

Imperial College London  
Department of Earth Science and Engineering

# **Full Waveform Inversion Procedures with Irregular topography**

Pei-Zhi Li

2020

Supervised by:

Prof Michael Warner

Prof Yanghua Wang

Submitted in part fulfilment of the requirements for the degree of Master of Philosophy of Imperial  
College London and the Diploma of Imperial College London

# Disclaimer

I declare that this thesis has been written and composed solely by myself, and all works that are not of my own doing have been appropriately acknowledged. I certify that this thesis have not been submitted, either wholly or partly for any previous applications for a degree.

Pei-Zhi Li

# Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

# Abstract

Full waveform inversion (FWI) is a form of seismic inversion that uses data residual, found as the misfit, between the whole waveform of field acquired and synthesized seismic data, to iteratively update a model estimate until such misfit is sufficiently reduced, indicating synthetic data is generated from a relatively accurate model. The aim of the thesis is to review FWI and provide a simplified explanation of the techniques involved to those who are not familiar with FWI.

In FWI the local minima problem causes the misfit to decrease to its nearest minimum and not the global minimum, meaning the model cannot be accurately updated. Numerous objective functions were proposed to tackle different sources of local minima. The ‘joint deconvoluted envelope and phase residual’ misfit function proposed in this thesis aims to combine features of these objective functions for a comprehensive inversion. The adjoint state method is used to generate an updated gradient for the search direction and is followed by a step-length estimation to produce a scalar value that could be applied to the search direction to reduce the misfit more efficiently.

Synthetic data are derived from forward modelling involving simulating and recording propagating waves influenced by the mediums’ properties. The ‘generalised viscoelastic wave equation in porous media’ was proposed by the author in sub-chapter 3.2.5 to consider these properties. Boundary layers and conditions are employed to mitigate artificial reflections arising from computational simulations. Linear algebra solvers are an efficient tool that produces wavefield vectors for frequency domain synthetic data.

Regions with topography require a grid generation scheme to adjust a mesh of nodes to fit into its non-quadrilateral shaped body. Computational co-ordinate terms are implemented within wave equations throughout topographic models where a single point in the model in physical domain are represented by cartesian nodes in the computational domains. This helps to generate an accurate and appropriate synthetic data, without complex modelling computations.

Advanced FWI takes a different approach to conventional FWI, where they relax upon the use of misfit function, however none of their proponents claims the former can supplant the latter but suggest that they can be implemented together to recover the true model.

# Acknowledgement

Before I continue, I want to give my utmost and sincerest appreciation to Professor Michael Warner and Professor Yanghua Wang, who have given me invaluable recommendations and academic guidance throughout my studies. Their advice and suggestions have helped light the way throughout my academic path, and for that I would like to thank them.

I would also like to show my gratefulness to the assistance expressed by my current and past colleagues inside my research group: Jinyue Liu, Yungui Xu, Jiashun Yao, Yao Gang, and Professor Shi, whom have played valuable role throughout my progression.

Throughout my time in Imperial College I have been fortunate to experience camaraderie and friendship with members of the department including but not only: Haifa Al-Salmi, Thomas Serafini De Oliveira, Jordan Schofield, and Yasir Khoja whom have made my experience enjoyable and memorable.

Last, but not least I would like to thank my parents who have supported me throughout my postgraduate research, studies, and life. Without them this would not be possible, and it is their patience, support, and encouragement that have helped me reach where I am today.

# Contents

Disclaimer .....	i
Copyright Declaration.....	i
Abstract.....	ii
Acknowledgement .....	iii
Chapter 1: Introduction.....	1
1.1 Seismic Exploration Background.....	1
1.2 Seismic Inversion.....	2
1.3 Full-Waveform Inversion Background .....	5
1.3.1. Time domain and frequency domain.....	7
1.4 Thesis structure .....	8
Chapter 2: Full Waveform Inversion .....	9
2.1 Objective functions .....	9
2.1.1 Joint deconvoluted envelope and phase residual misfit Function .....	15
2.2 Regularisation .....	15
2.3 Derivation of the Gradient .....	18
2.4 Optimisation algorithm .....	20
2.4.1 Search directions .....	20
2.4.2 Quasi-Newton method .....	23
2.4.3 Step-length .....	25
2.5 Summary .....	28
Chapter 3: Forward Modelling.....	29
3.1 Forward Modelling Introduction.....	29
3.2 Waves.....	29
3.2.1 Elastic wave equation.....	32
3.2.2 Acoustic wave equation .....	34
3.2.3 Viscoelastic wave equations .....	35
3.2.4 Isotropy and anisotropy .....	38
3.2.5 Generalised viscoelastic wave equation in porous media. ....	38
3.3 Modelling.....	42
3.3.1 Finite difference .....	42
3.3.2 Absorbing boundary layers .....	44
3.3.3 Boundary conditions .....	46
3.3.4 Free surface boundary condition.....	47
3.3.5 Source signature.....	47

3.4 Seismic data synthesis.....	48
3.5 Summary.....	52
Chapter 4: Topography .....	53
4.1 Body-fitted grid scheme.....	53
4.2 Wave equations for body-fitted grid mesh.....	56
4.3 Summation-by-parts formulation.....	58
4.4 SBP: Absorbing Boundary Layer .....	61
4.5 SBP: Other Boundary conditions, and Source .....	62
4.6 Summary .....	63
Chapter 5: Advanced FWI .....	64
5.1 Optimal Transport.....	64
5.2 Adaptive Waveform Inversion.....	69
5.3 Wavefield Reconstruction Inversion.....	72
5.4 Reflection Waveform Inversion.....	73
5.5 Summary .....	76
Conclusion .....	77
References.....	79

# List of Figures

<b>Figure 1.1</b> Flowchart of the processes involved in geophysical processing and imaging of the subsurface medium. ....	2
<b>Figure 1.2</b> Categories of seismic inversion. ....	5
<b>Figure 1.3</b> The cycle of forward modelling and seismic inversion. ....	6
<b>Figure 2.1</b> Pseudo-code to calculate descent direction via L-BFGS algorithm. ....	25
<b>Figure 2.2</b> Graph of misfit $\phi$ versus step length $\alpha$ , where the reduction of the former is relatively small compared to the change of the latter. ....	26
<b>Figure 2.3</b> Solid line indicates the resultant $\Phi$ when value of $\alpha$ is applied, dashed line indicates the basin of $\Phi$ when $\alpha$ is applied at 0.5 and 1. ....	26
<b>Figure 2.4</b> Pseudo-code of line search algorithm used to determine the step length. ....	27
<b>Figure 2.5</b> Pseudo-code for zoom function via the bisection interpolation. ....	27
<b>Figure 3.1</b> The 3D model illustrating the propagations of the body and surface waves. ....	30
<b>Figure 3.2</b> A seismic trace containing wave information. ....	31
<b>Figure 3.3</b> Travel path of propagating waves. ....	31
<b>Figure 3.4</b> Seismic data with the different waves recorded being indicated. ....	31
<b>Figure 3.5</b> Cartesian grid points of the finite-difference method. ....	43
<b>Figure 3.6</b> An example of computational model with the applied CPML region of 50 grid points width and reflection co-efficient value of $10^{-9}$ at 15Hz. ....	45
<b>Figure 3.7</b> Schematic map of the CPML displaying where the damping operator in their respective direction is being applied. ....	46
<b>Figure 3.8</b> Ricker wavelet in the frequency domain. ....	48
<b>Figure 3.9</b> Forward model wavefields for 5Hz to 30Hz incrementing in 5Hz interval from top to bottom. Wavefields (left) without absorbing boundary layer, (right) with the absorbing boundary layer applied. ....	50
<b>Figure 3.10</b> A schematic of wavefield data extraction from 5Hz to 30Hz linearly increase from top-right to bottom-left. Receiver depth position is highlighted in red. ....	51
<b>Figure 3.11</b> Synthetic seismic data produced (left) via the time domain time-marching method (right) via inverse Fourier transform of the complex frequency domain wavefield data. ....	51
<b>Figure 4.1</b> Mesh of grids for irregular topography (top) before body-fitted grid scheme applied (bottom) after applying the body-fitted grid scheme. ....	56

# Chapter 1: Introduction

## 1.1 Seismic Exploration Background

Extensive research on the seismic exploration techniques by the oil and gas industry have been ongoing for a significant time since near the early 20<sup>th</sup> century to produce accurate structural models of the subsurface (Green, 1979). As time passes conventional hydrocarbon reservoirs are becoming increasingly depleted, hence new reservoirs are needed to be explored to meet energy demands of an increasing global population. Unfortunately, reservoirs in regions easier for hydrocarbon extraction are getting rarer, where new survey may contain complex subsurface features. Exploration geophysics can be applied in various methods to image the subsurface structure, such as electromagnetic, geothermal, gravitational, and seismic reflection method. Oil and gas industries typically apply the seismic reflection method due to their effectiveness and economical costs.

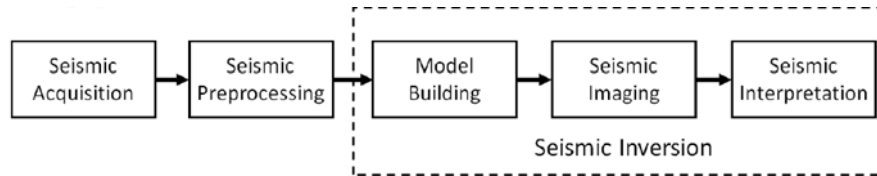
The seismic reflection method produces subsurface model estimates containing information of its properties, using the seismic waves returning to the receivers. This method is implemented over in the following steps of 1) seismic acquisition, 2) seismic pre-processing, 3) model building, 4) seismic imaging, and finally 5) seismic interpretation, where the latter three can be classed together as seismic inversion, as outlined in Figure 1.1.

The first action of the initial seismic acquisition stage is to conduct geometry planning to determine area of survey. The next action is to use a source to project waves throughout the subsurface for the waves to be recorded with receivers. For land seismic acquisition, the tools used are the vibroseis and geophones for the source and receivers respectively, whereas for marine seismic acquisition airguns and hydrophone respectively are used instead. In this process the recorded waves are typically the two body waves which are compressional P- and shear S-waves, where they reflect and refract throughout the subsurface structure to be recorded by the receivers based on their behaviour, however surface waves such as love waves and Rayleigh waves does appear though they are not as reliable as body waves and is outside the scope of this thesis. The final action in this stage involves converting the recorded data to a digital format to be used for seismic processing.

During seismic processing, the field acquired data is treated to produce meaningful image of the recorded signals, techniques involved includes muting direct arrivals (waves travelling directly from source to receiver), removal of noise usually via filtering, signal enhancement, velocity analysis, stacking, and migration. The resulting migrated seismic profile should detail accurate location of purely seismic events in the subsurface with respect to recorded times.



Techniques like velocity analysis and migration can be applied in a loop with seismic inversion, where velocity models reconstructed from model building and seismic imaging can be used as a reference model for velocity analysis, and migration, which in turn could build a more accurate seismic profile. Wave equation migration can even be applied to perform velocity analysis (Shen et al. 2003).



**Figure 1.1** Processing flowchart for a conventional seismic exploration project.

## 1.2 Seismic Inversion

Seismic inversion (SI) is a geophysical process that utilises processed seismic data to retrieve model estimates of quantitative physical properties that characterises subsurface structures in particular the reservoir. SI is performed in three stages of model building, seismic imaging, and seismic interpretation.

In the model building stage, the seismic profile is used to generate an initial model of velocity for seismic imaging, and there are two main approaches, reflection, and transmission (Wang et al, 2014). The reflection approach separates the velocity model into two scales, a long wavelength background velocity and short-wavelength reflectivity, which coupled together causes non-linearity that requires iterative optimisation method. Within the reflection approach there is the reflection travel-time tomography (Bishop et al, 1985) and migration-based travel-time tomography (Clement et al, 2001). Transmission approach are more linear with respect to the background velocity compared to the reflection approach (Trinks et al, 2005). Transmission approach can also be categorised into the ray-based travel-time tomography (RTT) and wave equation-based (WE) model building. RTT is a model building technique that utilises the relationship between travel-time and wave-speed via eikonal raytracing to generate a smooth low-frequency velocity model (Bording et al, 1987). WE model building are considered better than RTT because the latter uses ray theory that cannot correctly account for complex wave propagation phenomena, whereas the former takes it into account through full wave-equation modelling (Wang et al, 2014). Among WE model building techniques are the wave equation migration velocity analysis (WEMVA) and wave equation tomography (WET), (Stanford Exploration Project, 2002). WEMVA was originally developed to correct moveout inaccuracy of specular reflections to improve quality of migrated images, and it correlates perturbation in slowness corresponds to image perturbation (Sava and Biondi, 1999, 2004a,b). Recent advances in WEMVA involves extending to the angle domain (Zhang et al, 2012), and common-offset domain (Nandi and Albertin,

2020). Shen et al. (2003) uses wave-equation migration to measure differential semblance of the velocity. WET aims to minimise cross-correlation of travel-time differences between observed and synthetic data, in this regard WET overlaps between model building and seismic imaging where in the latter it would be categorised under inversion domain.

Thus, WET operates in the data space and WEMVA operates in the image space, the former was found to complement with FWI in producing a robust multiscale inversion strategy (Wang et al, 2014). Computational cost of WEMVA can be significantly reduced when its tomographic operator is approximated with the use of reverse time migration (RTM) (Nandi, 2018).

Seismic imaging can be categorised into two types of tomography based on the output in the process, reflectivity tomography, and model tomography. Reflectivity tomography aims to improve the subsurface reflectivity model, that is to improve resolution of geometrically located seismic events. Among the pure imaging methods are RTM and least squares migration (LSM), the former involves stacking the correlation between the forward source wavefield and the backpropagated receiver wavefield. LSM, initially introduced for cross-well data (Schuster, 1993), like inversion it tries to solve an optimisation problem with observed and synthetic data but differs in that it involves the born modelling of wave equation to generate synthetic data and updates the reflectivity model and could involve multiple iterations of migration and de-migration. Whilst LSM can produce higher quality images than RTM it is more computationally expensive (Nemeth et al., 1999, Huang et al., 2015).

Model tomography was born out of reflectivity tomography when the migration imaging principle of Claerbout (1971,196) was reconsidered as a local optimisation problem, where the misfit between the observed and synthetic data is iteratively reduced via the use of seismic reflection data and forward modelling involving acoustic and elastic wave equation (Lailly, 1983, Tarantola, 1984, 1986). The role of model tomography is to produce a high-resolution subsurface model of physical properties sampled by the propagating wave such as P-, and S- wave velocity. Developments of model tomography aims to expand it to include of rock properties (Yang and Malcolm, 2019, Li and Gu, 2019, Warner et al, 2013).

There are two types of model tomography based upon the parts of the seismic data used in the inversion, travel-time based inversion (TI), and full waveform inversion (FWI). Travel-time based inversion attempts to recover a subsurface model that would a generate synthetic data whose seismic event arrival times matches arrivals of the observed data. FWI aims to recover a subsurface model whose entire seismic signature matches the entire observed data. Both TI and FWI can be conducted in both frequency domain and time domain. TI in frequency domain would be known as phase inversion (where travel-times are Fourier transformed into phase) and uses complex angular frequency, where the real value represents wavelength, and the imaginary value represents attenuation. Phase inversion was found to be less effected by the starting model than travel-time based inversion, but phase inversion was found

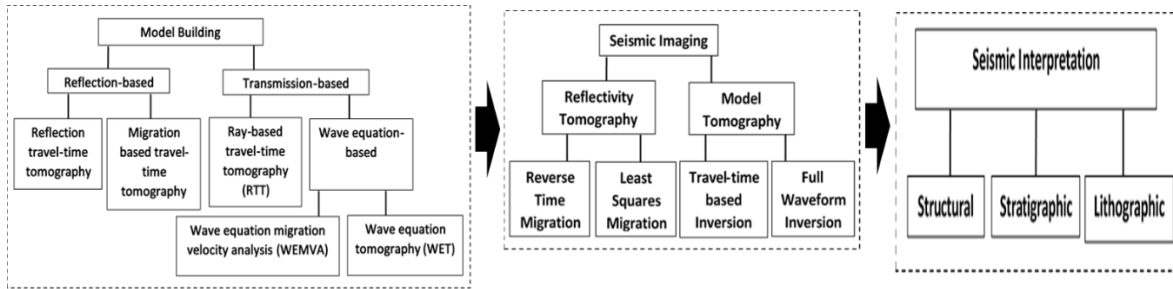
to perform poorly if the real part of the complex angular frequency was found to be outside the range determined by starting model travel-time residuals (Ellefsen, 2009). Comparing phase inversion to FWI, the former recovers a more accurate model when starting model is inaccurate, however the reduced influence of the waveform amplitude in the inversion limits its resolution capabilities (Fu et al, 2018) and inversion based purely on travel-time i.e. Full travel-time inversion (FTI), produced poorer inversion results than FWI, but further research in this area is recommended as this was introduced recently (Luo et al, 2016). FWI is a high-resolution imaging technique, and the following sub-chapter explores FWI in further details.

The goal of seismic imaging is to extract a subsurface wave model of reflectivity, absorption ( $Q$ -factor – briefly explained in sub-chapter 3.2.3) and wave velocity from the seismic data. Seismic imaging therefore does not directly tell us what is within the subsurface but merely its reaction to the wave propagating through it. The following step of seismic interpretation is to determine the physical rock-properties and structures of the sub-surface media to assist in seismic exploration and drilling for hydrocarbons. Within the seismic interpretation process there are three interrelated categories: structural, stratigraphic, and lithologic (AAPG, 2019). Structural interpretation is used to generate a structural map of the subsurface media that can include and help identify structural features such as faults, syncline, anticline. Stratigraphic interpretation is utilised to help seismic interpreters to understand the layering of sedimentary rocks based on the assumption that individual reflectors of seismic waves can be considered timelines of the depositional environment, which can be used as predictive tools for identifying hydrocarbon reservoirs, seal, and source rocks (SEG, 2018). Finally, lithologic interpretation is used to determine rocks physical properties such as elastic properties, porosity, and anisotropy, where fluid content can be determined that allows reservoir characterisation.

The scale of the work involved in seismic interpretation is vast and outside the scope of the thesis, instead a brief description of a powerful tool used in seismic interpretation known as amplitude variation with offset (AVO) analysis is mentioned here. AVO analysis is mainly utilised for lithologic interpretation but can also be applied to stratigraphic interpretation, it is based on the premise that amplitudes of a reflecting wave from a specific point in the subsurface media varies with the distance between source and receiver known as offset. This offset-based variation in response of the reflector point to the propagating waves can help seismic interpreters in predicting the type of rock and its contents. A key parameter of this response is the reflection co-efficient  $Z$ , which describes the amount of wave energy reflected due to waves travelling from one medium to another i.e. impedance at the boundary. There are three types of impedances, acoustic impedance (AI) product of density  $\rho$ , with the P-wave velocity  $V_p$ , elastic impedance (EI) product of  $\rho$  and S-wave velocity  $V_s$ , and finally ray impedance (RI) which defines the elastic impedance down a ray path using a constant ray parameter, where the wave velocity is obtained from seismic imaging.  $Z$  is then calculated with the difference between the impedance of the two mediums  $I_l$  and  $I_o$ , with the sum of the two impedances.

$$Z = \frac{I_1 - I_0}{I_1 + I_0}, \quad (1.1)$$

$Z$  can be related to the waves' angle of incidence in the Zoeppritz equation, which describes the partitioning the wave energy at the boundary between two mediums i.e. interface (Zoeppritz, 1919). However, the Zoeppritz equation does not specify on how reflection amplitudes vary with the rock properties involved and is largely angle dependent, for this the Shuey equation was developed to relate the angle dependence of reflection co-efficient to the elastic property, Poisson's ratio (Shuey, 1985). The reflection co-efficient at zero-offset  $Z_0$ , alongside AVO gradient (calculated in the Shuey equation), can be used as the axis within the AVO cross-plot to plot amplitudes of a trace with a line of best fit that is calculated to describe how the reflection amplitudes vary with offset. The expected trend is that amplitudes decrease with increasing offset indicated as a linear negative slope, however AVO anomalies are of interest when performing AVO analysis, these are often shown as increasing AVO or slower fall in  $Z_0$  than AVO gradient. Seismic interpreters can predict whether the sediment is oil-bearing or not using the AVO anomaly.



