this.

A.1.1 Theory of deconvolution

Deconvolution is the inverse process to convolution. The convolution of two time series is equal to the multiplication of their frequency domain representations, thus the deconvolution of one time series with respect to another is equal to the division of one frequency domain representation by the other. For the time series $A(t)$, $B(t)$, and $C(t)$; and their frequency domain representations $\tilde{A}(\omega)$, $\tilde{B}(\omega)$, and $\tilde{C}(\omega)$; this relationship can be written as follows:

$$A(t) = B(t) * C(t) \quad \longleftrightarrow \quad \tilde{A}(\omega) = \tilde{B}(\omega)\tilde{C}(\omega) \quad \textbf{Convolution} \tag{A.1}$$

$$C(t) = A(t) \circledast B(t) \quad \longleftrightarrow \quad \tilde{C}(\omega) = \frac{\tilde{A}(\omega)}{\tilde{B}(\omega)} \quad \textbf{Deconvolution} \tag{A.2}$$

It should be noted that whilst the symbol $*$ is commonly used to denote time domain convolution, the symbol $\circledast$ is not a standard symbol for deconvolution in the time domain.

In geophysics, the convolutional model states that the observed seismogram $d(t)$ is equal to the convolution of the source time function $s(t)$ with the Earth impulse response $g(t)$ plus an error $e(t)$:

$$d(t) = s(t) * g(t) + e(t) \tag{A.3}$$

The importance of this equation is that if the observed seismogram $d(t)$ is a recording of the pressure wavefield $u(t)$, then the Earth impulse response $g(t)$ is equal to the pressure Green's function $G(t)$ between the source and the observation point as defined by the acoustic wave equation. Thus the Green's functions of the medium used in chapter 4 are related to the observed pressure wavefield by deconvolution of the source time function $s(t)$ from the pressure wavefield $u(t)$. In the frequency domain, this is written

$$\tilde{G}(\omega) = \frac{\tilde{u}(\omega)}{\tilde{s}(\omega)} \tag{A.4}$$

If the error in equation A.3 is equal to zero, then the result of equation A.4 is exactly equal to the flat spectrum Green's functions used in chapter 4. In practice, the

error function is non-zero and the inverse problem solving equation A.4 is ill-posed. One reason for this is the presence of noise in real-world recordings of the pressure wavefield, another is the zero-division problem referenced in section 3.1.3. This occurs when the amplitude of the denominator in equation A.2 at frequencies ω_0 is equal to zero, $|\tilde{B}(\omega_0)| = 0$, since division by zero is undefined; for the problem at hand, this means that the source function in equation A.4 has frequencies with zero amplitude.

One solution to these problems is the so-called "water-level method" (e.g. Gubbins (2004)), whereby a small positive constant ϵ is added to the denominator of equation A.2 so that the zeros in the amplitude spectrum are made finite. This is also called "spectral whitening" since the characteristic of white noise is that it has constant amplitude over all frequencies. Difficulties arise if the denominator switches between positive and negative values at different frequencies, since addition of a positive constant moves the positive values of the denominator away from zero and the negative values of the denominator towards zero.

Another solution is to rationalise the denominator of equation A.2 by multiplication to the top and bottom of the equation by a factor of $\tilde{B}^*(\omega)$. The result of this is

$$\tilde{C}(\omega) = \frac{\tilde{A}(\omega)}{\tilde{B}(\omega)} = \frac{\tilde{A}(\omega)\tilde{B}^*(\omega)}{\tilde{B}(\omega)\tilde{B}^*(\omega)} \tag{A.5}$$

Since for a complex number z, $zz^* = |z|^2$, the denominator of equation A.5 is equal to the power spectrum of $\tilde{B}(\omega)$ which always has positive amplitude. This alone cannot solve the zero division problem since the square of zero amplitude is still zero, but when combined with the water-level method one can guarantee that the denominator of the deconvolution is positively valued and non-zero.

A.1.2 Results

To determine the effect of deconvolution on the inversion result as compared to using flat spectrum Green's functions, I compare the data as modelled by equation 4.2 using flat spectrum Green's functions to the data found by deconvolution of the Gabor wavelet from the data that has been previously convolved with said wavelet.

Figure A1 shows this comparison between data modelled with flat spectrum

Green's functions and data found by deconvolution of the Gabor wavelet as a minimum-phase wavelet and after zero-phasing with 1001 time samples. All three of these traces are found to be exactly equal to one another. When the number of time samples is increased to 64001, shown figure A2, although the spikes in the data traces are narrower than in figure A1 the result is the same as before. Since the data traces are found to be exactly equal, assuming the same pattern also holds true for the wavefield at all other points means that the kernel matrices will be exactly equal, and thus the inversion results also.

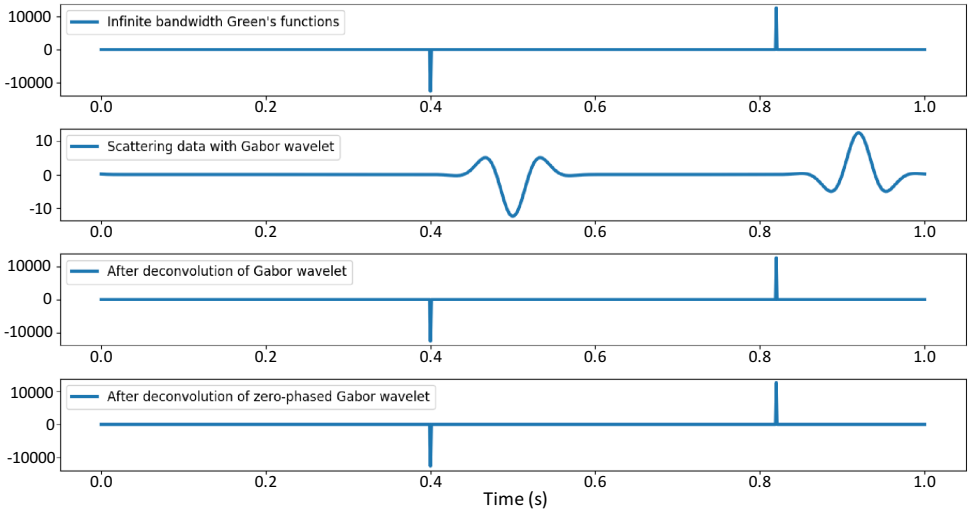


Figure A.1: Comparison of data traces modelled using 1001 time samples when (top) flat spectrum Green's functions are used, (second top) flat spectrum data is convolved with the Gabor wavelet, (second bottom) the convolved wavelet is removed by deconvolution, and (bottom) the convolved wavelet is removed by deconvolution with the Gabor wavelet after zero-phasing.