**Figure 1.2** Categories of seismic inversion.

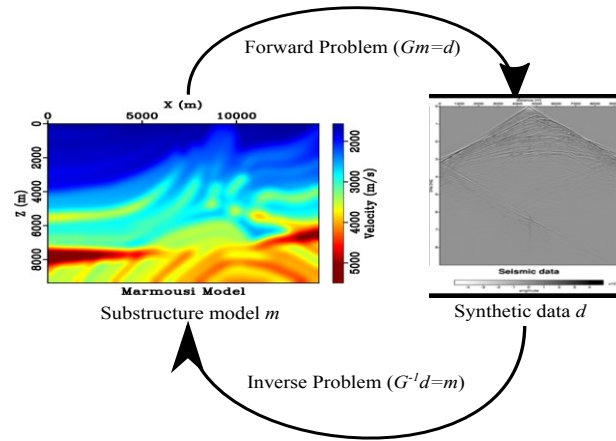
This sub-chapter has outlined the conventional procedure within geophysics exploration for hydrocarbon, however the subject of interest of this thesis is FWI where its future development could make it a unified tool to merge model building, seismic imaging and seismic interpretation delivering both high resolution subsurface images and attributes (Deal, 2014, Zhou et al, 2015, Martinez et al, 2016 and Cobo et al, 2018).

## 1.3 Full-Waveform Inversion Background

FWI is fundamentally an iterative tomographic technique that aims to recover a high-resolution and high-fidelity subsurface model detailing the physical property of its structures by utilising the whole signal waveform of the synthetic and observed seismic data. Application of FWI is not only limited to geophysical modelling and seismic data but can also be applied in fields such as medical screening,

non-destructive testing, and deep earth seismology, and can be utilised for magnetic, gravity, and electromagnetic fields as well as ultrasonic data (Seidl and Rank, 2017, Suzuki et al., 2019). In practice full waveform inversion has not been fully achieved as there are still many variables that are unaccounted for, especially during forward modelling and in calculating the local gradient based on the misfit function used.

FWI is commonly treated as a pure optimisation problem (Virieux and Operto, 2009). Misfit functions are utilised in the inversion to ultimately determine the search direction for model updating within each iteration, and to determine whether the model is sufficiently accurate. The processes for FWI are detailed in chapter 2.



**Figure 1.3** The cycle of forward modelling and seismic inversion.

Since FWI is a mathematical modelling of the physical events, there are three properties that the inverse problem must have to be considered well-posed (Hadamard, 1902).

- 1) Existence of a solution, i.e. a physical model, which exist in the case of seismic FWI, however the resulting model may not produce matching synthetic data due to noise.
- 2) Uniqueness of the solution, meaning that the observed data be re-enacted from only one geological physical model, as there could be the possibility of two significantly different models producing the same seismic data.
- 3) Stability of the solution, which whether a small perturbation in the observed data is linearly proportional to a small variation in model parameter, this is because it is ideal that suppression of noise in seismic data does not drastically change the model parameters.

It is evident that the inverse problem can satisfy properties 1) and 3) but not 2), and so is considered ill-posed, where in FWI the relations between the model and seismic data is non-linear. Measures to reduce non-linearity include formulation of a comprehensive wave equation, objective functions that employs

a more tactical approach to measure and reduce the model differences, utilisation of regularisation, and efficient optimisation algorithm.

### **1.3.1. Time domain and frequency domain**

An advantage of frequency domain FWI is the redundancy of wavenumber where accurate images of physical parameters can still be obtained from inverting with relatively fewer temporal frequencies as seismic data associated with the same location of the physical parameters over multiple frequencies and sources can be considered similar (Pratt, 1999). The redundancy also increases with aperture bandwidth due to the advantage of higher manifold coverage of a reflection point with larger offsets thus, larger offsets require fewer frequencies (Sirgue and Pratt, 2004). The reason for this redundancy is the fact that in the data space one aperture and one frequency can map one wavenumber in the model space. This leads to frequency-domain FWI having an advantage over its time-domain counterpart in that it is computationally efficient and can invert for a dense set of data. A guiding principle is that the maximum wavenumber imaged by the selected frequency matches the minimum vertical wavenumber imaged by the next frequency, which leads to frequency interval increasing with the frequency (Sirgue and Pratt, 2004). The inclusion of low frequencies in frequency-domain FWI and the wide apertures especially at long offsets can improve the modelling and inversion of intermediate and large wavelengths components of the medium. Furthermore, inclusion of these two factors can help image dipping structures by sampling small horizontal wavenumbers (Virieux and Operto, 2009).

Cycle-skipping and hence the local minima problem can be reduced with the use of the multi-scale method in time domain (Bunks et al., 1995) and with more efficiency in frequency domain (Pratt, 1999). This method involves conducting inversion over a range of increasing incremental frequencies and is generally effective because conventionally forward models are more linear low frequencies. Implementing this method in frequency domain with an appropriate frequency selection strategy (Sirgue and Pratt 2004), can produce a trade-off between computational complexity and model accuracy (Kim et al., 2011).

Inverting for low and high frequencies allows consideration of attenuation associated with visco-acoustic effects without the need for additional computation. It should be noted that low frequency waves attenuates less than their high frequency counterpart, because their longer wavelengths means that it has less oscillations and their energy is spread throughout the whole wavelength resulting in lower amplitudes. Thus, as low frequency waves propagate and encounters a contrasting media it loses less energy and can travel further, meaning it could be used to model and invert structures at lower depths. Higher frequency waves have more oscillation and higher attenuation meaning more energy is lost. A significant amount of this attenuated energy is reflections, which could be used for improvement of much finer structural details however due to the high attenuation this improvement in resolution is

limited in the depth of the earth. Thus, low frequency waves are useful for background velocity update, and high frequency waves are useful for improving model resolutions.

Forward modelling in time-domain FWI requires a time-step convolution integral to model these effects and could require significant computing time if utilising small-time steps. Frequency domain forward modelling utilises linear equation,  $Ax^*=b$ , where it applies LU decomposition and factorisation of the sparse matrix,  $A$ , and changes of vector  $b$  can be done in quick successions to obtain multiple solution vectors  $x^*$ , (explained in the forward modelling chapter). Unfortunately, FWI is limited to 2D in frequency domain because of the amount of memories needed for the large sparse matrix which would be even bigger if inverting for a 3D model, whereas such memory required isn't needed in time domain for 3D modelling. If focusing purely in 2D, conducting FWI in the frequency domain is ideal because the LU decomposition technique requires fewer computational cost, and it is easier to express and apply FWI mathematically and computationally.

## 1.4 Thesis structure

The aim of this thesis is to provide a supporting guide about FWI to those who are not familiar with it, as well as an insight into new developments of FWI. In addition, the author also included original formulations of forward modelling wave equation and objective function in sub-chapter 3.2.5 and 2.1.1 respectively. Each chapter contains a summary at the end for the purpose of briefly noting all the contents explained in the chapter.

Chapter 2 explains and makes a comparison of the various objective function utilised in FWI, as well as a brief look into regularisation that would help in mitigating the ill-posed problem of FWI mentioned before in sub-chapter 1.3. Following on, this thesis explains the steps to attain the components needed for FWI such as gradient, search direction as well as step-length.

Chapter 3 investigates the forward modelling method and the physics that generates of the wave equation that generates the synthetic data used in FWI. Within the forward modelling method, the finite difference modelling of the propagating wave, and the inclusion of absorbing boundary condition that mitigates errors within the computational simulation, as well as the source signature that is used to initiate the simulation, is examined.

Chapter 4 considers the effects of topography on seismic data, and the consequence of ignoring these effects when performing forward modelling on non-flat surfaces. Conventional flat surface forward modelling can be considered more direct, but forward modelling with topography would require spatial interpretation between physical space and computational space. The application of absorbing boundary condition within computational space is also considered.

Chapter 5 explores the recent advances in FWI development and describes the fundamentals in their application and method.

# Chapter 2: Full Waveform Inversion

This chapter explains the processes needed to conduct FWI with substantive review on the components of the optimisation problem central to FWI. As stated before, FWI seeks to retrieve parameter model(s) that truthfully reflect the spatial distribution of the physical property of the subsurface media by iteratively updating a starting model  $m^0$ , with the use of observed seismic data  $d_{obs}$ , that is retrieved in seismic acquisition or post processing. Thus, before applying FWI, forward modelling is performed to recreate seismic data  $d$ , in its synthetic variant  $d_{syn}$ , by applying a forward operator, which in seismic exploration is the geophysical wave equation  $G$ , onto the updating model representing the physical properties  $m$ , where it forms  $Gm=d$ . Details for forward modelling is described in chapter 3. Once  $d_{obs}$  and  $d_{syn}$  are available, a gradient model  $\nabla F(m^k)$ , where  $F$  stands for the forward operator function, is obtained via the adjoint state method to define the magnitude in the shift in model parameters (Plessix, 2006). The adjoint state method requires the use of the objective function  $\phi$ , and the adjoint source which is the derivative of the objective function with respect to the model  $\partial\phi/\partial m$ . When the gradient method is attained, an optimisation algorithm is performed to obtain the search direction  $D^k$ , which defines the direction for each element of the model to update. For efficient model updating and convergence (the rate to reach matching  $d_{obs}$  and  $d_{syn}$ ) an optimal step length  $\alpha^k$ . The superscript  $k$  denotes the current iteration step in FWI, where the scheme develops an updated model  $m^{k+1}$ , that is used for the next iteration, hence the iterations follow the form of  $m^{k+1}=m^k+\alpha^k D^k$ . The FWI iteration scheme is continuously performed until there is sufficient matching between  $d_{obs}$  and  $d_{syn}$ , which can be measured by the objective function.

## 2.1 Objective functions

The significance of misfit function is that it quantifies the discrepancies between  $d_{obs}$ , and  $d_{syn}$ , but more importantly it evaluates the accuracy of the final model. For frequency FWI, the objective function should be defined in frequency domain, this would involve the observed data undergoing Fourier transform.

The more conventionally used objective functions are the least absolute L1-norm  $\phi_{L1}$ , and the least squares L2- norm method  $\phi_{L2}$ , which is defined as being magnitude of differences between the datasets and squared differences between the datasets, respectively.

$$\begin{aligned}\phi_{L1}(x_s, x_r, \omega) &= \|d_{obs}(x_s, x_r, \omega) - d_{syn}(x_s, x_r, \omega)\|_1, \\ \phi_{L2}(x_s, x_r, \omega) &= \frac{1}{2} \|d_{obs}(x_s, x_r, \omega) - d_{syn}(x_s, x_r, \omega)\|_2^2,\end{aligned}\tag{2.1}$$

The values of the two-dataset selected for the objective function, are determined by the lateral position of its receivers  $x_r$ , and sources  $x_s$ . The norm  $\|\dots\|$  used in  $\phi_{L1}$  is simply the sum of the data differences



but for  $\phi_{L2}$  it is the SRSS (square root sum square) of the data being recorded at the individual receivers. The summation being conducted for the norm is the summation of data at every timestep in time domain for each receiver, but in frequency domain it sums the data along each frequency interval  $\omega$ . The adjoint state method utilises the back-propagated wavefield to determine the gradient model  $\nabla f(m^k)$  which was adapted to the theory of inverse problem (Chavent, 1974) from the control theory (Lions, 1972, Plessix, 2006). The gradient for  $\phi_{L2}$  is just the SRSS of the data differences, whereas for  $\phi_{L1}$  the gradient only involves the sign of the co-efficient of the function denoted by  $\text{sgn}$ .

$$\begin{aligned}\frac{\partial \phi_{L1}(x_s, x_r, \omega)}{\partial m} &= \text{sgn}\left(\|d_{obs}(x_s, x_r, \omega) - d_{syn}(x_s, x_r, \omega)\|_1\right) \frac{\partial d_{syn}(x_s, x_r, \omega)}{\partial m}, \\ \frac{\partial \phi_{L2}(x_s, x_r, \omega)}{\partial m} &= \|d_{obs}(x_s, x_r, \omega) - d_{syn}(x_s, x_r, \omega)\|_2 \frac{\partial d_{syn}(x_s, x_r, \omega)}{\partial m},\end{aligned}\quad (2.2)$$

It is generally easier to compute the derivative of the  $L_2$ -norm, which allows for easier computation of the gradient model. The  $L_2$ -norm is more likely to have a unique solution because the function reduces error by the singular mean value but can be negatively affected from the influence of outliers arising from measurement inaccuracies. Contrariwise the  $L_1$ -norm is more robust and is not affected as much by outliers since it reduces the error by median values between the two data. This unfortunately results in non-unique solutions making  $L_1$ -norm FWI non-linear.

Van Leeuwen and Mulder (2010) developed the cross-correlation objective function  $\phi_{cor}$  which it either maximises or minimises when relative phase shift or travel-time difference between the two dataset is zero. In frequency domain the cross-correlation can be simply calculated by multiplying two points within each dataset. Inconveniently this misfit function requires the Fourier transform  $FT$ , and inverse Fourier transform  $FT^{-1}$ , of the cross-correlated function  $C(x_s, x_r, \omega)$  to apply the penalty function  $P(\tau)$  which eliminates energy at travel-time differences greater than the maximum allowable time lag  $\tau_l$ .

$$\begin{aligned}\phi_{cor}(x_s, x_r, \omega) &= FT\left(\phi_{cor}(x_s, x_r, t)\right) = FT\left(\frac{1}{2}\|P(\tau)FT^{-1}(C(x_s, x_r, \omega))\|^2\right), \\ C(x_s, x_r, \omega) &= d_{obs}(x_s, x_r, \omega)\overline{d_{syn}(x_s, x_r, \omega)}, \\ P(\tau) &= \begin{cases} 1 & |\tau| \leq \tau_l \\ 0 & |\tau| > \tau_l \end{cases},\end{aligned}\quad (2.3)$$

The overhead bar above  $d_{syn}$  denotes its conjugate gradient. Since this function involves the  $L_2$ -norm of the cross-correlation, the derivative of this function can be obtained by applying the penalty function to the time domain cross-correlation and then performing Fourier transform upon it.

$$\frac{\partial \phi_{cor}(x_s, x_r, \omega)}{\partial m} = FT\left(\|P(\tau)FT^{-1}(C(x_s, x_r, \omega))\|\right) \frac{\partial d_{syn}(x_s, x_r, \omega)}{\partial m}, \quad (2.4)$$

On top of the inconvenient extra steps of applying both forward and inverse Fourier transform, another drawback is that once the data matches, the peak of the correlation function is centred but not confined to the zero-time lag, meaning that the model would still be updated even if the model is correct. The deconvolution misfit function  $\phi_{dec}$ , was proposed to solve this problem, where the cross-correlation is being deconvolved by the autocorrelations of the observed data (Luo and Sava, 2011). In the frequency domain the deconvolution is simply a division, where this helps to confine its energy to zero-time lag. The factor that makes deconvolution unique compared to other objective function is that it aims to measures and reduce errors to be as close to unity as possible, because it compares the relative difference rather than the absolute difference. Normalization helps to amplify low-frequency components, which makes the inversion less prone to cycle-skipping.

$$\phi_{dec}(x_s, x_r, \omega) = \frac{1}{2} \left\| \frac{\overline{d_{obs}(x_s, x_r, \omega)} d_{syn}(x_s, x_r, \omega)}{d_{obs}(x_s, x_r, \omega) d_{obs}(x_s, x_r, \omega) + \varepsilon^2} \right\|^2, \quad (2.5)$$

$\varepsilon$  is an arbitrary small value that is used in the case of deconvolution objective function as stabilising factor to prevent the cross-correlation from overcoming the autocorrelation, as it is supposed to be a fractional function. The overhead bar denotes the complex conjugate of the signal. Again, due to the  $L_2$ -norm of the deconvolution the derivative of the misfit function is simply a division.

$$\frac{\partial \phi_{dec}(x_s, x_r, \omega)}{\partial m} = \frac{d_{obs}(x_s, x_r, \omega) d_{syn}(x_s, x_r, \omega)}{\overline{d_{obs}(x_s, x_r, \omega)} d_{obs}(x_s, x_r, \omega) + \varepsilon^2} \frac{\partial d_{syn}(x_s, x_r, \omega)}{\partial m}, \quad (2.6)$$

However, the problem with this method lies within its aim to reduce the error to be as close to the unity as possible and hence it would be difficult to absolutely remove the errors. It was suggested to apply this method with the  $L_2$ -norm difference misfit function where the deconvolution misfit function is applied first to obtain an accurate starting model for the difference-based misfit function (Luo and Sava, 2011).

Previously stated objective function tries to minimise the misfit of either part of or the whole waveform. Interestingly the waveform can be deconstructed for the objective function that utilises a logarithmic wavefield  $\phi_{log}$  (Shin and Min, 2006). The waveform of  $d_{obs}$  and  $d_{syn}$  can be deconstructed into the amplitude  $A_{obs}$  and  $A_{syn}$ , and the phases  $\theta_{obs}$  and  $\theta_{syn}$ . Because of this separability inversion can be conducted and tested with the objective function based on amplitude only (Pyun et al, 2007), phase only (Bednar, 2007), and both (Shin et al, 2007).

$$\phi_{log} = \frac{1}{2} \left\{ \ln \left( \frac{A_{syn}(\omega)}{A_{obs}(\omega)} \right)^2 + (\theta_{syn}(\omega) - \theta_{obs}(\omega))^2 \right\}, \quad (2.7)$$

It should be noted that the phase obtained from the frequency-domain seismic data are phase wrapped i.e. constrained between  $-180^\circ$  and  $180^\circ$ , and when the actual phase is outside of this range it can be shifted by a multiple of  $360^\circ$  to keep them within these range. Wrapped phases  $\theta_w(\omega)$  occurs in the frequency domain due to a lack of low frequency data, which can cause phase shifts of seismic events represented in the data and cause local minima problem largely due to unavailability of global minima (Bunks et al, 1995). Hence the need for phase unwrapping where a multiple  $n$  of  $2\pi$  is added to reflect actual phase value  $\theta_w(\omega)$  of a continuous function i.e.  $\theta(\omega) = 2\pi n + \theta_w(\omega)$ . Should the starting model be assumed as already highly accurate, then the degree of phase unwrapping of both synthetic and observed data can be considered similar, i.e.  $2\pi n_{cal} + 2\pi n_{src} = 2\pi n_{obs}$ . Where  $n_{cal}$ ,  $n_{src}$ , and  $n_{obs}$  denote the integer multiple of  $2\pi$  used in phase unwrapping for green's function, source wavelet and observed wavefield respectively. The gradient of this objective function was found as follows:

$$\frac{\partial \phi_{\log}}{\partial m} = \frac{1}{2} \left\{ \ln \left( \frac{d_{syn}(\omega)}{d_{obs}(\omega)} \right)^* \frac{1}{d_{syn}(\omega)} \frac{\partial d_{syn}(\omega)}{\partial m} \right\}, \quad (2.8)$$

Where the two terms on the R.H.S. of equation 2.8 is complex and has a real term and an imaginary term times  $i$ .

$$\ln \left( \frac{d_{syn}(\omega)}{d_{obs}(\omega)} \right) = \ln \left( \frac{A_{syn}(\omega)}{A_{obs}(\omega)} \right) + i(\theta_{syn}(\omega) - \theta_{obs}(\omega)), \quad (2.9a)$$

and

$$\frac{1}{d_{syn}(\omega)} \frac{\partial d_{syn}(\omega)}{\partial m} = \frac{1}{A_{syn}(\omega)} \frac{\partial A_{syn}(\omega)}{\partial m} + i \frac{\partial \theta_{syn}(\omega)}{\partial m}, \quad (2.9b)$$

The value of amplitude can be obtained from the complex frequency domain seismic data as a means of finding the root-sum-square of the real  $\Re d(\omega)$  and imaginary component  $id(\omega)$ , whereas the phase can be obtained by trigonometric means.

$$\begin{aligned} A_{syn} &= \sqrt{\Re d_{syn}(\omega)^2 + id_{syn}(\omega)^2}, \\ A_{obs} &= \sqrt{\Re d_{obs}(\omega)^2 + id_{obs}(\omega)^2}, \end{aligned} \quad (2.10a)$$

$$\begin{aligned} \theta_{syn} &= \tan^{-1} \left( \frac{id_{syn}(\omega)}{\Re d_{syn}(\omega)} \right), \\ \theta_{obs} &= \tan^{-1} \left( \frac{id_{obs}(\omega)}{\Re d_{obs}(\omega)} \right), \end{aligned} \quad (2.10b)$$

It was found that this simultaneous inversion of both amplitude and phase produces better results in the shallow and deep part of the model with data containing relatively low frequencies, and the presence of random noise in the data has little to no effect on the inversion (Shin and Min, 2006). When inverting with both amplitude and phase it was found the logarithmic objective function was more robust in comparison to conventional  $L_2$ -FWI because of the scaling of the residual wavefield amplitude by the synthetic wavefield which allows the computation to be stable (Shin et al, 2007). Logarithmic Phase only inversion was found to produce clearer and more accurate image in comparison to the conventional phase only inversion (Bednar et al, 2007). Unlike the two comparisons that included phase information, the amplitude only logarithmic inversion produces poorer result even in comparison to conventional amplitude inversion, this is mainly due to the lack of phase/time-based information that defines structural locality.

Regardless, logarithmic inversion struggles when there is lack of frequencies below 5Hz in  $d_{obs}$  (Shin and Min, 2006). To deal with the lack of frequencies below 5Hz in seismic data, the Envelope misfit function  $\phi_{env}$ , was introduced (Bozdağ et al, 2011).

$$E = \sqrt{\Re d(\omega)^2 + i d_h(\omega)^2} = \sqrt{\Re d(\omega)^2 + \left( \Re d(\omega) \left( \tan \{ \theta_h(\omega) \} \right) \right)^2}, \quad (2.11)$$

Where the instantaneous envelope  $E$ , retrieves the low frequency data by hypothetically extrapolating data trends from the seismic data, and create an envelope over the dataset.  $E$  is calculated in a similar manner for the amplitude in the natural logarithm objective function equation 2.10a with the exception that it utilises the Hilbert transformed phase  $\theta_h(\omega)$ .

$$\begin{aligned} \phi_{env} &= \frac{1}{2} \left[ \ln \left( \frac{E_{obs}}{E_{syn}} \right) \right]^2, \\ \phi_{env} &= \frac{1}{2} \left\| \theta_{obs} - \theta_{syn} \right\|^2, \end{aligned} \quad (2.12)$$

The gradient of objective function:

$$\begin{aligned} \frac{\partial \phi_{env}}{\partial m} &= \ln \left( \frac{E_{obs}}{E_{syn}} \right) \frac{\partial d_{syn}(x_s, x_r, \omega)}{\partial m}, \\ \frac{\partial \phi_{env}}{\partial m} &= (\theta_{obs} - \theta_{syn}) \frac{\partial d_{syn}(x_s, x_r, \omega)}{\partial m}, \end{aligned} \quad (2.13)$$

Liu and Zhang, 2017 developed an objective function for joint travel-time and envelope inversion which is essentially the combined residual  $L_2$ -norm misfit function for both envelope and phase and aims to combine travel-time inversion with that of low frequency envelope inversion. A chosen weighting coefficient  $w$  determines whether the objective function leans completely to envelope or phase residual

when  $w=0$  or  $=1$  respectively, any value between 0 and 1 would allow the use of both. Combining inversion for both travel-time and envelope waveform inversion helps to reduce the cycle-skipping.

$$\begin{aligned}\phi_{env} &= \frac{1}{2} \left\{ (1-w) \|E_{obs} - E_{syn}\|^2 + w \|\theta_{obs} - \theta_{syn}\|^2 \right\}^2, \\ \frac{\partial \phi_{env}}{\partial m} &= (1-w) \|E_{obs} - E_{syn}\| \frac{\partial \|E_{obs} - E_{syn}\|}{\partial m} + w \|\theta_{obs} - \theta_{syn}\| \frac{\partial \|\theta_{obs} - \theta_{syn}\|}{\partial m},\end{aligned}\quad (2.14)$$

When applying an envelope over seismic data there is a risk that ultra-low frequency (ULF) signals could be lost, to account for this loss the instantaneous envelope is approximated  $E_{ULF}$ , for both synthetic and observed data, where it's formed from the summed seismogram of the direct arrivals and reflected series (Wu et al., 2014). Each seismogram is the product of a carrier signal and a modulator. The carrier signal for the direct arrivals and reflected series are the source wavelet  $s(\omega)$ , and the reflected series coefficient  $\gamma(\omega)$ . The modulators for the direct arrivals  $g_d$ , and reflected series  $G_{sr}$ , are green's function propagator where  $G_{sr}$ , is formed from the scalar multiplication of the green's function for the source  $g_s$ , and receiver  $g_r$ . Green's function are attained in linear algebra  $Ax^*=b$ , vector  $b$  equates to a dirac delta  $\delta$ , at the source and reflected points locations, and sparse matrix  $A$ , contains the forward wave differential operator, and the desired Green's function is the vector  $x^*$ , (Application of the linear algebra is explained in sub-chapter 3.4 in the case of forward modelling).

$$E_{ULF}^2 = (g_d)^2 E^2[s(\omega)] + (G_{sr})^2 E^2[\gamma(\omega)], \quad (2.15a)$$

$$G_{sr} = |g_s(\omega)| |g_r(\omega)|, \quad (2.15b)$$

Tests conducted using the envelope inversion (equation 2.15a) and  $L_2$ -FWI for full-band source and low-cut source wavelet, which is cut from 5Hz below, was found to produce no differential results between the two-source wavelet spectrum. The combination of envelope inversion and  $L_2$ -FWI can recover long wavelength structure i.e. background and deeper structure, better than using purely  $L_2$ -FWI when low frequency data is unavailable (Wu et al., 2014). The combined envelope and  $L_2$  inversion were found to be highly resistive to white gaussian noise, especially at high frequencies, and low seismic interference noise when signal-to-noise (SNR) ratio is around 1-2 but has difficulties with SNR at 0 (Wu et al., 2015). The improved SNR ratio was found to be aided when the envelope of both waveforms are squared.

$$\begin{aligned}\phi_{env} &= \frac{1}{4} \|E_{obs}^2 - E_{syn}^2\|^2, \\ \frac{\partial \phi_{env}}{\partial m} &= \frac{1}{2} \|E_{obs}^2 - E_{syn}^2\| \frac{\partial d_{syn}(x_s, x_r, \omega)}{\partial m},\end{aligned}\quad (2.16)$$

### 2.1.1 Joint deconvoluted envelope and phase residual misfit Function

Following the previous sub-chapter in regard to objective function, this research aims to utilise the multiple features stated before and develop a strong robust objective function  $\phi_J$ , to provide resistance to the noise and avoid cycle-skipping problems that can occur in the initial stages of FWI especially with a highly inaccurate starting model.

$$\phi_J = \frac{1}{4}(1-w) \left\{ \left\| \frac{E_{syn}^2}{E_{obs}^2 + \varepsilon^2} \right\| \right\}^2 + \frac{1}{2} w \left\{ \left\| \theta_{obs} - \theta_{syn} \right\| \right\}^2, \quad (2.17)$$

This misfit function utilises the envelopes of the waveforms to consider the lack of low frequency data within the seismic data, the square of these waveforms is to improve SNR ratio in order for the inversion to be resistive to gaussian white noise and seismic interference noise. Then the deconvolution would be taken of these waveform envelopes to assist with the bandlimited data and produce impulsive adjoint source where its sensitivity kernel displays more precise locale of the structural anomaly to be updated, and the gradient is more accurate (Luo and Save, 2011). The phase difference is to mitigate cycle-skipping that are associated with travel-time differences in seismic data which indicates inaccurate structural locations in the updating model. The same weighting co-efficient  $w$ , used in the joint travel-time and envelope inversion (Liu and Zhang, 2017), where initially it starts at 0 to focus purely upon the envelope for the low-frequency information and amplitude matching, then increments towards 1 where it gradually tends to the phase information.

The gradient of this objective function is simply the gradient of both factors, and during iterations it should focus initially to produce impulsive adjoint source for accurate gradient updating.

$$\frac{\partial \phi_J}{\partial m} = \frac{1}{2} \left\{ (1-w) \left\| \frac{E_{syn}^2}{E_{obs}^2 + \varepsilon^2} \right\| \right\} \frac{\partial d_{syn}(x_s, x_r, \omega)}{\partial m} + w \left\| \theta_{obs} - \theta_{syn} \right\| \frac{\partial d_{syn}(x_s, x_r, \omega)}{\partial m}, \quad (2.18)$$

Inversion with this objective function is recommended to be conducted in scale with  $L_2$ -FWI where initially the starting model is to be updated with the proposed joint misfit function in inversion until a suitably accurate model is attained, and  $L_2$ -FWI is implemented. users can designate a certain level of difference between  $d_{obs}$  and  $d_{syn}$  before transitioning from the jointed objective function to  $L_2$ -FWI.

## 2.2 Regularisation

When performing inversion that requires mathematical modelling of physical phenomena such as wave equation in geophysics, it is critical to ask if the problem is well-posed, and is defined as having the following three properties (Hadamard, 1902):

1. Existence of solution.
2. Uniqueness of solution.
3. Continuous Change of the solutions' behaviour because of the change in initial condition.

Whilst the first and third point are mostly achieved with straightforward FWI the second point is the most important yet difficult to attain. This means that FWI is inherently ill-posed due to the potential of multiple solutions in each inversion iteration, and hence is nonlinear. Regularisation is employed to suppress singularities that can cause erroneous solutions and produce well-behaved model. A model or data is considered well-behaved when the requirements used in the operators that generate the model and data is satisfied.

One of the methods to perform regularisation is by adding a penalty to the objective functions. One of the popular forms of these penalties is the model constraint, which limits significant changes to the model that would otherwise return highly erroneous results, in other words it limits the number of possible model solutions. In this thesis for simplicity reasons the  $L_2$ -norm objective function is used in equation 2.19 to demonstrate the formulation with regularisation  $R$ .

$$\phi_{L_2}(x_s, x_r, \omega) = \frac{1}{2} \|d_{obs}(x_s, x_r, \omega) - d_{syn}(x_s, x_r, \omega)\|_2^2 + \varpi R(m), \quad (2.19)$$

The simplest model constraint is the  $L_p$ -norm constraint where a trade-parameter  $\varpi$ , (parameter that controls the significance of the constraint in the regularised objective function) is applied to the least squares of the difference between the current model  $m$ , and the model estimate  $m_e$  that is derived from a priori information.

$$R(m) = \|m - m_e\|_p^p, \quad (2.20)$$

It should be noted that higher values of  $p$  results in the constraints being less tolerance to outliers, (Wang, 2016), conventionally  $p$  has a value of 1 (lasso regression) or 2 (ridge regression). The key difference between  $L_1$  and  $L_2$  is that  $L_1$  constraint can over-emphasise constraint in some elements within the model, whereas  $L_2$  considers constraint spread throughout the whole model, hence  $L_2$ -norm constraint is popular. The advantages of these model constraint are that they are straightforward to implement however it requires an accurate model estimate which if it is not it would result in highly erroneous results. Like finding the derivative of the objective function so too must the derivative be found for the model constraint, equation 2.21 displays the derivative of the  $L_1$ - and  $L_2$ - norm as the first and second formula, respectively.

$$\frac{\partial R_1(m_e)}{\partial m_e} = \text{sgn}(m_e), \quad \frac{\partial R_2(m_e)}{\partial m_e} = 2m_e, \quad (2.21)$$

The Tikhonov regularisation (Tikhonov, 1935) can be considered a more specific form of the  $L_2$ -norm regularisation, and the two are often used interchangeably, this method utilises a Tikhonov matrix  $\Gamma$ , to the model parameters. Tikhonov matrix are an explicit matrix, where a functional scale  $\alpha_f$ , is applied to an identity matrix i.e.  $\Gamma=I\alpha_f$  usually it doesn't involve model difference however if  $\alpha_f$  is the model difference then the first formula in equation 2.22 becomes equation second formula of equation 2.21.  $\alpha_f$  in the second formula in equation 2.22 can be thought of as an extended form of the Tikhonov regularisation, where  $r$  is the position of the model parameters and  $w$  denotes weighting function (Tikhonov and Arsenin, 1977, Tikhonov et al, 1995).

$$\begin{aligned} R(m) &= \|\Gamma m\|_2^2, \\ \alpha_f &= w_1(r) + w_2(r) \frac{\partial}{\partial r} \end{aligned} \tag{2.22}$$

The derivative of the Tikhonov regularisation is essentially similar to that of the derivative of  $L_2$ -norm regularisation, with the exception that 2 is scaled by  $\alpha_f$ . Advantages of the Tikhonov regularisation is that it can reconstruct smooth features in the model, and because it forces model recovered from inversion iterations to be consistent with a priori information given e.g. model estimate, or models inverted for at previous iterations (Aghazade et al, 2015). However, over-smoothing can pose a problem with Tikhonov regularisation where elements in the solution model fails to have the expected smoothness, and the regularisation would not have a discontinuous solution i.e. a final correct model solution (Hofmann and Mathé, 2017).

The other most popularly used model constraint is the total variation (TV) regularisation, this constraint is used to improve inversions of models by preserving sharp features whereas Tikhonov regularisation aims to recover smooth models (Guitton, 2012).

The advantages of TV regularisation are its ability to recover models with sharp boundaries without the smudging effect as seen from FWI without TV (Esser, 2018), demonstrating that inclusion of TV regularisation helps reduce certain unwanted artefacts in inversion results and preserve the true time-lapse model differences (model inverted at each iteration). Just as TV regularisation helps preserve sharp features in the model, it also has a negative side-effect in that discretisation of the above equation can introduce high-frequency artefacts (Estrela et al, 2016), and discretisation at the edge of the model may cause numerical difficulties, unless suitable boundary condition is introduced.

The Maximum entropy constraint though not as popular as Tikhonov and TV regularisations, is applied as a regularisation term based on the principle that random variables have uncertainties (disorders in a system) that should be constrained based upon the information one has at the moment (Jayne, 1968, Wang, 2016). This constraint involves summation of the natural logarithm of the model parameters, and motive for this constraint is to enhance details of low-intensity outliers within the model parameter misfit.



$$R(m) = \sum_{i,j} |m_{i,j}| \ln |m_{i,j}|, \quad (2.23)$$

The subscript  $i$  and  $j$ , denotes the horizontal and vertical co-ordinate of an element of a matrix that represents the spatial locations of physical properties. Derivative of the Maximum Entropy constraint is simply 1 added to the natural log of the model value.

$$\frac{\partial R(m)}{m} = \sum_{i,j} 1 + \ln |m_{i,j}| \frac{\partial d_{syn}(x_s, x_r, \omega)}{\partial m}, \quad (2.24)$$

The effectiveness of this model constraint is that like TV it utilises the available current model instead of having to rely on a model estimate (Phillips and Elith, 2013), as well as retrieving the background smooth model with fine structural details in the model (Wang, 2016). However, the limitations are that the constraint can cause over-fitting and introduce biases to the inversion (Phillips and Dudik, 2008).

Existence of other model constraints are also available, including but not limited to, the Cauchy constraint which is used to control sparseness of the inversion result (Liu et al, 2009, 2017), Modified TV regularisation which expand TV regularisation with first and second order spatial derivatives (Lin and Huang, 2015), as well as constraints on the forward modelling geophysical operators, (Backus and Gilbert, 1968). To cover all of regularisation is outside the scope of this thesis. The choice of regularisation technique depends upon the specific characteristics of the inversion solution model e.g., blocky, or smooth, or the dataset used in inversion e.g., flat angle gathers (Ramirez and Lewis, 2010). Although regularisation are effective tools there is still room for improvements where previous implementation of regularisation to FWI are inflexible, future research are conducted to allow regularisation to adapt to the data residual used in inversion (Aghamiry et al, 2020) or the wave equation utilised in forward modelling (Gao and Huang, 2019).

## 2.3 Derivation of the Gradient

The gradient describes the shift in  $m^k$ , to result in a model that reflects the true subsurface region as much as possible, where collaterally the misfit function would decrease towards zero, thus FWI is considered a local iterative scheme.

In conventional FWI the adjoint-state method is used to acquire functional gradients that are dependent upon state variables and adjoint state variables, where the former are solutions of linear systems used in forward modelling i.e. wave propagation, and the latter are solutions of the adjoint linear operator i.e. backpropagation. In geophysics the state variable is the wavefield vector  $\mathbf{u}$ , being perturbed by the wave equation matrix  $\mathbf{S}$ , that results in the source vector  $\mathbf{f}$ , described in the form of linear algebra  $Ax^* = b$ , which characterises the frequency domain wave equation (equation 2.25). The adjoint state variable is

the backpropagated wavefield  $\mathbf{u}_b$  obtained via reflective modelling of the gradient misfit functionals at the receivers.