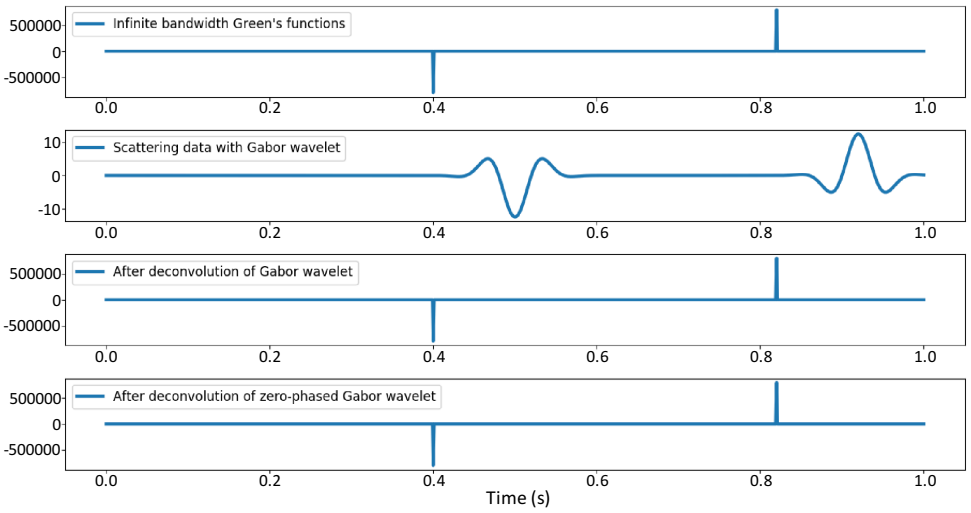


Figure A.2: Comparison of data traces modelled using 64001 time samples, shown truncated after 1001 samples. Subfigures are arranged as per figure A.1.

A.2 Inversion in the time domain and the effect of the 0Hz data component

In chapter 4 on page 71 I give the equations used by Beydoun and Tarantola (1988) to calculate the Born scattered wavefield and the scattering potential of a boxcar perturbation in the frequency domain. Unfortunately, the scattering potential in this analytic example is inversely proportional to frequency, and so the 0Hz component of the data cannot be used. Ordinarily this would be a minor detail, since many popular wavelets used in geophysics such as the Ricker wavelet have zero amplitude at 0Hz. However, the Gabor wavelet chosen by Beydoun and Tarantola (1988) has a significant amplitude at 0Hz which cannot be exploited by inversion in the frequency domain. As highlighted by one of the examiners, an inversion formula in the time domain is also specified which does not have this limitation, and so it should be possible to utilise the 0Hz frequency to improve the inversion and potentially eliminate the DC perturbation seen in the results attempting to recreate their experiment.

The time domain inversion formula is

$$\Delta u_B(z) = -4n_0^2 \int_0^{2(z-z_g)/v_0} \Delta P(z_g, t)\, dt$$

(Eq. A18 (Beydoun and Tarantola, 1988)) (A.6)

Here, $\Delta u_B(z)$ is the scattering potential under the Born approximation as a function of depth z, n_0^2 is the slowness-squared of the background medium, z_g is the depth at which the geophone is located, and $\Delta P(z_g, t)$ is the scattered wavefield observed at the geophone as a function of time t. The formula for the scattered wavefield as a function of frequency $\Delta P(z_g, \omega)$ is given in equation 4.2.

Using equation A.6, I now invert for the scattering potential in the time domain and including the 0Hz component of the data. For comparison, the inversion result obtained by Beydoun and Tarantola (1988) is demonstrated in figure 4.4a. Figure A.3 shows the results of time domain inversion using equation A.6 with flat spectrum Green's functions and a total of 1001 time samples. It shows that the locations of the first and second interfaces are both located exactly, but that the amplitude of the perturbation is incorrect. The amplitude of the perturbation can be correctly recovered at the first interface but not the second. Although I have used a model update whose amplitude has been normalised to the amplitude of the true model perturbation, with a suitable optimisation method it should be possible to obtain the same results (albeit by taking more time and at greater computational expense than normalising the amplitude). Interestingly, if one ignores the integration limits in equation A.6 and sums over all available time samples, one recovers the model perturbation exactly.

Comparing now to when the data is convolved with the Gabor wavelet, figure A.4 shows the subsequent velocity models post-inversion when the wavelet was left in the data and when the wavelet was deconvolved from the data. When the wavelet is not removed from the data, the first and second interfaces can again be identified but the wavelet is also imprinted on the inversion solution. The second interface can be identified from the maximum amplitude point of the wavelet, whereas the first interface is identifiable at the vertex between zero update and the wavelet imposed