The advantages of this method are that only a single extra linear system equation is needed because the adjoint-state variable can be seen as measurements of derivative of the forward model perturbation i.e. misfit functional of the dataset, with respect to state variables, and thus are not considered as derivatives based upon model parameters.

$$\mathbf{S}\mathbf{u} = \mathbf{f}, \quad (2.25)$$

The gradient of the objective functions is derived as the first-order derivative of linear equation 2.25 with respect to model parameters, via product rule differentiation of the linear system to a null right-hand side vector.

$$\mathbf{S} \frac{\partial \mathbf{u}}{\partial m^k} + \frac{\partial \mathbf{S}}{\partial m^k} \mathbf{u} = 0, \quad (2.26)$$

Equation 2.26 is rearranged to obtain the Fréchet matrix  $\partial \mathbf{u} / \partial m^k$ , which is essentially the Jacobian matrix,  $\mathbf{J}$ , used for the minimization of the error (misfit) functional.

$$\mathbf{J} = \frac{\partial \mathbf{u}}{\partial m^k} = -\mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial m^k} \mathbf{u}, \quad (2.27)$$

Substituting the wavefield  $\mathbf{u}$  with that of the objective function can in turn generate the gradient  $\partial \phi / \partial m$ , and applies a perturbation in wavefield  $\Delta u$ . The perturbation also causes  $\mathbf{u}$ ,  $\partial \mathbf{S} / \partial m^k$  and  $\mathbf{S}$ , to undergo (Hermitian) conjugate transpose (superscript  $H$ ). Physically the conjugate represents time reversal, and the transpose represents backpropagation.

$$\frac{\partial \phi}{\partial m^k} = -\mathbf{u}^H \left( \frac{\partial \mathbf{S}}{\partial m^k} \right)^H \left( \mathbf{S}^{-1} \right)^H \Delta u, \quad (2.28)$$

The functional gradients outlined in misfit function sub-chapter 2.1 are calculated at the receivers to obtain the adjoint source for the adjoint operator, where  $\mathbf{u}_b$  can be defined as the conjugate transpose of the forward operator.

$$\mathbf{u}_b = -\left( \mathbf{S}^{-1} \right)^H \Delta u, \quad (2.29)$$

Equation 2.29 can be rearranged to obtain the adjoint of the linear operator for the backpropagation model similar to that of equation 2.25, where the adjoint source vector can be considered as the complex conjugate of the objective function at receivers' location, the complex conjugate is denoted as an overhead bar. Hence  $\mathbf{u}_b$  can be obtained by the simply reversing the sign of the imaginary component of the complex conjugated backpropagation  $\overline{\mathbf{u}_b}$ .

$$\mathbf{S}\overline{\mathbf{u}_b} = \overline{\Delta\mathbf{u}}, \quad (2.30)$$

Ultimately to obtain the global gradient adjoint state method involves the multiplication between the wavefield of the complex conjugate of the forward modelling  $u_{forward}$ , with that of the back-propagation model  $u_{back}$ , and the summation of these multiplications over temporal frequency and source location. The following equation for the gradient is taken as the gradient with respect to the P-wave velocity model  $V_p$ , (sub-chapter 3.1). The gradient can be taken over the model which the user would like to update e.g., S-wave velocity model, or density model.

$$\nabla F(m^k) = \frac{\partial \phi}{\partial m^k} = \sum_{\omega} \sum_s \frac{2\omega^2}{V_p^3} u_{forward}(x_s, \omega) u_{back}(x_r, \omega), \quad (2.31)$$

Because of the temporal frequency, utilising the code library MUMPS for LU decomposition (explained in sub-chapter 3.4) is practically more time efficient to solve for all the sources, before moving onto the following temporal frequency.

## 2.4 Optimisation algorithm

Optimisation algorithms are used to find the most ideal change of elements in the minimisation or maximisation of a functional. In seismic inversion case it is to find the minimal value within the objective function, because in doing so would find the smallest data misfit. In this sub-chapter various optimisation algorithms are described and evaluated to find the most viable technique for convergence, whether it is for accuracy, time consumption, or stability. Optimisation algorithms are designed to specifically find the roots of the functional that indicates zero-valued functional thus zero-valued data misfit, and a successful inversion.

### 2.4.1 Search directions

Smooth Optimisation methods assume that the functional is continuously differentiable, what that means is that the functional is assumed to be an unbroken (no null gap) continuous function with a variable functional gradient for every point in the model. The most straightforward method is the

gradient method of where the negative gradient (equation 2.31) is directly used to iteratively update the model.

$$D_{grad}^k = -\nabla F(m^k), \quad (2.32)$$

The difficulty with the pure gradient method is that the model update does not consider the size of the minimization of the misfit, and depends on a constant valued step length,  $\alpha^k$ . Should the step length be too large it would overshoot past the minimum whereas if it is too small, convergence would be slow, due to requirement of many iterations.

One of the earliest methods to quicken convergence is the steepest descent method which aims to find the descent in the largest directional derivative. This method begins with the calculation of the pseudo-misfit function for the next step  $\phi(m^{k+1})$ , with the use of the Taylor-Lagrange expansion of the current misfit function  $\phi(m^k)$ , (Virieux and Operto, 2009).

$$\phi(m^{k+1}) = \phi(m^k + \alpha^k D^k) \approx \phi(m^k) - \alpha^k \langle \nabla F(m^k), \nabla F(m^0) \rangle + \frac{1}{2} (\alpha^k)^2 \nabla F(m^0)^T (H_a^k) \nabla F(m^0), \quad (2.33)$$

This expansion utilises inner product  $\langle, \rangle$ , between the gradient of current  $k$  iteration and gradient of starting model  $m^0$ , and the approximate hessian matrix  $H_a^k$ , containing the second order derivative of the functional with respect to model parameters, calculated with equations 2.39b. By setting  $\phi$  in equation 2.33 to zero, and rearrange  $\nabla F(m^k)$  and  $H_a^k$ , within the first iteration of FWI i.e.  $m^k = m^0$ , the step length  $\alpha^k$ , for the steepest descent can be derived. However,  $H_a^k$  can be obtained as the product of two Jacobian matrices of the starting model  $J(m^k)$ , i.e.  $H_a^k = J(m^0)^T J(m^0)$ . The Jacobian matrix, also known as the Fréchet derivative, is the first order derivative of the state variable with respect to the model,  $J^0 = \partial u / \partial m$ . Thus,  $\alpha^k$  can be calculated as follows.

$$\alpha^k = 2 \frac{\langle \nabla F(m^0), \nabla F(m^0) \rangle}{\nabla F(m^0)^T H^k \nabla F(m^0)} = 2 \frac{\langle \nabla F(m^0), \nabla F(m^0) \rangle}{\langle J(m^0)^T \nabla F(m^0), J(m^0)^T \nabla F(m^0) \rangle}, \quad (2.34)$$

The term  $J(m^0)^T \nabla F(m^k)$  can be calculated as the first order finite difference derivative of the forward operator  $f$ , over a small value  $\varepsilon$  that is used as the discretised step length for the steepest descent gradient.

$$J(m^k)^T \nabla F(m^k) \approx \frac{f(m^k + \varepsilon \nabla F(m^k)) - f(m^k)}{\varepsilon}, \quad (2.35)$$

Once  $\phi(m^{k+1})$  is obtained then the adjoint source must be derived from it to calculate the gradient which is used as the search direction for steepest descent  $D_{sd}^k$ . Pratt et al., 1998 tested the steepest descent and

concluded that this method could fail to converge regardless of how many iterations, because it is difficult to estimate a reliable pseudo-step length,  $\varepsilon$ . The conjugate gradient method was developed as local optimisation to converge in fewer iterations (Hestenes and Stiefel, 1952), and was tested by Mora et al., 1987. This method derives the current search direction  $D_{cj}^k$ , from the gradient  $\nabla f(m^k)$ , and the direction at previous iteration  $D_{cj}^{k-1}$ , multiplied by the scalar  $\beta^k$ , this scalar ensures that the  $D_{cj}^k$  and  $D_{cj}^{k-1}$  are conjugate.

$$D_{cj}^k = \begin{cases} \nabla F(m^k) & k = 0, \\ \nabla F(m^k) + \beta^k D_{cj}^{k-1} & k \geq 1, \end{cases} \quad (2.36)$$

There are multiple methods to compute  $\beta^k$ , however it was found that earlier incarnations would result in convergence failing, negligible updating or even descent loss, where the misfit increase instead. Thus, a hybrid scheme was developed to determine an optimal scalar  $\beta_{HZ}^k$ , which selectively chooses from the Hestenes-Stiefel scalar  $\beta_{HS}^k$ , and the Dai-Yuan scalar  $\beta_{DY}^k$ , where it exploits the strengths of previous scalars and can be independent from the line search (Hestenes and Stiefel, 1952, Dai and Yuan, 1999, Hagel and Zhang, 2006).

$$\begin{aligned} \beta_{HZ}^k &= \max(0, \min(\beta_{HS}^k, \beta_{DY}^k)), \\ \beta_{HS}^k &= \frac{\langle \nabla F(m^k), \nabla F(m^k) - \nabla F(m^{k-1}) \rangle}{\langle D^k, \nabla F(m^k) - \nabla F(m^{k-1}) \rangle}, \\ \beta_{DY}^k &= \frac{\langle \nabla F(m^k), \nabla F(m^k) \rangle}{\langle D^{k-1}, \nabla F(m^k) - \nabla F(m^{k-1}) \rangle}, \end{aligned} \quad (2.37)$$

Instead of using the equation 2.34 to calculate step length it can be derived from a line search which attempts to test initial step length until conditions are satisfied, explained in sub-chapter 2.3.3. The conjugate gradient method is ideal for large-scale problems as it uses limited computational memory and does not need to solve a numerical linear algebra equation. Thus, each iteration is conducted at a fast rate, however they still require many iterations typically in the hundreds due to poor estimation of the step length.

An alternative to the conjugate gradient method is the Newton method, which is more effective at finding the global optimal solution. Newton methods involve the division of the gradient over the full hessian matrix.

$$D_{newton}^k = \frac{\nabla F(m^k)}{H^k}, \quad (2.38)$$

Two Newton methods that are available are categorised based on the hessian utilised, the full-newton method employs the full hessian  $H_{full}^k$ , and the gaussian-newton method employs only  $H_a^k$ . In addition to  $H_a^k$ , the full-newton method derives the hessian matrix from differentiation by product of the derivative of the objective function with respect to the model.

$$H_{full}^k = \Re \left( \frac{\partial J(m^k)^T}{\partial m} \Delta d + H_a^k \right), \quad (2.39a)$$

Gaussian-newton method argues that computation of the full hessian is unnecessarily computationally expensive as  $H_a^k$  constitute most of the full hessian value, and so the extra term is negligible.  $H_a^k$  is based upon truncating the Taylor series of the residual and is the square of the Jacobian matrices, which are calculated the same way as in equation 2.27, where the Jacobian can be considered as the first order derivative of the wavefield with respect to the model.

$$H_a^k = J(m^k)^T J(m^k), \quad (2.39b)$$

However, computation of both Jacobian and hessian matrices requires extensive computer memory storage especially for a large-scale problem, an alternative is to use the quasi-newton methods.

## 2.4.2 Quasi-Newton method

Quasi-Newton methods are generalisations of the secant method, which are algorithms that finds approximate roots to a function with a succession of roots to a secant line. Quasi-Newton methods are used when the Jacobian or Hessian matrix are too expensive to compute and instead calculates its approximation predictively with preceding information. Formally it attempts to find the approximation of the second derivatives of the functional (hessian matrix) via finite difference. One of the most well-known methods used is the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm and is outlined in the following five steps that are repeated until the model converges to the solution. The initial iteration utilises  $H_a^k$  as the initial hessian matrix  $H_{BFGS}^k$ .

1. Calculate the search direction  $D^k$ , in the 1<sup>st</sup> iteration it is done by using the newton method.
2. Utilise line search to find step length to find  $\alpha^k$ , (explained in 2.3.3), so that  $\alpha^k = \arg \min F(m^k + \alpha^k D^k)$ .
3. Update model i.e.  $m^{k+1} = m^k + \alpha^k D^k$ , where model difference  $z^k$ , can be determined from model shift i.e.  $z^k = \alpha^k D^k$ , instead of normally from  $z^k = m^{k+1} - m^k$ .
4. Compute the new gradient  $\nabla f(m^{k+1})$  and derive the gradient difference as  $y^k = \nabla f(m^{k+1}) - \nabla f(m^k)$ .
5. Use equation 2.40 to obtain  $H_{BFGS}^k$  for step 1 in the next iteration.

$$H_{BFGS}^{k+1} = H_{BFGS}^k - \frac{H_{BFGS}^k z^k (z^k)^T (H_{BFGS}^k)^T}{(z^k)^T H_{BFGS}^k z^k} + \frac{y^k (y^k)^T}{(y^k)^T z^k}, \quad (2.40)$$

The BFGS algorithm may be computationally inexpensive since it requires only the approximate hessian matrix, however the feasibility of this method is limited to high performing CPU if the FWI involves a large square approximate hessian matrix as it would require significant memory to store the matrix. To solve this memory problem the limited memory-BFGS (L-BFGS) method was formulated, where the hessian matrix doesn't need to be explicitly stored, instead only the gradient and models of previous iterations are called to calculate the descent direction directly (Nocedal, 1980). In this method the inverse hessian  $B^k=1/H^k$  in equation 2.41 is calculated and applied to the gradient.

$$B^k = \left( I - \rho^{k-1} z^{k-1} (y^{k-1})^T \right) B^{k-1} \left( I - \rho^{k-1} y^{k-1} (z^{k-1})^T \right) + \rho^{k-1} z^{k-1} (z^{k-1})^T, \quad (2.41)$$

Where  $\rho$  in the case of L-BFGS is the algebraic dot product of model and gradient differences (not to be confused with the density in the case of wave equation), and the identity matrix  $I$ , is used.

$$\rho^{k-1} = \frac{1}{(y^{k-1})^T z^{k-1}}, \quad (2.42)$$

Hence instead of the hessian matrix only the difference vector  $V^k=(I-\rho^k z^k (y^k)^T)$ , of the gradient and model of previous iteration steps are stored. Equation 2.43 is derived as the recursive form of equation 2.41.

$$\begin{aligned} B^k &= (V^{k-1})^T (V^{k-2})^T \dots (V^{k-m})^T B^0 (V^{k-m}) \dots (V^{k-2}) (V^{k-1}) \\ &\quad + \rho^1 (V^{k-1})^T (V^{k-2})^T \dots (V^{k-m+1})^T y^1 (y^1)^T (V^{k-m+1}) \dots (V^{k-2}) (V^{k-1}) \\ &\quad \vdots \\ &\quad + \rho^{k-2} (V^{k-1})^T y^{k-2} (y^{k-2})^T (V^{k-1}) \\ &\quad + \rho^{k-1} y^{k-1} (y^{k-1})^T, \end{aligned} \quad (2.43)$$

In L-BFGS the user decides the number of iterations of difference vectors to be stored  $m$ , i.e. how far back in iterations is the information retrieved to attain inverse hessian matrix. It is easier to visualise this in a pseudo-code, where L-BFGS is applied as a two-loop recursion as highlighted in figure 2.1, where gradient  $g^k$ , is immediately applied to the pseudo-initial inverse hessian matrix  $B^0$ , of the iteration

$k$ , to attain search direction  $D^k$ , directly. L-BFGS can be utilised for either a minimising or maximising function. Because we are trying to minimise the misfit function, the search direction should be negative i.e.  $D^k = -g^k$  and the ‘initial’ inverse approximate hessian  $B^{k,0}$  at the FWI iteration  $k$ , must be positive definite, meaning that it’s a symmetrical, and all its pivot (first nonzero in the rows of a diagonal matrix) values are positive, hence  $B^{k,0}$  is a diagonal matrix. To ensure the initial inverse hessian is positive definite, the criteria  $(y^{k-1})^T z^{k-1} > 0$  is used.

---

**Algorithm 1:** L – BFGS method to find descent direction

---

**Input:**  $g^k$ ,  
 $q = -g^k$ ,  
**Output:**  $D^k$

---

**For**  $i = k-1, k-2, \dots, k-m$  **do**  
 $\alpha^i = \rho^i (z^i)^T q$ ,  
 $q = q - \alpha^i y^i$ ,  
**End For**

$$B^{k,0} = s^{k-1} (y^{k-1})^T \left( (y^{k-1})^T y^{k-1} \right) \approx \frac{(s^{k-1})^T y^{k-1}}{(y^{k-1})^T y^{k-1}} I,$$

$$D^k = B^{k,0} q,$$

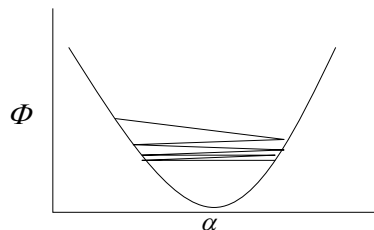
**For**  $i = k-m, k-(m-1), \dots, k-1$  **do**  
 $\beta^i = \rho^i (y^i)^T D^k$ ,  
 $D^k = D^k + z^i (\alpha^i - \beta^i)$ ,  
**End For**

---

**Figure 2.1** Pseudo-code to calculate descent direction via L-BFGS algorithm.

### 2.4.3 Step-length

A reasonable step length is needed to provide enough decrease in the function per iteration, to get it closer to the global minimum of a nonlinear function. The inexact line search method is recommended instead of the exact line search method because it is efficient for an unconstrained minimisation problem as it requires step length  $\alpha^k$ , to reduce the function only ‘sufficiently’ as opposed to an exact reduction to the global minimum, thus inexact line search is faster. However, there are two ways where step length does not sufficiently reduce the function. The first is that the step length is too large, that relatively the reduction of the function is too little, illustrated in figure 2.2.



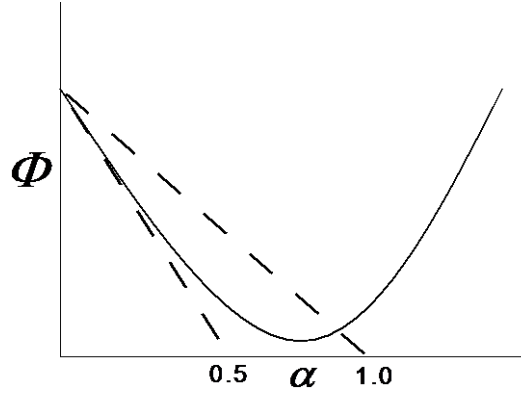


**Figure 2.2** Graph showing the reduction of the misfit  $\phi$  (vertical) along the step length  $\alpha$  (horizontal) in a minima basin. The zig-zag line indicates a significant shift in  $\alpha$  causes the function to overshoot the minimum, leading to relatively small reduction of the former compared to the change of the latter.

To avoid this problem a set of inequalities called the Wolfe conditions must be satisfied to attain an acceptable step length  $\alpha^k$ , (Wolfe, 1969). To determine if  $\alpha^k$  decreases the function sufficiently the first of the two Wolfe conditions, the Armijo rule is used, as displayed below.

$$F(m^k + \alpha^k D^k) \leq F(m^k) + c_1 \alpha^k (D^k)^T \nabla F(m^k), \quad c_1 \in (0,1) \quad (2.44)$$

This rule requires that the graph of the function on the left-hand side is either equal or lower than that of the right-hand side, so to ensure that  $\alpha^k$  is small enough that the low reduction associated with high step length is avoided. This is illustrated in figure 2.3, where for the step length to satisfy equation 2.44, it must be of a value where graph of the misfit function is below graph of the non-tangent dashed line.



**Figure 2.3** Solid curve indicates  $\Phi$ , the tangent dashed line indicates  $\Phi$  when  $\alpha$  is 0 i.e.  $c_1=0$ , and the other dashed line occurs when  $c_1$  has a value of 0.5, the latter is an arbitrary value to demonstrate the Armijo condition which requires that  $\alpha$  is between 0.5 and 1.