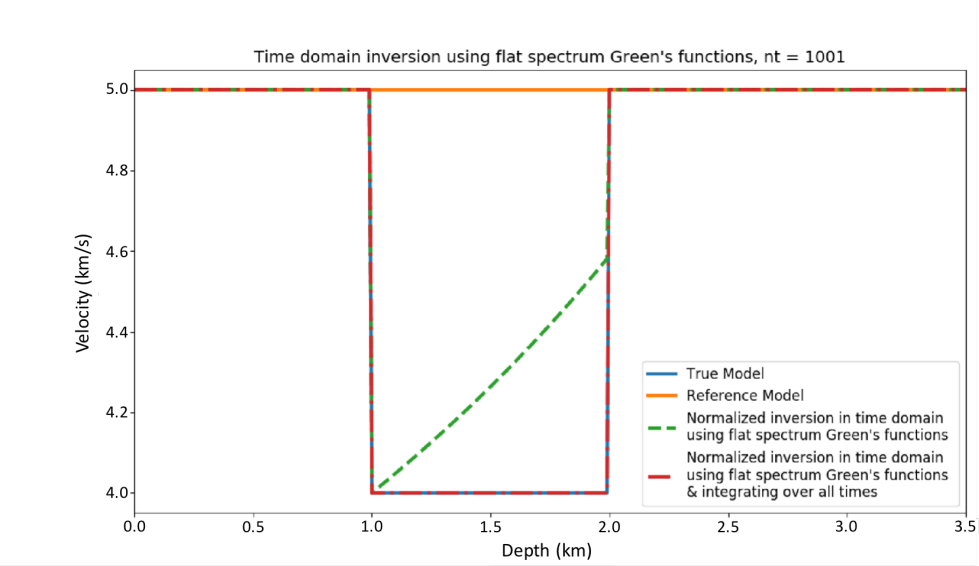


Figure A.3: 1D Born inversion of a boxcar perturbation in the time domain using the analytic method of Beydoun and Tarantola (1988) with flat spectrum Green's functions.

on the solution. The maximum amplitude point of the wavelet around the first interface does not occur at the location of the first interface. When the wavelet is deconvolved from the data before inversion, either with or without zero-phasing, the solution matches the flat spectrum Green's function solution. This agrees with the claim made in A.1.2 that the agreement of the flat spectrum Green's function data and the deconvolved data would result in matching inversion models.

Given the exact inversion result seen when flat spectrum Green's functions are inverted by replacing the integration limits in equation A.6 by integration over all time, I decided to determine if this was a property of using flat spectrum Green's functions or a fundamental property of the analytic inversion method of Beydoun and Tarantola (1988). To this end, figure A.5 shows the results of integrating over all time when the Gabor wavelet is present in the data and when that wavelet is deconvolved from the data. It is seen that the inversion results for both cases are again an exact match to the true velocity model.

To determine whether time redundancy also improves the inversion results in the time domain, figure A.6 shows the inversion results when the number of time samples is increased from 1001 to 64001. From figure A.6a, increasing the number of time samples is shown to improve the accuracy of the inversion amplitude when

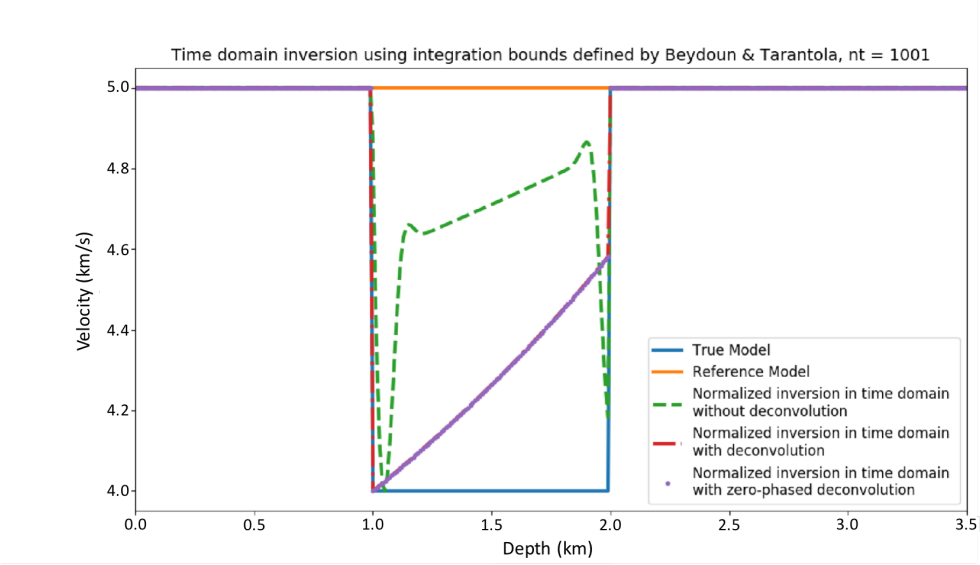


Figure A.4: 1D Born inversion of a boxcar perturbation in the time domain using the analytic method of Beydoun and Tarantola (1988) with data convolved with the Gabor wavelet.

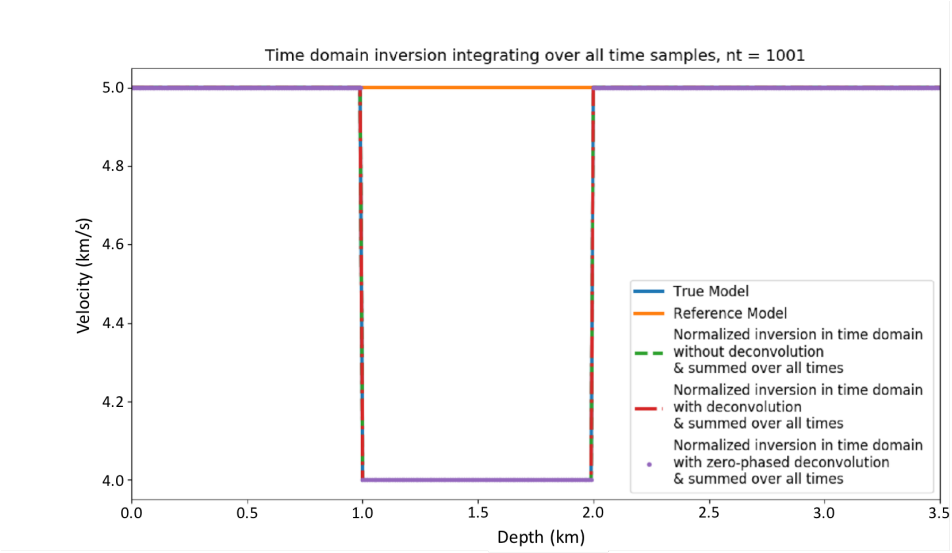


Figure A.5: 1D Born inversion of a boxcar perturbation in the time domain by replacing the integration limits of the analytic method of Beydoun and Tarantola (1988) with a sum over all time samples.

flat spectrum Green's functions are used, such that the inversion is almost exact when the integration bounds in equation A.6 are used. When the integration bounds are ignored and replaced by an integral over all time, the inversion is again exact. Figure A.6b shows that when the Gabor wavelet is not deconvolved from the data, the amplitude difference at the first and second interfaces has been almost entirely eliminated. When the wavelet has been deconvolved from the data, the inversion solution again matches the flat spectrum Green's function solution. If the inversion limits are replaced by an integral over all time, the inversion is exact even if the wavelet has not been removed from the data, as seen in figure A.6c.

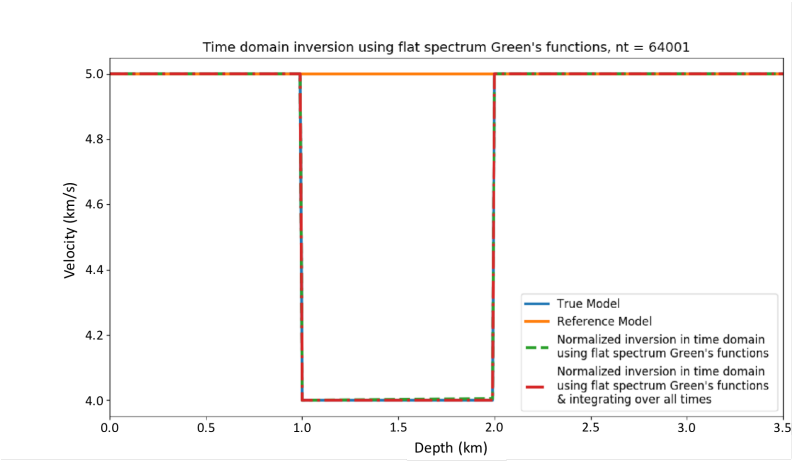
Two questions immediately come to mind when looking at these results:

1. *Why does the inversion solution become exact (with proper scaling) when the integration limits in equation A.6 are replaced with an integral over all time?* And
2. The solution found by Beydoun and Tarantola (1988) (figure 4.4a) appears to have the correct amplitude above and below the perturbation, implying that the 0Hz component of the data has been utilised to set the DC component of the model update. This is impossible with the frequency domain inversion formula of equation 4.1 since it is inversely proportional to the frequency. But the results so far in the time domain have sharp features without the ringing observed in their result, which implies that inversion was performed in the frequency domain. *Can this apparent contradiction be resolved by combining the time and frequency domain solutions, inverting the 0Hz component in the time domain and the remaining frequencies in the frequency domain? And does this improve the inversion result?*

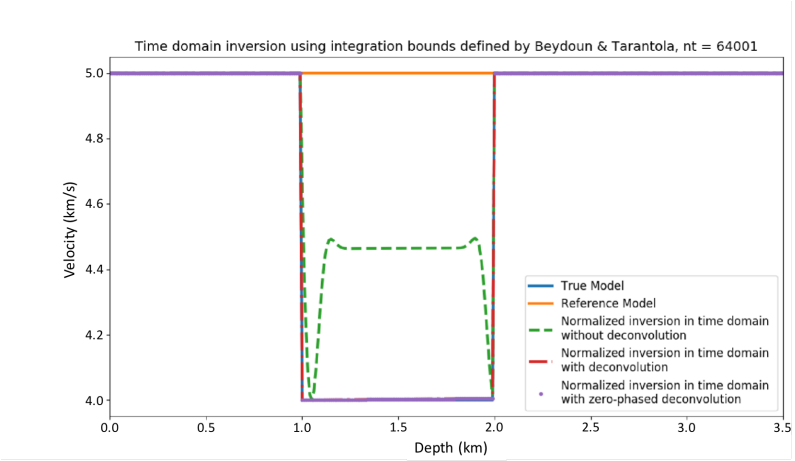
To answer the first question, I replace the integral in equation A.6 with a causal integration. Causal integration is defined as (Claerbout, 2014):

$$y(t) = \int_{-\infty}^{t} x(\tau) d\tau \tag{A.7}$$

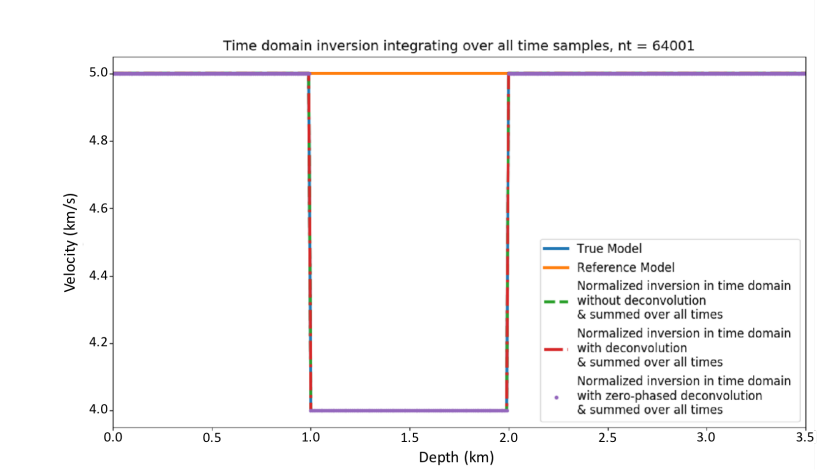
As a result, the inversion results for all integral upper limits at all depths are obtained. This allows the change in the amplitude of the inversion result with different