The second rule known as the curvature condition is needed to prevent  $\alpha^k$  from being too small and is introduced below, which ensures sufficient reduction of the slope of the function.

$$-(D^k)^T \nabla F(m^k + \alpha^k D^k) \leq -c_2 (D^k)^T \nabla F(m^k), \quad (2.45)$$

Combination of 2.44 and 2.45 gives us the weak Wolfe condition, which may result in a step length value that is not close to the functions ( $\phi$ ) minimiser (critical point). To force  $\alpha^k$  to be as close as possible to the critical point of  $\phi$ , the strong Wolfe condition is used where it differs in that it applies curvature condition only to the magnitudes of both sides of equation 2.45.

$$\left| (D^k)^T \nabla F(m^k + \alpha^k D^k) \right| \leq c_2 \left| (D^k)^T \nabla F(m^k) \right|, \quad (2.46)$$

Both conditions have their uses, the weak Wolfe condition is utilised for non-smooth optimisation whereas the strong Wolfe condition is used for smooth optimisation.

---

**Algorithm 2:** Line Search Algorithm

---

**Input:**  $\alpha_i > 0$  and  $\alpha_{\max}$   
 $\alpha_0 = 0$ ,  $i \leftarrow 1$ ,  
**Output:**  $\alpha$

---

```

1  While true do
2      Evaluate  $\phi(\alpha_i)$ 
3      If  $[\phi(\alpha_i) > \phi(0) + c_1 \alpha_i \partial \phi(0)]$  or  $[\phi(\alpha_i) > \phi(\alpha_{i-1}) \text{ \& } i > 1]$ 
4           $\alpha \leftarrow \text{zoom}(\alpha_{i-1}, \alpha_i)$ ,
5          Return  $\alpha$ 
6      End if
7      Evaluate  $\partial \phi(\alpha_i)$ 
8      If  $|\partial \phi(\alpha_i)| \leq -c_2 \partial \phi(0)$ ,
9          Return  $\alpha$ 
10     Else if  $\partial \phi(\alpha_i) \geq 0$ ,
11          $\alpha \leftarrow \text{zoom}(\alpha_{i-1}, \alpha_i)$ ,
12         Return  $\alpha$ 
13     Else
14         Choose  $\alpha_{i+1}$  such that  $\alpha_i < \alpha_{i+1} < \alpha_{\max}$ 
15     End if
16      $i \leftarrow i + 1$ ,
17 End while
```

---

**Figure 2.4** Pseudo-code of line search algorithm used to determine the step length.

---

**Algorithm 3:** Zoom function for line search algorithm

---

**Input:**  $\alpha_{\text{low}}, \alpha_{\text{high}}$   
 $j \leftarrow 0$ ,  
**Output:**  $\alpha$

---

```

1  While true do
2      Find  $\alpha_j$  via  $\alpha_j = \frac{\alpha_{\text{low}} + \alpha_{\text{high}}}{2}$ ,
3      Evaluate  $\phi(\alpha_j)$ 
4      If  $[\phi(\alpha_j) > \phi(0) + c_1 \alpha_j \partial \phi(0)]$  or  $[\phi(\alpha_j) > \phi(\alpha_{\text{low}})]$ 
5           $\alpha_{\text{high}} \leftarrow \alpha_j$ ,
6      Else
7          Evaluate  $\partial \phi(\alpha_j)$ 
8          If  $|\partial \phi(\alpha_j)| \leq -c_2 \partial \phi(0)$ ,
9               $\alpha = \alpha_j$ ,
10             Return  $\alpha$ 
11          Else if  $\partial \phi(\alpha_j) * (\alpha_{\text{high}} - \alpha_{\text{low}}) \geq 0$ 
12               $\alpha_{\text{high}} \leftarrow \alpha_{\text{low}}$ ,
13          End if
14           $\alpha_{\text{low}} \leftarrow \alpha_j$ ,
15      End if
16       $j \leftarrow j + 1$ 
17 End while
```

---

**Figure 2.5** Pseudo-code for zoom function via the bisection interpolation.

## 2.5 Summary

This chapter investigates the multiple objective functions that are of vital importance in FWI, primarily in the frequency domain. It looks specifically into the mathematics that are central to each misfit function and the thesis' original objective function takes into consideration the strengths of the reviewed misfit functions. The most robust objective function considers both waveform amplitude envelope and phase to fully satisfy the performance of full waveform inversion. Deconvolution is utilised to provide a strict confinement of energy peaks when both synthetic and observed seismogram matches, this is applied to the square of the envelope function in order to improve signal-to-noise ratio of enveloping wavelets of both data which can help retrieve low frequency signal information from the seismograms. The phase difference is utilised to reduce cycle-skipping and is heavily emphasized in the initial stages of the FWI, with the co-efficient  $w$  starting at 1 and gradually diminishes as the iterations progresses to shift the focus towards amplitude matching. This chapter briefly covers regularisation which are used to treat the non-linearity of FWI, by applying constraints in the objective functions that limits its variation using the a priori information of the model being inverted for. The latter part of this chapter details the adjoint state methodology (Plessix, 2006) needed to attain the gradient, and several optimisation algorithm for the search direction used to update the model, followed by a sub-chapter detailing the conditions of which the step length is chosen in order to improve convergence rates.

# Chapter 3: Forward Modelling

## 3.1 Forward Modelling Introduction

Forward modelling is crucial for a recreation of the seismic data otherwise known as the synthetic data, which involves simulating a propagating seismic wave going through the computational model of the physical properties of the media. This method attempts to simulate behavioural response of waves on the solid structure or the fluid, such as carbonate reservoirs, throughout the media (Hodgetts et al, 2000). Consequently, forward modelling simulates seismic acquisition where the generated traces in synthetic data are sorted as shot gathers like that of the observed seismic data. This simulation would involve the use of wave equations, which can determine the accuracy of the physical model, where the disparity between the synthetic data generated from forward modelling using the physical model and the observed data are compared in the misfit function for FWI. Wave equation can be applied in 2D or 3D depending on the degree of freedom of the wave equation used, but for the purpose of simple explanation this thesis describes the 2D formulation of the wave equation.

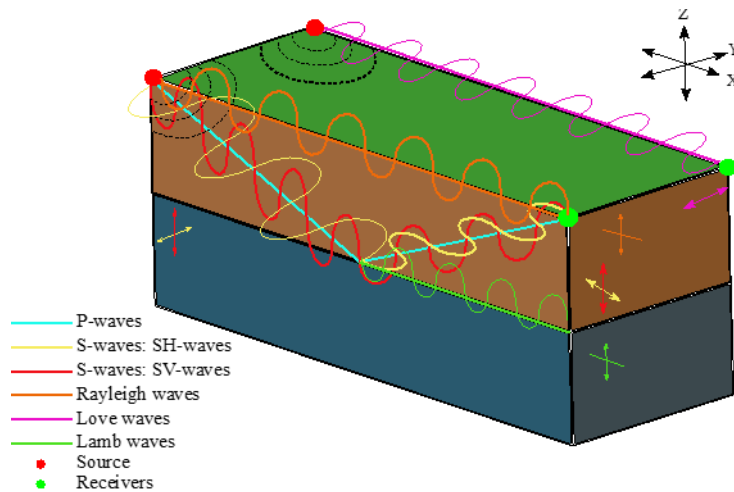
The formulation of synthetic data is done by taking the data points within the wavefield where the receivers are located at either in a time step for time domain or temporal frequency in frequency domain. However, the disparity for computationally simulating a propagating wave is that there are boundaries to the set model, but these boundaries do not exist in the practical seismic acquisition of the observed data, hence an absorbing boundary condition and layer would be required to apply to the wave equation at the periphery of the model of the medium. The free surface boundary condition is applied computationally to represent the physics behaviour of the wave when it reaches medium above the surveyed area e.g. air for land seismic data, and water for ocean bottom seismic data. Another component that is crucial to forward modelling is the wavelet that introduces the oscillation of which causes the wave to travel through the media. Seismic wavelets are conventionally designed with specific properties that accurately displays the reaction of the medium to the wave propagating through it. The last step is to describe how to extrapolate from the field to generate seismic data.

## 3.2 Waves

The fundamental principle of forward modelling is that the wave is a disturbance of the field, where the values of the field oscillate continually around a stable equilibrium, these are described by wave equations. In geophysics the waves described are seismic waves which are mechanical waves that describe transfer of energy by oscillation throughout the earth. Seismic waves are an important feature when generating synthetic data because it describes the vibrational reaction of the medium to the waves traveling through it. There are more than one type of seismic wave and of course the more waves

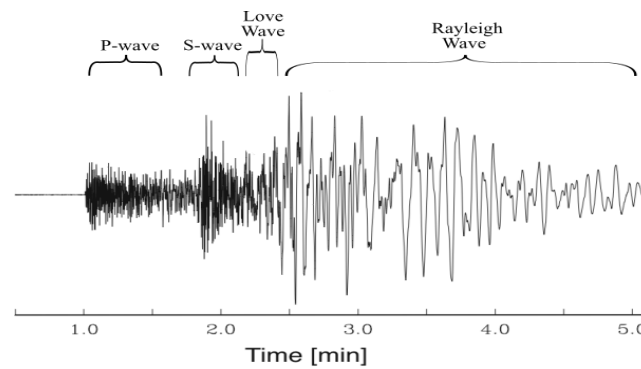
included the more detailed and accurate the synthetic data would be. Seismic waves are categorised into two groups, body waves where it traverses throughout the interior of the Earth (the model), and surface waves where they travel along the surface of the model. Body waves are typically affected by the material's physical properties, such as density and their velocities which can provide an insightful detail into the composition and phase of the material that is traversed by the waves.

There are typically two types of body waves determined by the particle motions of the material, first of which is the primary waves (P-waves) because they would usually be the first to arrive at the receivers due to their high velocity, and displacement of the medium is in the same direction of wave propagation. The second body wave are the S-waves which involve particle displacement at the two perpendicular directions to wave travel, this is illustrated in Figure 3.1 where displacement in SH-waves are in the  $x$ - $y$  dimensions whereas for SV-waves they are in the  $x$ - $z$  dimensions.



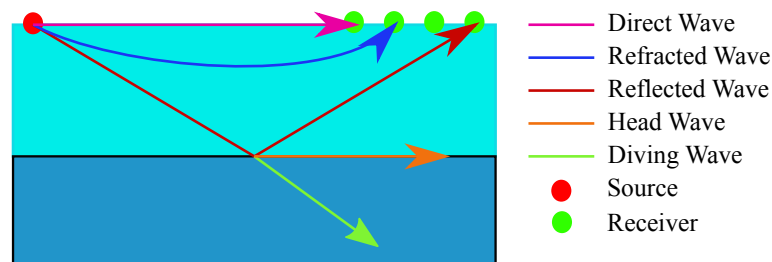
**Figure 3.1** 3D model illustrating wave propagations. Pointers with arrowheads indicate direction of oscillation.

When the waves reach the receivers, it could be clearly noted that P-waves arrive first due to their high velocity and that S-wave would arrive afterwards with greater amplitude. Both surface waves would arrive last but with Love waves slightly faster than Rayleigh waves, but their amplitudes are comparable (Tanimoto et al, 2015).

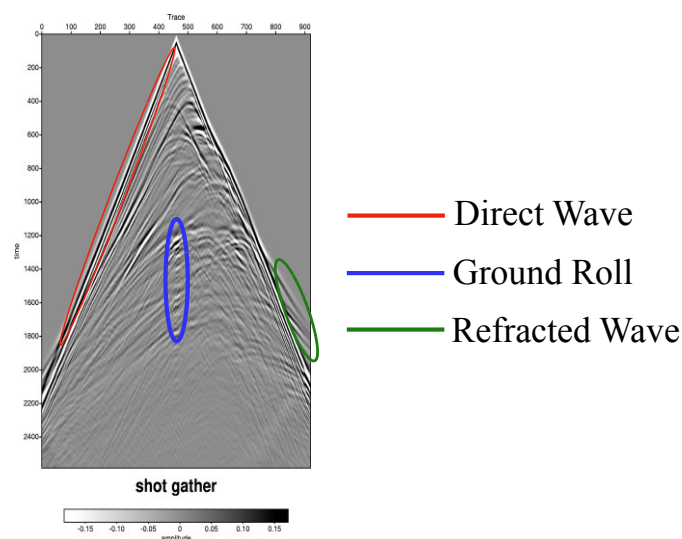


**Figure 3.2** A seismogram containing the amplitude and phase information from body and surface waves. Seismogram from Lecture notes of University of California San Diego (2018).

In addition to the displacement of the medium, the travel path of the waves propagating throughout the medium can also affect seismic data, usually results in additional events in the gather of the seismic data (group of seismic traces). Most notable of which is the direct waves which are waves that travels directly from the source to receivers along a straight line above the paths of the reflected seismic data.



**Figure 3.3** Travel path of propagating waves.



**Figure 3.4** Seismic data with the different waves recorded being indicated (generated by the author).

### 3.2.1 Elastic wave equation

The wave equations used are decided upon by the a priori knowledge of the media, which assumes how the material behaves when the wave travels through its body. The most common denoted wave equation due to material behaviour is the elastic wave equation, that is when a wave travels the material is expected to deform elastically, which means the material returns to its original shape and form as soon as pressure of the wave is removed. To demonstrate the linear algebra formulation of the forward modelling wave equation (equation 2.25), a characterised elastic wave equation for homogeneous isotropic media was used and rearranged (Aki and Richards, 2002).

$$-\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} + (\lambda + 2\mu) \nabla (\nabla \cdot \mathbf{u}) - \mu \nabla \times \nabla \times \mathbf{u} = -\mathbf{f}, \quad (3.1)$$

Where the non-bold operator on the L.H.S. of equation 3.1 denote the operators in the sparse matrix  $\mathbf{S}$ , upon the wavefield vector  $\mathbf{u}$ , resulting in the external body force vector  $\mathbf{f}$ . The operators included are the second order derivative of  $\mathbf{u}$  with respect to time  $t$ , and the gradient  $\nabla(\cdot)$ , divergence  $\nabla \cdot$ , and curl  $\nabla \times$ , describes the physical displacements of  $\mathbf{u}$ . Whenever  $\mathbf{f}$  is applied on an the medium, the propagating waves are influenced by the mediums' material-dependent physical properties including the Lamé first parameter  $\lambda$ , shear modulus  $\mu$ , and density  $\rho$ . The velocity-stress formulation of the elastic wave equation was first solved with the finite-difference method in time domain (Alterman and Karal, 1968), and adapted for layered media (Kelly et al., 1976), to generate a synthetic seismogram for P-SV wave propagation (Virieux, 1986), while its derivation in the frequency domain was initially used for cross-hole imaging (Pratt, 1990). In 2 dimensions this formulation consists of 5 formulas, the first two of which is the first order force balance, and the last three are the constitutive equations (read Alterman and Karal, 1968 for source information). Since solving each of the 5 formulas is computationally expensive and complicated to model, it can instead be expanded and summed to a second order time derivative of the vector displacement formula for easier modelling. The second order time derivative is derived by the insertion of the three stress-displacement formulas into the first two velocity-stress formulas in 2D.

$$\begin{aligned} \rho \frac{\partial^2 u}{\partial t^2} &= \frac{\partial}{\partial x} \left( \frac{\partial \sigma_{xx}}{\partial t} \right) + \frac{\partial}{\partial z} \left( \frac{\partial \sigma_{xz}}{\partial t} \right) \rightarrow \rho \frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial x} \left( (\lambda + 2\mu) \frac{\partial u}{\partial x} + \lambda \frac{\partial v}{\partial z} \right) + \frac{\partial}{\partial z} \left( \mu \left( \frac{\partial v}{\partial x} + \frac{\partial u}{\partial z} \right) \right), \\ \rho \frac{\partial^2 v}{\partial t^2} &= \frac{\partial}{\partial z} \left( \frac{\partial \sigma_{zz}}{\partial t} \right) + \frac{\partial}{\partial x} \left( \frac{\partial \sigma_{xz}}{\partial t} \right) \rightarrow \rho \frac{\partial^2 v}{\partial t^2} = \frac{\partial}{\partial z} \left( (\lambda + 2\mu) \frac{\partial v}{\partial z} + \lambda \frac{\partial u}{\partial x} \right) + \frac{\partial}{\partial x} \left( \mu \left( \frac{\partial v}{\partial x} + \frac{\partial u}{\partial z} \right) \right), \end{aligned} \quad (3.2)$$

The two formulas that the elastic wave equation splits into corresponds to the horizontal displacement  $u$ , and vertical displacement  $v$ , of the wavefield. The time derivative of the stress tensors  $\sigma_{xx}$ ,  $\sigma_{zz}$ , and  $\sigma_{xz}$ , can be expanded to the directional derivatives of displacement to the horizontal  $x$ , and vertical  $z$ , directions.