(a) Inversion with flat spectrum Green's functions



(b) Inversion with the Gabor wavelet intact

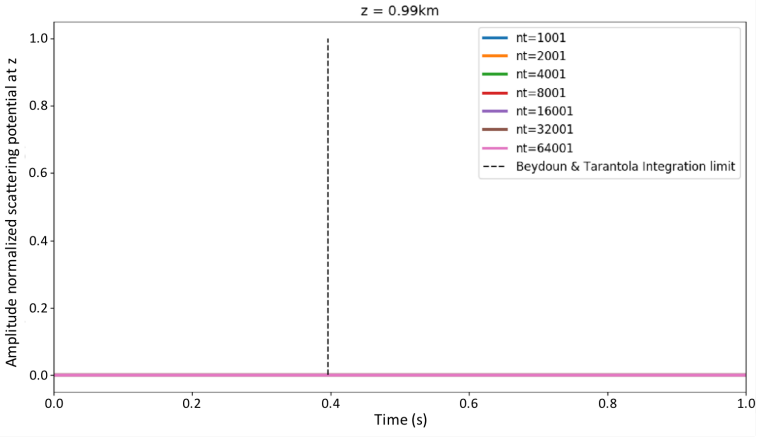


(c) Inversion with the Gabor wavelet integrating over all times

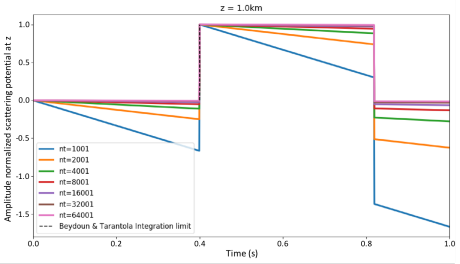
Figure A.6: 1D Born inversion of a boxcar perturbation in the time domain using the analytic method of Beydoun and Tarantola (1988) and data modelled with 64001 time samples.

integral limits to be determined at any given depth. Figure A.7 shows this for depths above, within, and below the velocity perturbation, using data modelled with flat spectrum Green's functions for different numbers of time samples modelled. It is seen from figures A.7b - A.7f that the inversion result is depth invariant at depths where the true model perturbation is non-zero, whilst above and below the perturbation (figures A.7a and A.7g) the inversion result is zero for all integration limits. Looking now at the effect of increasing the number of time samples modelled, it is clear that the solution at a given depth is slanted along the integration time axis, and that increasing the number of time samples modelled reduces the gradient of this slant. This explains why the amplitude of the perturbation becomes more balanced between the interfaces when the number of time samples increases. Since the amplitude of the inversion result only varies with the upper integration limit and not with depth (within the perturbation), because the integration limit for the first interface is lower than for the second interface, increasing the number of time samples decreases the slant of the solution in time and reducing the change in amplitude between the integration limits at the two interfaces. It also explains why the inversion converges on the exact solution when one integrates over all time - because the inversion amplitude is depth invariant within the perturbation and zero outside the perturbation, setting the integration limit equal at all points forces the amplitude to be equal everywhere within the perturbation and zero outside this range. In fact, for a given number of time samples modelled, the total number of time samples is not a special number. Setting the upper integration limit to any number except zero will give the same result.

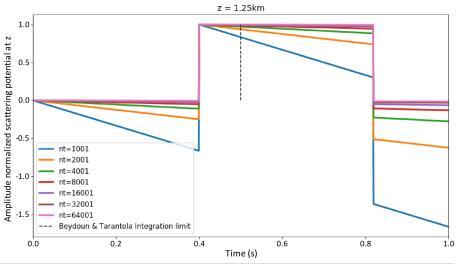
Repeating the experiment for data where the wavelet has not been removed by deconvolution, figure A.8 shows that the inversion result is also depth invariant within the perturbation when the wavelet is not removed from the data. It also explains why the identifiable wavelet features at the interfaces of the perturbation in figure A.4 are different at the two interfaces. Looking at figure A.8b, at the first interface ($z = 1$km) the black dashed line demarcating the upper integration limit defined by Beydoun and Tarantola (1988) coincides with the anti-symmetry point of the waveform, whereas in figure A.8f at the second interface ($z = 2$km), the upper integration limit coincides



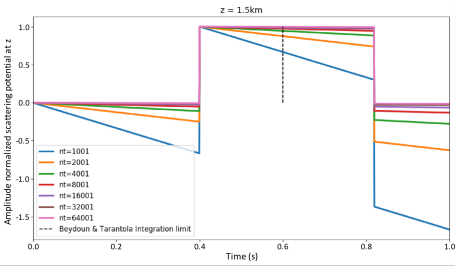
(a) Depth = 0.99km



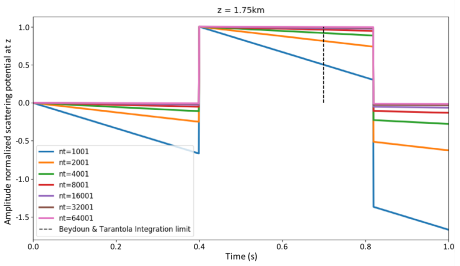
(b) Depth = 1km



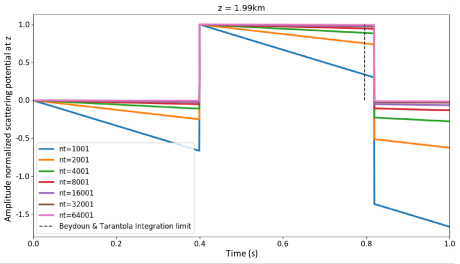
(c) Depth = 1.25km



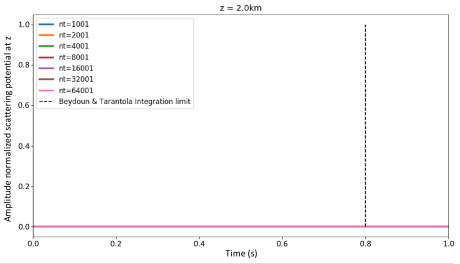
(d) Depth = 1.5km



(e) Depth = 1.75km



(f) Depth = 1.99km



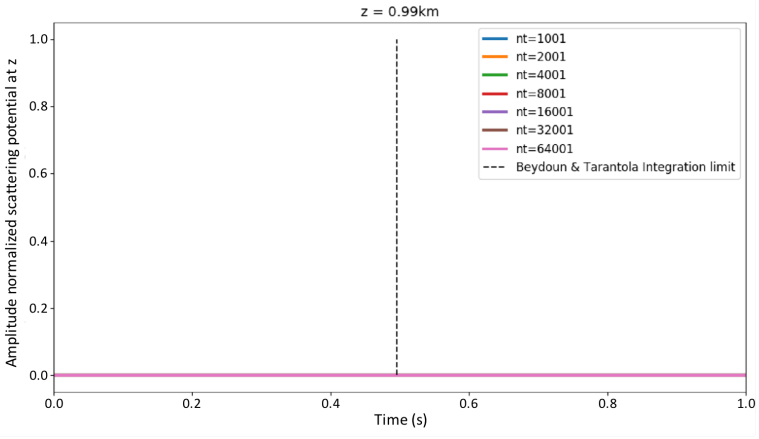
(g) Depth = 2km

Figure A.7: Comparison of the inversion amplitude with the upper integration limit in equation A.6 evaluated at various depths above, within, and below the velocity perturbation using flat spectrum Green's functions.

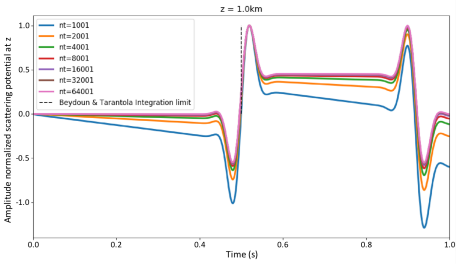
with the maximum of the waveform. This waveform is the derivative of the Gabor wavelet[1], and since this is the centre of the waveform it is actually the first interface that is exactly located not the second. If the time domain solution permitted it, the predicted depth of the second interface would be greater than the true depth.

Moving onto the second question, I begin by inverting the 0Hz component of the data in the time domain. Figure A.9 shows the inverted for scattering potential using only the 0Hz component of the data, the solution has a sharp boxcar shape with zero amplitude outside the perturbation. It is clear then that adding the 0Hz inversion solution to a frequency domain inversion cannot correct for the observed non-zero DC component outside of the perturbation in the frequency domain solution. Thus it remains unclear as to how Beydoun and Tarantola (1988) obtained their solution.

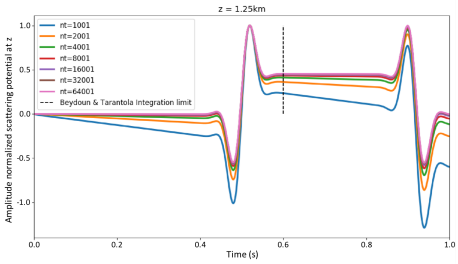
[1]Modelling in 1D transforms the source wavelet into its first time derivative.



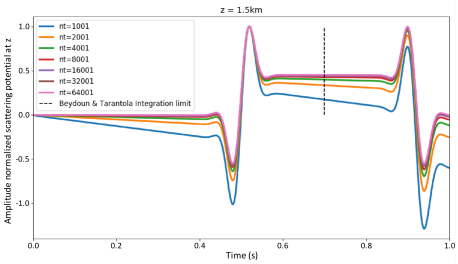
(a) Depth = 0.99km



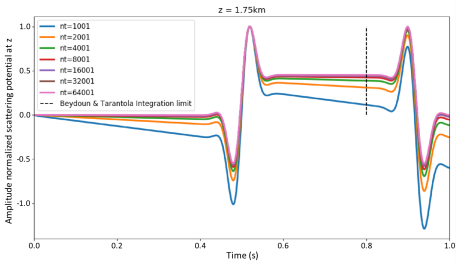
(b) Depth = 1km



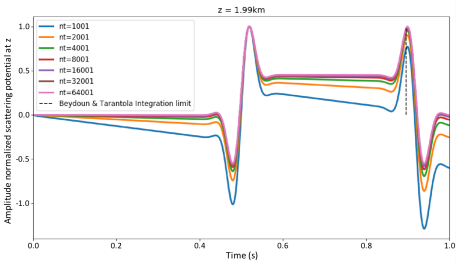
(c) Depth = 1.25km



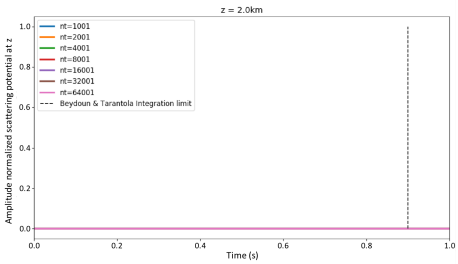
(d) Depth = 1.5km



(e) Depth = 1.75km



(f) Depth = 1.99km



(g) Depth = 2km

Figure A.8: Comparison of the inversion amplitude with the upper integration limit in equation A.6 evaluated at various depths above, within, and below the velocity perturbation using data convolved with the Gabor wavelet.

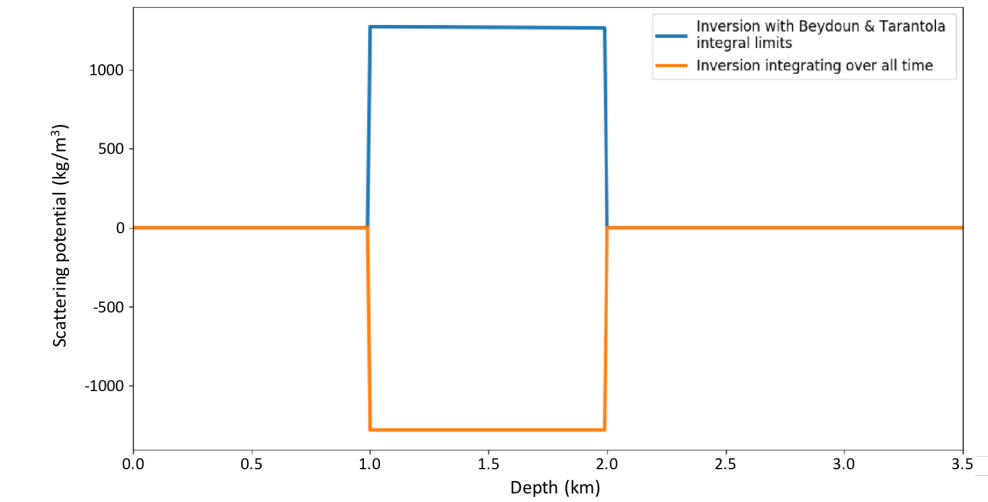


Figure A.9: 1D Born inversion of a boxcar perturbation in the time domain using only the 0Hz component of the data with the analytic method of Beydoun and Tarantola (1988).