If the subsurface medium is homogeneous i.e. it has uniform physical properties throughout its body, and the Lamé parameters can be taken out of the secondary first-order derivative, this is because the elastic moduli is independent of the position of the stress displacement.

$$\begin{aligned}\rho \frac{\partial^2 u}{\partial t^2} &= (\lambda + 2\mu) \frac{\partial^2 u}{\partial x^2} + (\lambda + \mu) \frac{\partial^2 v}{\partial x \partial z} + \mu \frac{\partial^2 u}{\partial z^2}, \\ \rho \frac{\partial^2 v}{\partial t^2} &= (\lambda + 2\mu) \frac{\partial^2 v}{\partial z^2} + (\lambda + \mu) \frac{\partial^2 u}{\partial x \partial z} + \mu \frac{\partial^2 v}{\partial x^2},\end{aligned}\quad (3.3)$$

By shifting  $\rho$  in equation 3.3 to the right-hand side, the Lamé parameters are substituted with the squared velocities of the two types of elastic body waves,  $V_p$  the speed of P-waves and  $V_s$  the speed of S-wave.

$$V_p = \sqrt{\frac{\lambda + 2\mu}{\rho}} \quad V_s = \sqrt{\frac{\mu}{\rho}} \quad V_p^2 - V_s^2 = \frac{\lambda + \mu}{\rho}, \quad (3.4a)$$

The co-efficient velocities of second order spatial derivation is corresponds to directions of displacement vectors such as for the horizontal displacement vector the squared velocity of P-waves amplifies the second order horizontal spatial derivative. The resulting equation is the elastic wave equation in the homogeneous media.

$$\begin{aligned}\frac{\partial^2 u}{\partial t^2} &= V_p^2 \frac{\partial^2 u}{\partial x^2} + (V_p^2 - V_s^2) \frac{\partial^2 v}{\partial x \partial z} + V_s^2 \frac{\partial^2 u}{\partial z^2}, \\ \frac{\partial^2 v}{\partial t^2} &= V_p^2 \frac{\partial^2 v}{\partial z^2} + (V_p^2 - V_s^2) \frac{\partial^2 u}{\partial x \partial z} + V_s^2 \frac{\partial^2 v}{\partial x^2},\end{aligned}\quad (3.4b)$$

If the subsurface media is heterogeneous i.e. inhomogeneous, the product rule is applied to the second derivative resulting in the spatial derivative of the Lamé parameters, this is because the position of stress displacement effects the elastic moduli.

$$\begin{aligned}\rho \frac{\partial^2 u}{\partial t^2} &= (\lambda + 2\mu) \frac{\partial^2 u}{\partial x^2} + (\mu) \frac{\partial^2 u}{\partial z^2} + (\lambda + \mu) \frac{\partial^2 v}{\partial z \partial x} + \frac{\partial(\lambda + 2\mu)}{\partial x} \frac{\partial u}{\partial x} + \frac{\partial \mu}{\partial z} \frac{\partial v}{\partial x} + \frac{\partial \lambda}{\partial x} \frac{\partial v}{\partial z} + \frac{\partial \mu}{\partial z} \frac{\partial u}{\partial z}, \\ \rho \frac{\partial^2 v}{\partial t^2} &= (\lambda + 2\mu) \frac{\partial^2 v}{\partial z^2} + (\mu) \frac{\partial^2 v}{\partial x^2} + (\lambda + \mu) \frac{\partial^2 u}{\partial x \partial z} + \frac{\partial(\lambda + 2\mu)}{\partial z} \frac{\partial v}{\partial z} + \frac{\partial \lambda}{\partial z} \frac{\partial u}{\partial x} + \frac{\partial \mu}{\partial x} \frac{\partial v}{\partial x} + \frac{\partial \mu}{\partial x} \frac{\partial u}{\partial z},\end{aligned}\quad (3.5)$$

Generally, the starting models utilised in forward modelling and FWI are the velocity and density model, thus the values of these models can substitute the Lamé parameters for easier calculation.



$$\rho V_p^2 = \lambda + 2\mu \quad \rho V_s^2 = \mu \quad \rho(V_p^2 - 2V_s^2) = \lambda, \quad (3.6a)$$

$$\begin{aligned} \frac{\partial^2 u}{\partial t^2} &= V_p^2 \frac{\partial^2 u}{\partial x^2} + (V_p^2 - 2V_s^2) \frac{\partial^2 v}{\partial x \partial z} + V_s^2 \left( \frac{\partial^2 v}{\partial z \partial x} + \frac{\partial^2 u}{\partial z^2} \right) \\ &\quad + \frac{1}{\rho} \left( \frac{\partial(\rho V_p^2)}{\partial x} \frac{\partial u}{\partial x} + \frac{\partial(\rho(V_p^2 - 2V_s^2))}{\partial x} \frac{\partial v}{\partial z} + \frac{\partial(\rho V_s^2)}{\partial z} \frac{\partial v}{\partial x} + \frac{\partial(\rho V_s^2)}{\partial z} \frac{\partial u}{\partial z} \right), \\ \frac{\partial^2 v}{\partial t^2} &= V_p^2 \frac{\partial^2 v}{\partial z^2} + (V_p^2 - 2V_s^2) \frac{\partial^2 u}{\partial z \partial x} + V_s^2 \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 u}{\partial x \partial z} \right) \\ &\quad + \frac{1}{\rho} \left( \frac{\partial(\rho V_p^2)}{\partial z} \frac{\partial v}{\partial z} + \frac{\partial(\rho(V_p^2 - 2V_s^2))}{\partial z} \frac{\partial u}{\partial x} + \frac{\partial(\rho V_s^2)}{\partial x} \frac{\partial v}{\partial x} + \frac{\partial(\rho V_s^2)}{\partial x} \frac{\partial u}{\partial z} \right), \end{aligned} \quad (3.6b)$$

### 3.2.2 Acoustic wave equation

Materials that are able to deform elastically are generally considered to have some form of solidity, however some seismic acquisitions involve a layer of fluid, such as ocean-bottom-seismic (OBS) data, that if elastic wave equation is applied it would give erroneous data. This is because fluids cannot support shear stress due to the non-fixed positions of fluid particles, as they do not have the same rigid bodies as that of solids. Therefore, shear terms in the wave equation such as  $\mu$  and  $V_s$  co-efficient would be omitted, and only compressional stress is supported which replaces the displacement wavefield with pressure field  $P$ . Combining the two formulas of equation 3.4b and omitting terms containing shear properties would yield the acoustic wave equation within the homogeneous media.

$$\frac{\partial^2 P}{\partial t^2} = V_p^2 \left( \frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial z^2} \right), \quad (3.7)$$

Combining the two formulas in Equation 3.6b omitting terms containing shear properties would yield acoustic wave equation for pressure displacement in an inhomogeneous media. It should be noted that despite the simplicity and effectiveness of the acoustic wave equation in generating synthetic data, it would leave out information involving elastic properties within the sub-surface media that would otherwise give more accurate image of the structure.

$$\rho \frac{\partial^2 P}{\partial t^2} = \rho V_p^2 \left( \frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial z^2} \right) + \frac{\partial(V_p^2 \rho)}{\partial x} \frac{\partial P}{\partial x} + \frac{\partial(V_p^2 \rho)}{\partial z} \frac{\partial P}{\partial z}, \quad (3.8)$$

### 3.2.3 Viscoelastic wave equations

However, these equations only consider the resistance to deformation by largely solid materials, whereas hydrocarbons are fluid by nature (e.g., natural gas and crude oil), hence, to generate synthetic data that contain information regarding fluids in the subsurface regions the viscosity is considered. Viscosity  $\eta$ , is the measurement of a fluid's resistance to deformation at a given time rate and describes the attenuative properties of the wave model. S-waves attenuate more in fluids, and they typically have higher amplitude compared to that of the P-waves. Thus, viscoelastic wave equation is used to describe the viscous and elastic reaction of subsurface materials to propagating waves. Viscoelastic wave equation was originally conceived for medical ultrasound imaging of soft biological tissues (Charlier and Crowet, 1985), but in geophysical exploration the elastic components and viscous components describes the storage and dissipation of the waves' energy, respectively.

Viscoelastic media are described either as linear or non-linear, the former indicates the fluid is Newtonian where the stress applied is linearly proportional to the strain rate due to its viscosity being constant. Concurrently non-linear viscoelasticity indicates a non-Newtonian fluid where its viscosity changes when force is applied, and so exhibits a more solid or fluid nature, thus there is no linear relationship between stress and strain. Fluids that are Newtonian include air, water, light hydrocarbon oils and natural gases, but most heavy crude oils are non-Newtonian fluids and rocks are also not fully elastic. It is generally very difficult to perform forward modelling of wave equation in non-linear viscoelastic media, as there are several factors that contributes to the complexity, such as the wave equation requiring different treatments in the presence of fluids and solids (Mahieux, 2006). Research conducted to address the issue of forward modelling with non-linear viscoelastic media includes using blow-up solutions (Liu et al., 2018, Song and Zhong, 2010), Galerkin method (Soufyane and Feng, 2020), Faedo-Galerkin method (Alodhaibi and Zennir, 2020), and via auxiliary differential equation method (Martin et al., 2019).

To describe the attenuative properties of a viscous material the  $Q$ -factor is used which gives information regarding attenuation of body (P- and S-) waves due to anelasticity. Unfortunately, it is also difficult to link  $Q$ -factor to viscosity, a two-step solution was proposed where the  $Q$ -estimation method is used to produce a  $Q$  tomography model followed by the Biot/Squirt flow (BISQ) model to address the missing link in the case of pore fluids (Lines and Vashegani, 2012).

The BISQ model combines Biot theory (1956a,b) and the squirt-flow mechanism (Dvorkin and Nur, 1993) where the former describes macroscopic fluid flow due to macroscopic rock properties, and the latter describes fluids squirting out of pores due to deformation caused by compressional waves. Unfortunately, the BISQ model was found to be unstable due to the time-exponentially exploding solutions, that occurs when the squirt-flow co-efficient is negative or has a non-zero imaginary part, this seems to be always the case at low-angular frequencies (Liu and Yong, 2016). It is suggested that

further research should be done either to build upon the BISQ model or extend the Biot's theory to address this issue.

Regardless, the study of  $Q$ -estimation is wide and varied, and is outside the scope of this thesis, the reader is advised to read the book by (Wang, 2008), instead a simplified definition of  $Q$  is provided.

$$Q = 2\pi \frac{\text{Energy of seismic wave}}{\text{Loss of energy per wave cycle}} = 2\pi \frac{e}{\delta e}, \quad (3.9)$$

With the availability of viscosity derived from  $Q$ -estimation, the constitutive models of linear viscoelasticity could be used to establish the viscoelastic wave equation. As mentioned earlier the viscoelastic model is composed of the elastic component  $\sigma_e$ , which follows stress-strain interpretation of the Hooke's Law, and the viscous component  $\sigma_v$ , which follows fluids resistance to stress rate, which together form the overall  $\sigma$ , which comprises of spatial derivatives (strain)  $\varepsilon_s$ , and elastic modulus  $E_m$ .

$$\sigma_e = E_m \varepsilon_s, \quad \sigma_v = \eta \frac{\partial \varepsilon_s}{\partial t}, \quad (3.10)$$

First of the viscoelastic models is the Maxwell model established in 1867, where elastic deformation occurs coincidentally with the stress applied to the media, thereby elastic deformation ceases with removal of stress. This model assumes viscous deformation would continue to grow if stress is applied, which is true in the presence of heat but is inaccurate in geophysics exploration of settled hydrocarbon reservoirs at lower temperatures. In the Maxwell model the stress for both components is the same, but the overall strain is the sum of viscous and elastic strain. The overhead dot denotes time derivative.

$$\sigma + \frac{\eta}{E_m} \dot{\sigma} = \eta \dot{\varepsilon}, \quad (3.11)$$

Whilst it is complex to model wave propagation through the maxwell viscoelastic model in time domain (equation 3.12a), it is more straightforward to model this in frequency domain (equation 3.12b-d), because time derivative is Fourier transformed to imaginary angular frequency,  $i\omega$ , where  $\kappa_v = E/\eta$ .

$$\frac{E}{\eta} \sigma + \frac{\partial \sigma}{\partial t} = E \frac{\partial \varepsilon}{\partial t}, \quad (3.12a)$$

$$(\kappa_v + i\omega) \sigma = (i\omega) E \varepsilon, \quad (3.12b)$$

$$\sigma = \frac{(i\omega)}{(\kappa_v + i\omega)} E \varepsilon, \quad (3.12c)$$

$$\sigma = \frac{(i\omega)}{(\kappa_v + i\omega)} E \varepsilon \times \frac{(\kappa_v - i\omega)}{(\kappa_v - i\omega)} = \left[ \frac{\omega^2}{(\kappa_v^2 + \omega^2)} + \frac{\kappa_v (i\omega)}{(\kappa_v^2 + \omega^2)} \right] E \varepsilon, \quad (3.12d)$$

In time domain the three constitutive formula in the velocity-stress equation is expanded with the addition of the integral of the spatial derivative over the time of the wave propagation. The time domain time integral is conventionally attained by first performing an elastic wave equation with finite difference and sum the deformation throughout each time step, and then perform a second wave propagation this time with the viscoelastic term. The problem with this method is that additional computational memory storage is required to store the time integral and longer wave simulation time due to two wave propagations are needed instead of one. It is simple in frequency domain as the time derivative in equation 3.11 becomes a single imaginary term.

$$\begin{aligned}
\frac{\partial \sigma_{xx}}{\partial t} &= \left( \frac{\omega^2 + \frac{\lambda + 2\mu}{\eta}(i\omega)}{\left(\frac{\lambda + 2\mu}{\eta}\right)^2 + \omega^2} \right) (\lambda + 2\mu) \frac{\partial u}{\partial x} + \left( \frac{\omega^2 + \frac{\lambda}{\eta}(i\omega)}{\left(\frac{\lambda}{\eta}\right)^2 + \omega^2} \right) \lambda \frac{\partial v}{\partial z}, \\
\frac{\partial \sigma_{zz}}{\partial t} &= \left( \frac{\omega^2 + \frac{\lambda + 2\mu}{\eta}(i\omega)}{\left(\frac{\lambda + 2\mu}{\eta}\right)^2 + \omega^2} \right) (\lambda + 2\mu) \frac{\partial v}{\partial z} + \left( \frac{\omega^2 + \frac{\lambda}{\eta}(i\omega)}{\left(\frac{\lambda}{\eta}\right)^2 + \omega^2} \right) \lambda \frac{\partial u}{\partial x}, \\
\frac{\partial \sigma_{xz}}{\partial t} &= \left( \frac{\omega^2 + \left(\frac{\mu}{\eta}\right)(i\omega)}{\left(\frac{\mu}{\eta}\right)^2 + \omega^2} \right) \mu \left( \frac{\partial v}{\partial x} + \frac{\partial u}{\partial z} \right),
\end{aligned} \tag{3.13}$$

The Kelvin-Voigt/Voigt model considers the viscoelastic deformation to be reversible and deformation rate decreases with constant stress, a typical characteristic behaviour of soft solids including organic polymers and sediments, but not that of stronger materials like rocks and metals. Voigt models have the opposite stress-strain relationship to that of the Maxwell model, its' overall strain is identical for both viscous and elastic component, but its' overall stress is the sum of viscous and elastic stresses.

Voigt model is more suited to seismic wave propagation, especially considering the typical depth of the subsurface region containing hydrocarbon reservoirs and the lithology, because it indicates that both viscous and elastic component undergo deformation when stress is applied, whereas the Maxwell model assumes sequential deformation with firstly the elastic strain followed by the viscous strain.

$$\sigma = E_m \varepsilon_s + \eta \dot{\varepsilon}_s, \tag{3.14}$$

The advantage of this model is that it is the most straightforward to implement into the stress-strain time domain because it does not require a time integral field to model displacement and applying the time derivative of the strain only requires the previous field of displacement. Equation 3.14 is simpler to implement in the frequency domain than that of equation 3.11.

### 3.2.4 Isotropy and anisotropy

Isotropy is the quality of the medium to have the same physical properties in all directions of the wave travelling through it, whereas anisotropy considers that the medium material would have different properties depending on the direction of wave travel. The most basic interpretation of anisotropy is the transverse isotropy, which would explicitly state that one plane encountered by the propagating wave in the medium is considered isotropic whereas the other plane is anisotropic. The first example is the Vertical Transverse Isotropy (VTI) whereby it has a vertical axis of symmetry, this means that properties of the medium is the same when waves travel through a specific point horizontally and is commonly representative of media with layering and shale (Huang et al., 2011, Bloor et al., 2013, Xu et al., 2015). Second example is the Horizontal Transverse Isotropy (HTI) which has a horizontal axis of symmetry, meaning a specific point in the medium would have the same physical property if the wave travels through it vertically, and is commonly associated with cracks and fractures caused by regional stress (Bale, 1999, Andrey et al., 2000, Bale and Margrave., 2005, Yang et al., 2015). However, anisotropy is not as clear cut as that of purely horizontally and vertically isotropic, but could be at an angle, known as Tilted Transversely Isotropy (TTI), which involves the use of trigonometric functions to distinguish the physical properties that would be identified at certain angles of a structure within the subsurface medium such as shale layering resting on dipping salt flanks, (Zhang et al., 2011). Unfortunately, due to the large scope of the subject, this sub-chapter is written as a passing mention as to a possible future expansion to research.

### 3.2.5 Generalised viscoelastic wave equation in porous media.

The author of this thesis proposes a wave equation propagating throughout a homogeneous sub-surface media based on the combination of poro-elastic and visco-elastic properties. The motivation is to provide a more comprehensive synthetic data that contains information regarding the medias' viscoelastic and its poro-elastic properties, as well as to scale between these two properties, for the purpose of more accurate data matching in FWI. The interpolation co-efficient  $\beta_v$ , and porosity  $\phi$ , is used to determine the extent of the effects of the two properties.

As this thesis proposes the FWI in frequency domain, it would require Fourier transform  $FT\{\}$ , to be applied to the time derivatives in time domain wave equation, where time-based terms becomes term based on the angular temporal frequency  $\omega=2\pi f$ , which is dependent upon the temporal frequency  $f$ .

$$FT\left\{\frac{\partial}{\partial t}\right\}=i\omega, \quad (3.15)$$

The imaginary unit  $i$ , has the equivalent value of  $\sqrt{-1}$ , Fourier transform of second order time derivative yields the square of the imaginary unit and would result in the value of  $-\omega^2$ , it should be noted that

Fourier transform does not affect the spatial derivatives as it decomposes the time function into functions of its constituent frequencies.

The proposed wave equation utilises the generalised visco-elastic wave equation by Wang, 2016 where the Kelvin-voigt visco-elastic constitutive relation of equation 3.14 is applied into the velocity-stress wave equation 3.2, resulting in the frequency domain equation 3.16, and  $\beta_v$  describes the behavioural gradient of the medium as being fully elastic at 0, to being fully viscoelastic at 1.

$$\begin{aligned} -\rho\omega^2(u) &= \left((\lambda + 2\mu) + \beta_v(\eta^{\beta_v} i\omega^{\beta_v})\right) \frac{\partial^2 u}{\partial x^2} + \left(\mu + \beta_v(\eta^{\beta_v} i\omega^{\beta_v})\right) \frac{\partial^2 u}{\partial z^2} + \left((\lambda + \mu) + \beta_v(\eta^{\beta_v} i\omega^{\beta_v})\right) \frac{\partial^2 v}{\partial x \partial z}, \\ -\rho\omega^2(v) &= \left((\lambda + 2\mu) + \beta_v(\eta^{\beta_v} i\omega^{\beta_v})\right) \frac{\partial^2 v}{\partial z^2} + \left(\mu + \beta_v(\eta^{\beta_v} i\omega^{\beta_v})\right) \frac{\partial^2 v}{\partial x^2} + \left((\lambda + \mu) + \beta_v(\eta^{\beta_v} i\omega^{\beta_v})\right) \frac{\partial^2 u}{\partial x \partial z}, \end{aligned} \quad (3.16)$$

To simplify later equations, three complex viscoelastic coefficient terms  $A_v$ ,  $B_v$ , and  $C_v$  representing directions of viscoelastic effect that is parallel, perpendicular, and the mixed partial spatial derivative to the displacements, respectively.

$$\begin{aligned} A_v &= (\lambda + 2\mu) + \beta_v(\eta^{\beta_v} i\omega^{\beta_v}), \\ B_v &= \mu + \beta_v(\eta^{\beta_v} i\omega^{\beta_v}), \\ C_v &= (\lambda + \mu) + \beta_v(\eta^{\beta_v} i\omega^{\beta_v}), \end{aligned} \quad (3.17)$$

The poro-elastic wave equation utilised takes inspiration upon the work of Yang and Mao, 2017, who derived their isotropic poroelastodynamic equations from the stress-strain relations in anisotropic poroelastic media by Carcione, 1996. Where equations 3.18a and 3.18b represents the dynamic equation of wave propagating throughout a porous media and equations 3.18c and 3.18d represents the generalised Darcy's law which describes wave induced fluid flow through a porous media, Biot, 1962a.