Appendix B

Refractions, Head Waves & Diving Waves

The terminology surrounding refracted waves is often confused or used interchangeably without precision. Here then, I attempt to clarify the varied terminology.

Refraction is the bending of the wavefront due to a change in the refractive index of the medium; when the wavespeed changes in a medium, the wavelength of the wave must also change since $c = f\lambda$. Thus, when a wavefront is incident upon a change in wavespeed, the wavelength of the transmitted wavefront changes and the wavefront bends at the interface between the different wavespeeds.

Refraction is a purely geometric effect, it is governed by Snell's law:

$$\frac{\sin\theta_{in}}{c_1} = \frac{\sin\theta_{out}}{c_2} \tag{B.1}$$

When the outward angle of propagation of the wavefront is 90° from the axis perpendicular to the interface between layers, the wave is said to have been critically refracted. The angle above which any incident wave will be critically refracted is called the critical angle, by setting $\theta_{out} = 90^\circ$ in equation B.1 it is found to be

$$\theta_{critical} = \sin^{-1}\left(\frac{c_1}{c_2}\right) \tag{B.2}$$

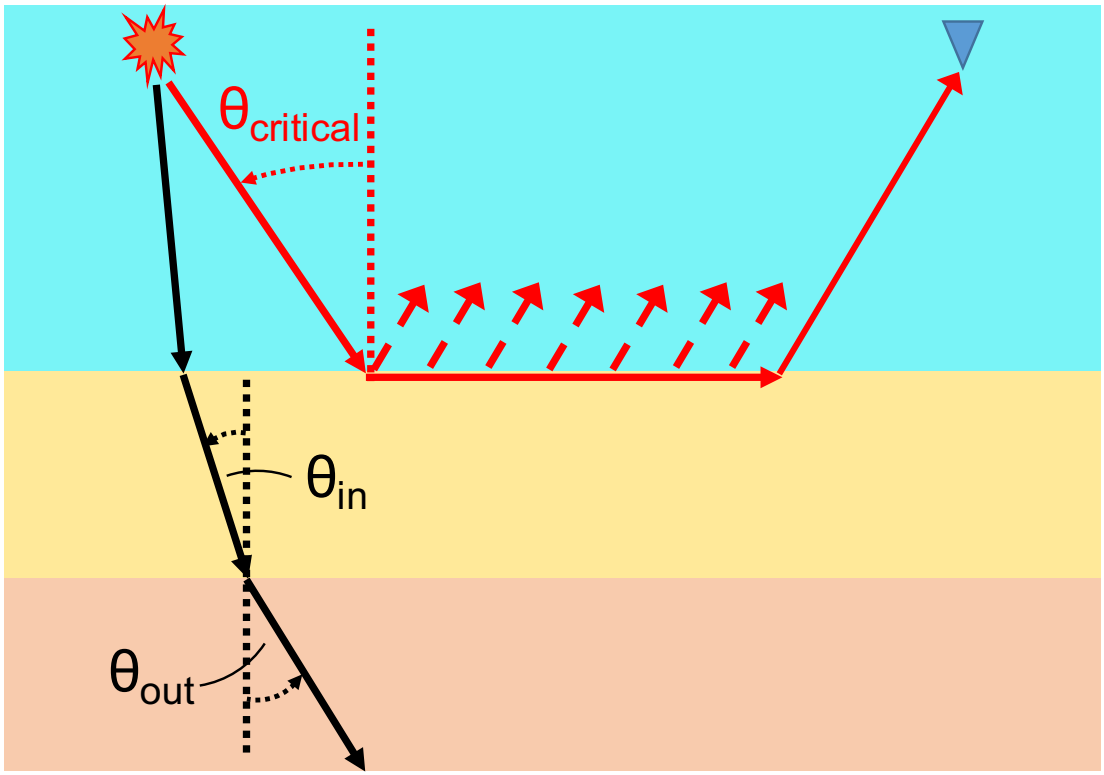


Figure B.1: A schematic of refractions in a layered medium. For incident angles below the critical angle $\theta_{critical}$ (black arrows), the propagation direction of the wave changes at the interfaces between velocity layers. For incidence angles equal to or greater than the critical angle (red arrows), the wave is critically refracted at the velocity interface and continues to propagate perpendicular to the interface in the second layer. Per Snell's law, the critically refracted wavefront acts as a source of waves propagating upwards at the critical angle (dashed red arrows). The recorded refraction is called a head wave (upwards red arrow).

Snell's law also tells us that the critically-refracted wave propagates in the second layer just below the interface between different velocity layers, and that wavefront is refracted back into the first velocity layer again according to Huygen's principle at the critical angle of incidence. When recorded at the surface of the Earth, such waves are labelled as refracted waves, or simply *refractions*. Such waves are also often called *head waves*, since at long offsets from the source they arrive ahead of other wave types.

Another type of refraction is called a *diving wave*. Diving waves occur when the velocity of the medium increases smoothly with depth. In such a model, the wavefront will be continuously refracted according to Snell's law until it is eventually refracted