$$\begin{aligned} &\left((\lambda + 2\mu + \alpha_\phi(\alpha_\phi - \phi)E_m) \frac{\partial^2}{\partial x^2} + (\mu) \frac{\partial^2}{\partial z^2} + \omega^2 \rho_1\right) u_s + \left((\lambda + \mu + \alpha_\phi(\alpha_\phi - \phi)E_m) \frac{\partial^2}{\partial x \partial z}\right) v_s \\ &+ \left(\alpha_\phi \phi E_m \frac{\partial^2}{\partial x^2} + \omega^2 \rho_2\right) u_f + \left(\alpha_\phi \phi E_m \frac{\partial^2}{\partial x \partial z}\right) v_f = 0, \end{aligned} \quad (3.18a)$$

$$\begin{aligned} &\left((\lambda + 2\mu + \alpha_\phi(\alpha_\phi - \phi)E_m) \frac{\partial^2}{\partial x^2} + (\mu) \frac{\partial^2}{\partial z^2} + \omega^2 \rho_1\right) v_s + \left((\lambda + \mu + \alpha_\phi(\alpha_\phi - \phi)E_m) \frac{\partial^2}{\partial x \partial z}\right) u_s \\ &+ \left(\alpha_\phi \phi E_m \frac{\partial^2}{\partial x^2} + \omega^2 \rho_2\right) v_f + \left(\alpha_\phi \phi E_m \frac{\partial^2}{\partial x \partial z}\right) u_f = 0, \end{aligned} \quad (3.18b)$$

$$\begin{aligned}
F_u = & \left( (\alpha_\phi - \phi) E_m \frac{\partial^2}{\partial x^2} + \omega^2 (\rho_f - \phi \tilde{\rho}) \right) u_s + \left( (\alpha_\phi - \phi) E_m \frac{\partial^2}{\partial x \partial z} \right) v_s \\
& + \left( \phi E_m \frac{\partial^2}{\partial x^2} + \phi \tilde{\rho} \omega^2 \right) u_f + \left( \phi E_m \frac{\partial^2}{\partial x \partial z} \right) v_f = 0,
\end{aligned} \tag{3.18c}$$

$$\begin{aligned}
F_v = & \left( (\alpha_\phi - \phi) E_m \frac{\partial^2}{\partial x^2} + \omega^2 (\rho_f - \phi \tilde{\rho}) \right) v_s + \left( (\alpha_\phi - \phi) E_m \frac{\partial^2}{\partial x \partial z} \right) u_s \\
& + \left( \phi E_m \frac{\partial^2}{\partial x^2} + \phi \tilde{\rho} \omega^2 \right) v_f + \left( \phi E_m \frac{\partial^2}{\partial x \partial z} \right) u_f = 0,
\end{aligned} \tag{3.18d}$$

Equations 3.18 a-d expands the displacement vectors with displacements in the solid structure denoted by subscript  $s$  in the horizontal  $u_s$ , and vertical  $v_s$  directions, along with displacements in the fluid denoted by subscript  $f$  as well in the horizontal  $u_f$  and vertical  $v_f$  directions. Poro-elastic properties included are the bulk moduli of material grains  $K_s$ , drained media  $K_m$ , and fluid  $K_f$ , Biot's theory inertial effect  $\tilde{\rho}$ , (Pride et al, 2004), fluid density  $\rho_f$ , solid density  $\rho_s$ , poro-elastic coefficient of effective stress  $\alpha_\phi$ , elastic modulus  $E_m$ , bulk modulus variable  $B_m$ , unsaturated density  $\rho_1$  and saturated density  $\rho_2$ .

$$\begin{aligned}
\alpha_\phi = 1 - \frac{K_m}{K_s}, \quad E_m = \frac{K_s^2}{B_m - K_m}, \quad B_m = K_s \left( 1 + \phi \left( \frac{K_s}{K_f} - 1 \right) \right), \\
\rho_1 = (1 - \phi) \rho_s, \quad \rho_2 = \phi \rho_f, \quad \tilde{\rho} = \frac{i\eta}{\omega \kappa(\omega)},
\end{aligned} \tag{3.19}$$

The inertial effect  $\tilde{\rho}$  describes the flow resistance of the fluid within the pores and itself is calculated with the dynamic permeability  $\kappa(\omega)$ , which describes the ability for fluid to flow through the porous rock dependent upon the angular frequency of the travelling waves,  $\omega$  (Johnson et al, 1987). To calculate  $\kappa(\omega)$ , three properties must be determined, first is the steady-flow limit of permeability  $\kappa_0$ , where it indicates the permeability extent of the rocks which allows time-independent fluid flow. The second property is the relaxation frequency or Biot's critical frequency  $\omega_{BIOT}$ , an indication of whether viscous friction and inertial coupling effects are dominant at the low and high frequency range respectively, typically located in orders of tens of kHz (Yang and Mao, 2017). The third property is the cementation exponent  $c$ , which follows Archie's law about the relation between the porosity of a sedimentary rock and its electrical conductivity, which can be set as 1 for simplicity (Yang and Mao, 2017).

$$\begin{aligned}
\kappa(\omega) = \kappa_0 \left[ \sqrt{1 - i \frac{\omega}{2\omega_{Biot}}} - i \frac{\omega}{\omega_{Biot}} \right]^{-1}, \\
\omega_{Biot} = \frac{\eta}{\rho_f \kappa_0 \phi^{-c}},
\end{aligned} \tag{3.20}$$

$\varphi$  can also be considered an interpolation term whereby if its 0 the medium is not porous and can behave either in either a viscoelastic or fully elastic material, however if it is not zero then the media exhibits the effects of a porous structure. The Dirac delta  $\delta$  is used to indicate whether the effects of poro-elasticity from the media is applied upon the waves. Here the thesis proposes to superimpose the two-displacement kelvin-voigt viscoelastic wave equations upon the two poro-elastic wave equations 3.18a and 3.18b, and the two generalised Darcy's law equation only applies if there is porosity, represented for the horizontal and vertical displacement by the function  $F_u$  and  $F_v$ .

$$\begin{aligned} & \left[ \left( A_v + \delta(\varphi)(\alpha_\varphi(\alpha_\varphi - \varphi)E_m) \right) \frac{\partial^2}{\partial x^2} + B_v \frac{\partial^2}{\partial z^2} + \omega^2 \rho_t(\varphi) \right] u(\varphi) + \left[ \left( C_v + \delta(\varphi)\alpha_\varphi(\alpha_\varphi - \varphi)E_m \right) \frac{\partial^2}{\partial x \partial z} \right] v(\varphi) \\ & + \delta(\varphi) \left( \alpha_\varphi \varphi E_m \frac{\partial^2}{\partial x^2} + \omega^2 \rho_2 \right) u_f + \delta(\varphi) \left( \alpha_\varphi \varphi E_m \frac{\partial^2}{\partial x \partial z} \right) v_f = 0, \end{aligned} \quad (3.21a)$$

$$\begin{aligned} & \left[ \left( A_v + \delta(\varphi)(\alpha_\varphi(\alpha_\varphi - \varphi)E_m) \right) \frac{\partial^2}{\partial z^2} + B_v \frac{\partial^2}{\partial x^2} + \omega^2 \rho_t(\varphi) \right] v(\varphi) + \left[ \left( C_v + \delta(\varphi)\alpha_\varphi(\alpha_\varphi - \varphi)E_m \right) \frac{\partial^2}{\partial x \partial z} \right] u(\varphi) \\ & + \delta(\varphi) \left( \alpha_\varphi \varphi E_m \frac{\partial^2}{\partial z^2} + \omega^2 \rho_2 \right) v_f + \delta(\varphi) \left( \alpha_\varphi \varphi E_m \frac{\partial^2}{\partial x \partial z} \right) u_f = 0, \end{aligned} \quad (3.21b)$$

$$F_u(\varphi) = \begin{cases} F_u & \varphi \neq 0 \\ 0 & \varphi = 0 \end{cases}, \quad (3.21c)$$

$$F_v(\varphi) = \begin{cases} F_v & \varphi \neq 0 \\ 0 & \varphi = 0 \end{cases}, \quad (3.21d)$$

Equations 3.21a and 3.21b exhibit the two displacement and density functions of  $\varphi$ , where the presence of porosity would retrieve the corresponding vectors and absence of porosity would result in the viscoelastic vectors. The terms regarding fluid displacement are omitted from equations 3.21a and 3.21b if porosity is absent.

$$\delta(\varphi) = \begin{cases} 0 & \varphi = 0 \\ 1 & \varphi \neq 0 \end{cases}, \quad (3.22a)$$

$$\rho_t(\varphi) = \begin{cases} \rho & \varphi = 0 \\ \rho_1 & \varphi \neq 0 \end{cases}, \quad (3.22b)$$



$$u(\varphi) \begin{cases} u \\ u_s \end{cases} \begin{cases} \varphi = 0 \\ \varphi \neq 0 \end{cases}, \quad (3.22c)$$

$$v(\varphi) \begin{cases} v \\ v_s \end{cases} \begin{cases} \varphi = 0 \\ \varphi \neq 0 \end{cases}, \quad (3.22d)$$

Thus, this sub-chapter introduces a generalised viscoelastic wave equation in porous media, as result of the superposition of the generalised viscoelastic wave equation (Wang, 2016) and the poro-elastic wave equation (Yang and Mao, 2017). Using porosity as a co-efficient allows the equation to determine whether the subsurface media displays both poroelastic and viscoelastic behaviour in the presence of pores or exhibiting elastic-viscoelastic behaviour in the absence of it. In non-porous areas the equation follows the generalised viscoelastic wave equation by Wang, 2016. The reason for porous areas to exhibit both visco-elastic and poro-elastic behaviour is based upon the assumption that fluids in sufficient number of pores reflects a visco-elastic reaction to the wave. Due to this assumption, it is believed the proposed wave equation would be effective in highly porous rocks near the reservoir, such as limestones and sandstones, with significant number of the latter having porosity around the range of 12-22% (Stück et al, 2013). Because of the use of kelvin-voigt model, this wave equation is primarily suited within linear viscoelastic model.

The difference between this wave equation and that of others proposed for poro-viscoelastic wave equation (Sharma and Gogna, 1991, Santos et al, 2004, Picottia et al, 2010, Liu and Greenhalgh, 2018) is in its simplicity and flexibility to portray elastic, viscoelastic, and poro-elastic behaviour separately or together.

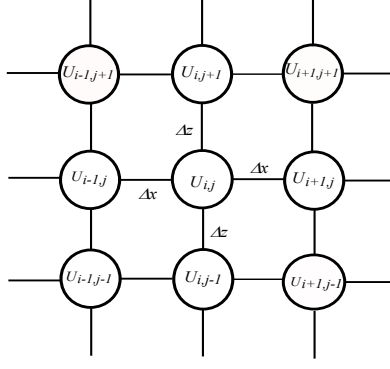
### 3.3 Modelling

#### 3.3.1 Finite difference

Implementing finite difference method involves applying differences between two points within the domain over the distance between as a means of discretising the PDE of the wave equation, the direction of the difference operator centres around the grid point of which the PDE must be solved.

Figure 3.5 illustrates a mesh of cartesian grid points that describes the model of the medium in the computational domain, when the wave equation is applied it simulates the wavefield at the node and is influenced by its neighbouring grid points who represent the neighbouring points within the medium.

The position of grid points  $U$  is indicated by its computational horizontal  $i$  and vertical  $j$  co-ordinates and are separated by the horizontal  $\Delta x$  and vertical  $\Delta z$  spatial interval, who represent their physical distances  $\partial x$  and  $\partial z$ , respectively. If  $\Delta x = \Delta z$ , then the  $h$  is used to denote the discretised steps.



**Figure 3.5** Cartesian grid points of the finite-difference method.

The finite difference formulation used for the discretisation of the spatial derivatives are organised into three categories, forward, backward, and central. The first order finite difference operator for the three categories involves finding the differences of two points over the distance between them.

$$\begin{aligned}
 \text{Forward:} \quad \frac{\partial u}{\partial x} &= \frac{U_{i+1,j} - U_{i,j}}{\Delta x}, & \frac{\partial u}{\partial z} &= \frac{U_{i,j+1} - U_{i,j}}{\Delta z}, \\
 \text{Backward:} \quad \frac{\partial u}{\partial x} &= \frac{U_{i,j} - U_{i-1,j}}{\Delta x}, & \frac{\partial u}{\partial z} &= \frac{U_{i,j} - U_{i,j-1}}{\Delta z}, \\
 \text{Central:} \quad \frac{\partial u}{\partial x} &= \frac{U_{i+1,j} - U_{i-1,j}}{2\Delta x}, & \frac{\partial u}{\partial z} &= \frac{U_{i,j+1} - U_{i,j-1}}{2\Delta z},
 \end{aligned} \tag{3.23a}$$

There are multiple ways to apply the finite difference operator to formulate the second order finite difference, namely applying the forward and backward operator twice is the least accurate as it would discount the dispersion of energy in the opposite direction. However, the two-fold application of the central finite difference operator also reduces accuracy because it omits the neighbouring grid points to the central node in both directions. The application that would return the greatest accuracy from discretisation is applying a backward operator onto a forward operator, because it would utilise the two neighbouring grid points on both sides of the central node. Mixed second order finite difference applies central finite difference operator in both directions.

$$\begin{aligned}
 \frac{\partial^2 u}{\partial x^2} &= \frac{U_{i+1,j} + U_{i-1,j} - 2U_{i,j}}{\Delta x^2}, \\
 \frac{\partial^2 u}{\partial z^2} &= \frac{U_{i,j+1} + U_{i,j-1} - 2U_{i,j}}{\Delta z^2}, \\
 \frac{\partial^2 u}{\partial x \partial z} &= \frac{U_{i+1,j+1} + U_{i-1,j-1} - U_{i+1,j-1} - U_{i-1,j+1}}{4\Delta x \Delta z},
 \end{aligned} \tag{3.23b}$$

### 3.3.2 Absorbing boundary layers

As the wave propagation through a computational grid is simulated it would eventually reach the ‘arbitrary’ edges of the domain, causing spurious reflections back into the grid, where there are no such boundaries physically. To mitigate these reflections, an absorbing boundary condition is needed around the grid to attenuate outgoing waves and eliminate inward reflections.

Clayton and Engquist (1977) introduced the absorbing boundary condition, which takes into consideration of waves diverging at small angles from the boundary wave equation i.e. one-way wave equation that travels perpendicular to the boundary axis. In this condition attenuation of wave amplitudes occurs primarily in parallel to the direction of waves travelling away from the boundary perpendicularly, as well as lessening the impact of waves not in parallel to these waves also known as paraxial approximation. The problem with this method is that the corner of the domain can act as sources for wave reflection because the absorbing boundary condition deals with waves travelling purely in the horizontal axis and vertical axis. Engquist and Majda (1977) would later derive an absorbing boundary condition to deal with the corner sources. Despite better accuracy with higher orders of the paraxial approximation, reflections still occur from waves propagated at wide divergent angles that are closer to the perpendicular direction of the travelling waves. Berenger (1994) developed the perfectly matched layer (PML), where propagating wave in the boundary layer is replaced with decaying waves by expanding the spatial derivative of the wave equation in the boundary layer to complex co-ordinates with the complex stretching formula  $\acute{S}$ . PML aims to resolve reflections from the incident waves at any frequency.

$$\frac{\partial}{\partial x} = \frac{1}{\acute{S}_x} \frac{\partial}{\partial x}, \quad \frac{\partial}{\partial z} = \frac{1}{\acute{S}_z} \frac{\partial}{\partial z}, \quad (3.24)$$

Traditional PML method requires a non-physical splitting of the wave equation along the horizontal and vertical axis, involving the summation of two-directional PML fields generated by two simulations of wave propagation, essentially doubling the computational forward modelling time. To circumvent non-physical splitting the convolutional perfectly matched layer (CPML), applies a stretched-coordinate formulation (in time domain, recursive convolution) onto the complex-frequency shifted perfectly matched layer (CFS-PML), (Kuzuoglu and Mittra, 1996, Roden and Gedney, 2000). CFS-PML involves shifting the damping factor from real to complex, using stretching function  $\acute{S}$ .

$$\acute{S}_x = \chi_x + \frac{d_x^*}{\acute{\alpha}_x + i\omega}, \quad \acute{S}_z = \chi_z + \frac{d_z^*}{\acute{\alpha}_z + i\omega}, \quad (3.25)$$

For most applications, the parameter  $\chi=1$  is generally acceptable to use (Gao et al, 2017). The frequency attenuation  $\dot{\alpha}$  is a spatially dependent variable where it decreases the further the distance to the innermost layer of the PML and  $d^*$  is the damping operator.

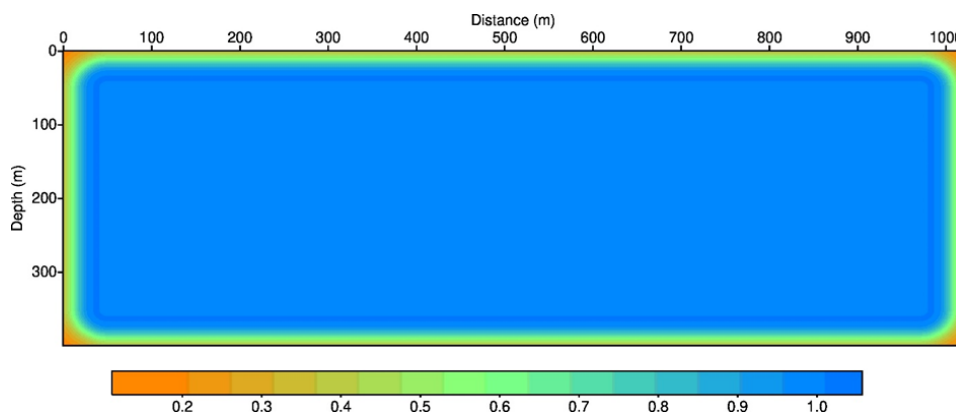
$$\dot{\alpha}_x = \dot{\alpha}_{\max} \left( 1 - \frac{|L_{pml} - L|}{L_{pml}} \right), \quad \dot{\alpha}_z = \dot{\alpha}_{\max} \left( 1 - \frac{|L_{pml} - L|}{L_{pml}} \right), \quad (3.26)$$

$\dot{\alpha}_{\max} = \pi f_p$ , is dependent on the peak frequency,  $f_p$  the frequency of which the wavelet chosen for the source peaks, along with the distance ratio determined by the PML layer grid-point  $L$ , and total PML layer width  $L_{pml}$ , helps determine the frequency attenuation.

$$d_x^* = d_0 \left( \frac{|L_{pml} - L|}{L_{pml}} \right)^2, \quad d_z^* = d_0 \left( \frac{|L_{pml} - L|}{L_{pml}} \right)^2, \quad (3.27a)$$

$$d_0 = \frac{3V_{\max}}{2L_{pml}} \log \left( \frac{1}{R} \right), \quad (3.27b)$$

The damping operator is more complex since it involves the additional damping co-efficient  $d_0$ , that is dependent upon the highest velocity of the model used in simulation  $V_{\max}$ , and the user determined value for the theoretical reflection co-efficient,  $R$  (Collino and Tsogka, 2001). From the authors' testing it is recommended for  $R=10^{-9}$  where  $L_{pml}=5h$ ,  $h$  represents the spatial interval steps where  $\Delta x=\Delta z$ . A demonstration is shown in figure 3.6 where the parameter  $\chi_x$  is set as 1 and the PML utilises 50 grid points, it should be noted that increasing the number of grid points slows down the factorization (method described in sub-chapter 3.4), The interior of the computational domain in light blue colour represents the digitised physical model and hence no complex stretching were no applied.



**Figure 3.6** An example of computational model with the applied CPML region of 50 grid points width and reflection co-efficient value of  $10^{-9}$  at 15Hz. With the scale representing the damping from high damping (brown) to no damping (blue).