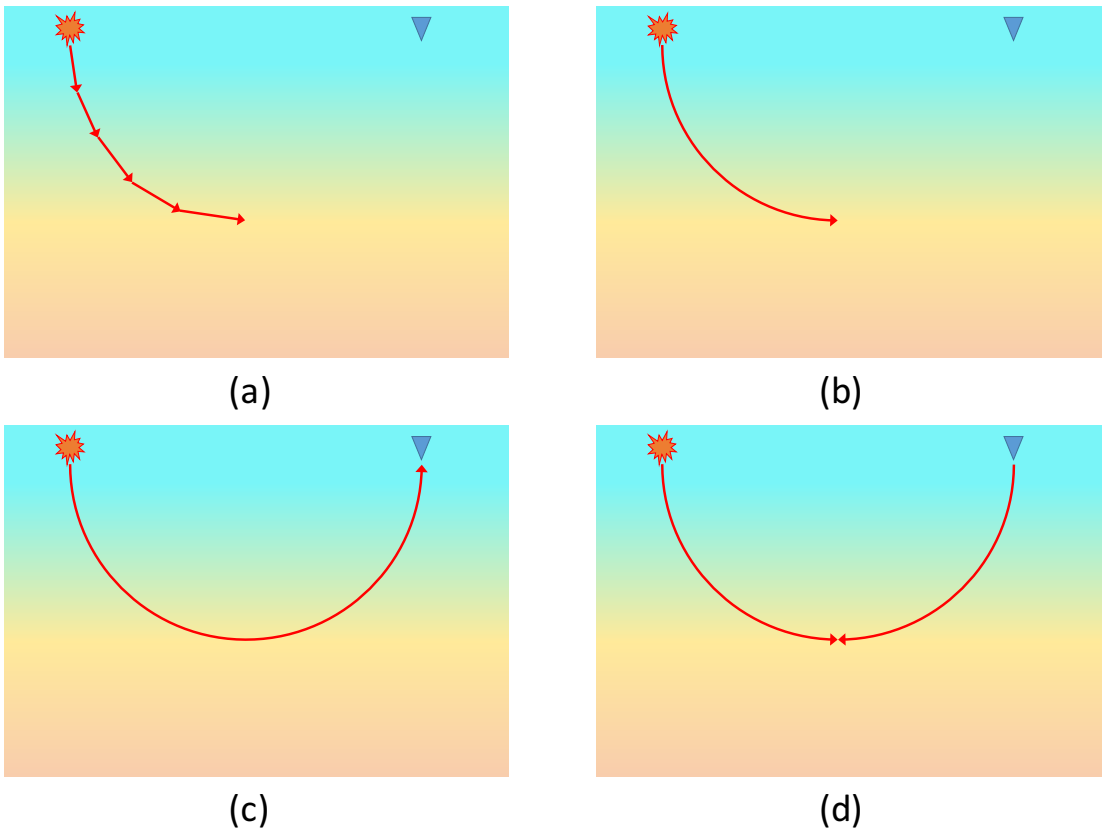


Figure B.2: A schematic of diving wave formation. In a smooth model where velocity increases linearly with depth, the wavefield is continuously refracted since the velocity changes continuously. In a finitely discretized model, it thus refracts after each unit of length propagated (a). In the limit of infinitesimal unit size, the wavefront is refracted along a smoothly curving trajectory until it reaches the criticial incidence angle (b), at which point it propagates horizontally emitting Huygen's wavelets at the critical angle upwards. Since the medium smoothly increases with depth, the upward trajectory of the wavefront curves smoothly towards the vertical (c), this is the diving wave. We can image using diving waves by backpropagating the wavefield from the receiver through the same model (d).

to or beyond the critical angle. Snell's law dictates that the wavefront will again turn upwards as before, even though there is no discernible interface in the model to which the refraction can be attributed. Diving waves are of interest to geophysicists because they cannot be attributed to any interface in the Earth, rather they are sensitive to the velocity structure between interfaces. The process by which a diving wave occurs is described in Figure B.2.

Appendix C

Analytic Green's Function Solutions to the Helmholtz Equation in Two- & Three- Dimensions

In chapter 3, it is stated that the 2D Green's function solution to the Helmholtz equation is phase rotated by $\pi/4$ compared to the 3D Green's function solution, in this appendix I briefly outline the mathematics behind this.

The Helmholtz equation for a constant density medium with wavespeed $c(\boldsymbol{x})$ is (also equation 2.3)

$$\nabla^2 G\left(\boldsymbol{x}, \boldsymbol{x}'; \omega\right) + \frac{\omega^2}{c^2\left(\boldsymbol{x}\right)} G\left(\boldsymbol{x}, \boldsymbol{x}'; \omega\right) = -\delta\left(\boldsymbol{x} - \boldsymbol{x}'\right) \tag{C.1}$$

where the Green's function $G\left(\boldsymbol{x}, \boldsymbol{x}'; \omega\right)$ is defined between points $\boldsymbol{x}'$ and $\boldsymbol{x}$ for frequency ω. The analytic Green's function solutions in two- and three-dimensions are (e.g. Berkhout, 1980; DeSanto, 1992; Snieder and Van Wijk, 2015)

$$G^{2D}\left(\boldsymbol{x}, \boldsymbol{x}'; \omega\right) = -\frac{i}{4} H_0^{(2)}(k|\boldsymbol{x} - \boldsymbol{x}_0|) \tag{C.2}$$

$$G^{3D}\left(\boldsymbol{x}, \boldsymbol{x}'; \omega\right) = \frac{1}{4\pi} \frac{e^{-ik|\boldsymbol{x} - \boldsymbol{x}_0|}}{|\boldsymbol{x} - \boldsymbol{x}_0|} \tag{C.3}$$

where $k = \omega/c$ is the wavenumber and $H_0^{(2)}$ is the zeroth Hankel function of the second kind. According to Abramowitz and Stegun (1964), the n^{th} Hankel function

of the second kind can be approximated as

$$H_n^{(2)}(z) \sim \sqrt{\frac{2}{\pi z}} \exp\left[-i\left(z - \frac{1}{2}n\pi - \frac{\pi}{4}\right)\right] \tag{C.4}$$

Taking $n = 0$ and $z \to |\boldsymbol{x} - \boldsymbol{x}_0|$ in equation C.4, one finds that

$$H_n^{(2)}(z) \sim \sqrt{\frac{2}{\pi z}} \exp\left[-i\left(z - \frac{\pi}{4}\right)\right] \tag{C.5}$$

Substituting equation C.5 into equation C.2, and using that $-i = e^{-i\pi/2}$, the Green's function solution to the Helmholtz equation in 2D is

$$G^{2D}\left(\boldsymbol{x}, \boldsymbol{x}'; \omega\right) = \sqrt{\frac{2\pi}{k|\boldsymbol{x} - \boldsymbol{x}_0|}} \exp\left[-i\left(k|\boldsymbol{x} - \boldsymbol{x}_0| + \frac{\pi}{4}\right)\right] \tag{C.6}$$

Comparing equations C.6 and C.3, one observes a $\pi/4$ phase shift between the Green's function solutions in 2D and 3D.

"So long, and thanks for all the fish"

Douglas Adams