The second order spatial derivative in wave equation within the PML layer is expanded using the quotient rule and chain rule with respect to the computational horizontal and vertical direction, however these are not needed when summation-by-parts is applied (explained in sub-chapter 4.2) which includes the complex stretching applied within the boundary layers. In the corner regions that are not perpendicular to the edges of the interior, the complex stretching factor is applied in both directions, illustrated in the schematic map of figure 3.7.

$d_z \neq 0$ $d_x \neq 0$	$d_x = 0$ $d_z \neq 0$	$d_z \neq 0$ $d_x \neq 0$
$d_z = 0$ $d_x \neq 0$	<i>Interior Domain</i>	$d_z = 0$ $d_x \neq 0$
$d_z \neq 0$ $d_x \neq 0$	$d_x = 0$ $d_z \neq 0$	$d_z \neq 0$ $d_x \neq 0$

**Figure 3.7** Schematic map of the CPML displaying where the damping operator in their respective direction is being applied.

### 3.3.3 Boundary conditions

However, when the wave propagation reaches the edges of the computational model external to the PML layer, the finite difference operator is inapplicable because it would theoretically require a non-existent grid-point outside the domain. Hence a boundary condition is required, for example the Dirichlet boundary condition where the value of the ‘outside’ grid-point is set to a specific value, this method naturally occurs for the sparse matrix when we solve linear algebra  $Ax^* = b$ , for the forward modelling (sub-chapter 3.4) and typically resembles a hard wall at the edge of the model, which may introduce small reflection back through the boundary layer into the domain. An alternative is the Neumann boundary condition which sets flux (spatial derivative) at the boundary to a specific value. These boundary condition can be homogeneous where the value set is 0, or inhomogeneous for non-zero values. The homogeneous Neumann boundary condition is applied when there is no wave propagation towards outside of the computational domain, whereas inhomogeneous assumes a 1 directional flow. The homogeneous Neumann boundary condition requires a ghost point where the artificial grid point outside of the computational model is equivalent to the nearest grid point at the edge, this is derived from setting a first order centre derivative at the edge of the domain to zero. Equation 3.28 illustrates the application of the Neumann boundary condition in the horizontal direction at the edge on the left side of the computational model where horizontal co-ordinate is 0.

$$\frac{\partial U}{\partial x} = \frac{U_{1,j} - U_{-1,j}}{2\Delta x} = 0, \quad \therefore U_{1,j} = U_{-1,j}, \quad (3.28)$$

### 3.3.4 Free surface boundary condition

Typically during seismic acquisition, the travelling waves behave differently when it reaches the surface of the recorded region, hence a free surface boundary condition is applied when displacement on interfaces is not constrained such as, air-water interface and air-land interface, this is because air doesn't have a defined surface where waves can travel through, therefore the air has no surface tension. Since there is no external pressure vertically above the surface, traction is removed this means that stress normal to the surface is effectively zero.

$$\sigma_{xz} = 0, \quad \sigma_{zz} = 0, \quad (3.29)$$

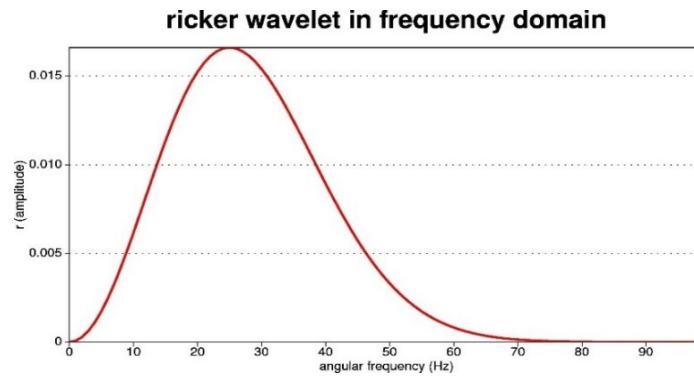
### 3.3.5 Source signature

When performing forward modelling it is important to consider the source from which waves propagate from, and an important emphasis is placed upon the seismic wavelet, known as the source signature. Seismic wavelets represent the source function, the most effective of which can capture changes in signals that represents reaction of the materials within the subsurface to waves propagating throughout the medium and are typically estimated after seismic acquisition in the field, where they are attained after deconvolving from seismic traces, and tied-in with well log data. There are many established seismic wavelets such as the minimum-phase wavelet (Mehta et al, 1991), constant-phase wavelet (Liu et al., 2018), Morelet wavelet (Bernadino and Santos-Victor, 2005) and many more, including ongoing research on wavelet estimation (Zhou et al, 2017), the topic of which is too broad for this thesis to cover. One of the most well-known wavelets used for forward modelling is the Ricker wavelet, also known as the Mexican hat wavelet due to shape of its analytical response, the reason is that the Ricker wavelet displays typical response from viscoelastic homogeneous materials to seismic waves (Ricker, 1944), and it's formulation in the frequency domain is shown in equation 3.30.

$$\begin{aligned} r(\omega) &= \frac{2f^2}{\pi^2 f_c^3} \exp\left(-\frac{f^2}{\pi f_c^2}\right), \\ f &= \omega / (2\pi), \\ f_c &= f_{co} / (3\sqrt{\pi}), \end{aligned} \quad (3.30)$$

The amplitude of the Ricker source  $r(\omega)$ , measured in units of angular frequency (rad/s)  $\omega$ , is determined by the temporal source frequency  $f$ , and the wavelets critical frequency  $f_c$ , the latter of which is dependent upon the cut-off frequency  $f_{co}$ , set by the user, where any frequency above this value, the signal wavelet ceases to show any significant amplitude. Below is the illustration of the Ricker wavelet

in frequency domain, where the dominant frequency can be controlled by the selection of  $f_{co}$  because as shown the dominant frequency lies around 1/3 of value of  $f_{co}$ , and as demonstrated below in order to set the dominant frequency at 25Hz,  $f_{co}$  must be set at 75Hz.



**Figure 3.8** Ricker wavelet in the frequency domain.

A second factor for the horizontal and vertical source term  $f_x, f_z$ , is the location of source point(s) in the horizontal  $x_s$ , and vertical location  $z_s$ , where the source vector  $\mathbf{f}$  (equation 3.33), is constrained by applying dirac-delta function  $\delta(\cdot)$ , onto the models' horizontal  $r_x$ , and vertical  $r_z$ , spatial coordinate, to match source location. Thus, the source terms would only apply at the source location.

$$-f_x = -f_z = -\delta(r_x - r_{x_s}, r_z - r_{z_s})r(\omega), \quad (3.31)$$

$$-\delta(r_x - r_{x_s}, r_z - r_{z_s}) \begin{cases} 1 & (r_x - r_{x_s}) \wedge (r_z - r_{z_s}) = 0 \\ 0 & (r_x - r_{x_s}) \vee (r_z - r_{z_s}) \neq 0 \end{cases} \quad (3.32)$$

### 3.4 Seismic data synthesis

In forward modelling, the wave equation coupled with the source term can be numerically solved in the matrix form systems of linear equation  $Ax^* = b$ ,

$$\mathbf{S}\mathbf{u} = \mathbf{f}, \quad (3.33)$$

Where sparse complex (due to applied PML layers) wave equation matrix  $\mathbf{S}$ , data wavefield vector  $\mathbf{u}$ , and source vector  $\mathbf{f}$ , corresponds to  $A$ ,  $x^*$ , and  $b$ , respectively. Practically the goal of the linear algebra equation in forward modelling and FWI is to retrieve the vector  $x^*$ .

There are many solvers available on the internet to solve this linear algebra computationally, one of which is the Multi-frontal Massively Parallel Solver (MUMPS) package developed by the CERFACS group (Amestoy et al, 2006). The scale of the complex matrix  $\mathbf{S}$  is  $ns * ns$ , where number of samples  $ns$ ,

is the total grid-points within the updating model.  $\mathbf{S}$  consists of diagonal bands and the number of diagonal bands is dependent on the number of neighbouring grid points used in discretising the wave equation (Figure 3.5). MUMPS uses a multifrontal technique, a direct method based on the  $LU$  or  $LDL^T$  factorisation of the sparse matrix  $\mathbf{S}$ , which is in the compressed row format, that is it vectorises the matrix row by row. To use MUMPS,  $\mathbf{S}$  can be entered in the ‘centralized assembled format’, where it requires three arrays to be filled, the array containing function of the grid-point, and the two arrays containing its horizontal co-ordinate ( $i$ ) and vertical co-ordinate ( $j$ ). This compressed sparse matrix format vectorizes matrix row by row and has the advantage that the elements can be in any order within the array of function, so long as it corresponds to its  $i$  and  $j$  within the other two arrays at the same sequence position.

The sparse matrix for an acoustic wave equation (equation 3.7) is demonstrated, where ghost points  $g_p$ , represent points outside of the domain of the model that depends upon the boundary condition used (sub-chapter 3.3.3).  $V_{re}$  and  $V_{rs}$  in the main diagonal denotes the last value of a row at the edge of the right of velocity model, and the first value of the subsequent row below, respectively.

$$\mathbf{S} = \begin{bmatrix} \left(-\frac{\omega^2}{v_1^2} + \frac{1}{s 4h^2}\right) & -\frac{1}{s 4h^2} & 0 & 0 & -\frac{1}{s 4h^2} & 0 & 0 & 0 \\ -\frac{1}{s 4h^2} & \left(-\frac{\omega^2}{v_2^2} + \frac{1}{s 4h^2}\right) & -\frac{1}{s 4h^2} & 0 & 0 & -\frac{1}{s 4h^2} & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & -\frac{1}{s 4h^2} & \left(-\frac{\omega^2}{v_n^2} + \frac{1}{s 4h^2}\right) & g_p & 0 & 0 & -\frac{1}{s 4h^2} \\ -\frac{1}{s 4h^2} & 0 & 0 & g_p & \left(-\frac{\omega^2}{v_n^2} + \frac{1}{s 4h^2}\right) & -\frac{1}{s 4h^2} & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & -\frac{1}{s 4h^2} & 0 & 0 & -\frac{1}{s 4h^2} & \left(-\frac{\omega^2}{v^2} + \frac{1}{s 4h^2}\right) & -\frac{1}{s 4h^2} \\ 0 & 0 & 0 & -\frac{1}{s 4h^2} & 0 & 0 & -\frac{1}{s 4h^2} & \left(-\frac{\omega^2}{v_m^2} + \frac{1}{s 4h^2}\right) \end{bmatrix} \quad (3.34)$$

In the sparse matrix, the main diagonal consists of central node of wave equation ( $i,j$ ), and the 8 other diagonal bands represent surrounding nodes (figure 3.5). The band immediately right of main diagonal band contains nodes to the right of the central node ( $i+1,j$ ) immediate left band corresponds to nodes on the left ( $i-1,j$ ). Furthest bands on the left and the right represents nodes above ( $i,j-1$ ) and below ( $i,j+1$ ) the central nodes.

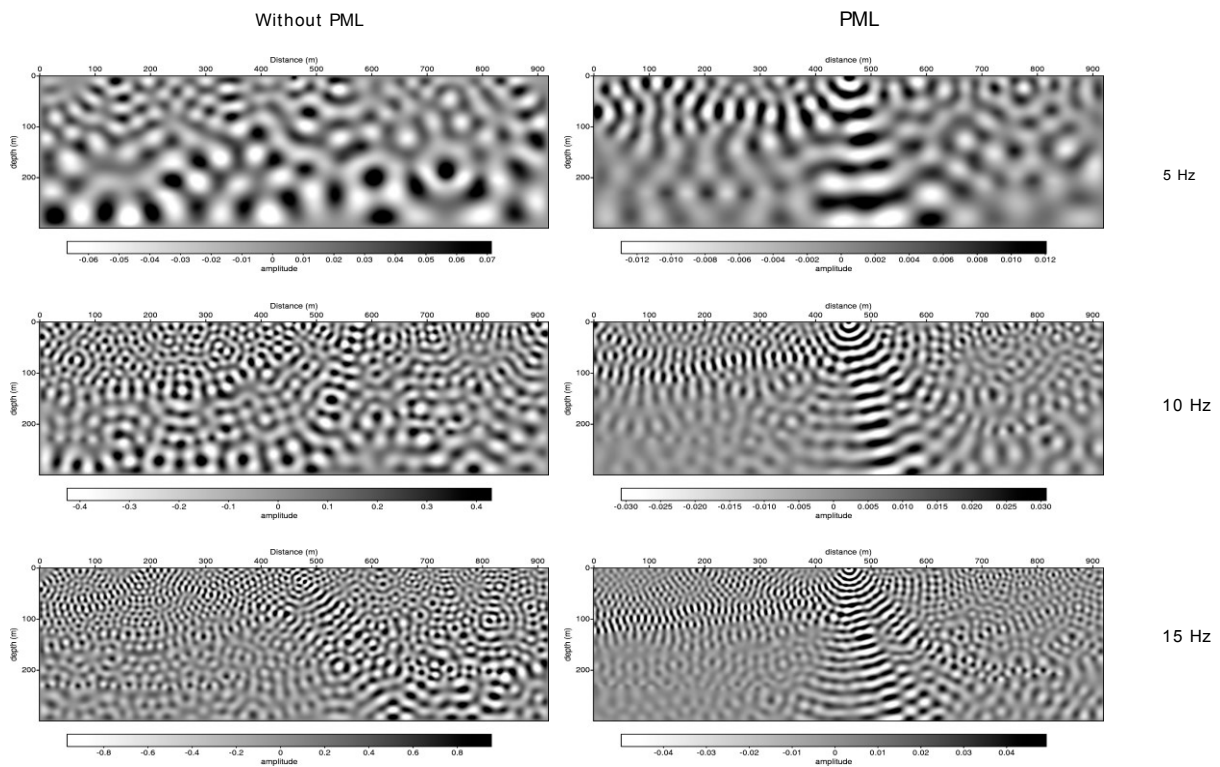
$$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_m \\ \vdots \\ u_{ns} \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} r(\omega)_{source} \\ \vdots \\ r(\omega)_{source} \\ 0 \\ 0 \end{bmatrix} \quad (3.35)$$



In source term vector  $\mathbf{f}$ , the ricker wavelet  $r(\omega)$  is applied at the grid-point location of the source within model matrix (note: seismic acquisition may contain multiple source). The model format is converted to the compressed row format vector, unless there are sources below surface any values below the first row is zero. The column vector of the wavefield  $\mathbf{u}$  is the compressed row format vector of the wavefield data, subscript  $m$  denotes any element between 1 and  $ns$ , and is obtained as follows:

$$\mathbf{u} = \mathbf{S}^{-1}\mathbf{f} \quad (3.36)$$

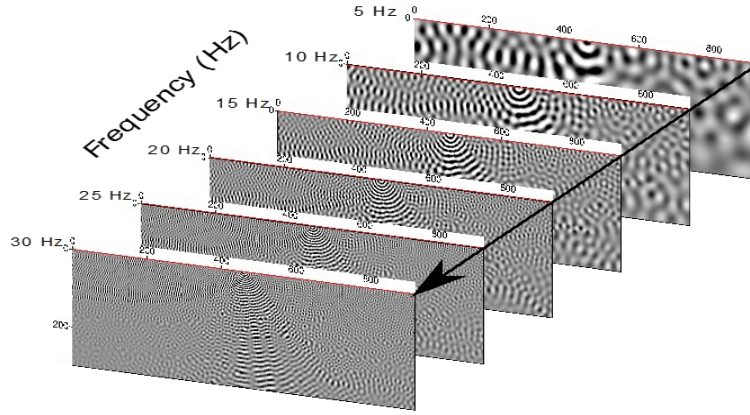
The figure below was generated by the author to demonstrate how different frequencies can affect the reflection wavefield recovered, where the L.H.S column is a result of forward modelling without PML layer and shows no coherent reflections, and the R.H.S. demonstrates forward modelling wavefield with PML layer included. The forward modelling demonstration below is generated using the Marmousi velocity model, a popular model used in FWI research that was originally devised by the Institut Francais du Petrole (IFP), (Society of Exploration Geophysicists, 2020), and later extended to elastic domain (Martin, 2004).



**Figure 3.9** Forward model wavefields for 5Hz to 15Hz incrementing in 5Hz interval from top to bottom. Wavefields without absorbing boundary layer show (left), with the absorbing boundary layer applied (right).

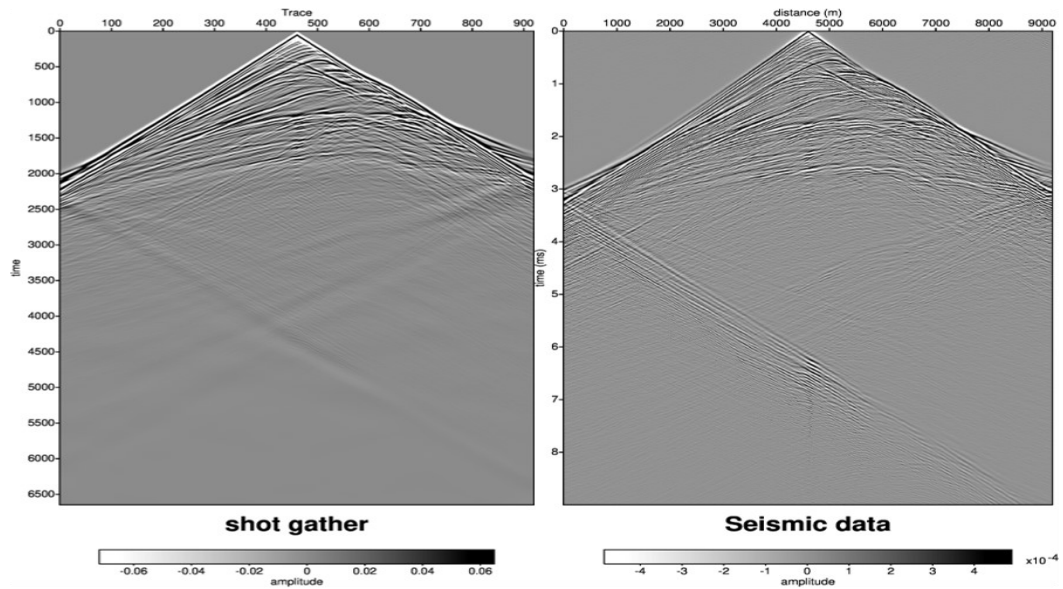
$d_{syn}(x_s, x_r, \omega)$  for each angular frequency increment  $\omega$  in frequency domain, and  $d_{syn}(x_s, x_r, t)$  for each time step  $t$ , in time domain, is obtained by extracting wavefield data  $u(x, z, \omega)$ , at the receiver  $x_r$ , on the surface where  $z=0$ , whereas the remaining wavefield terms of the vertical points  $nz$  isn't utilised. Thus  $d_{syn}(x_s, x_r, \omega)$  is essentially a collection of extracted wavefield data at the receiver location from a stack of wavefields arranged in incremental temporal steps (Figure 3.10).

$$d_{syn}(x_s, x_r, \omega) \begin{cases} u(x, z, \omega) & z = 0 \\ 0 & 0 < z < nz \end{cases} \quad (3.37)$$



**Figure 3.10** A schematic of wavefield data extraction from 5Hz to 30Hz linearly increase from top-right to bottom-left. Receiver depth position is highlighted in red.

In frequency domain wavefield extraction is done for both real and imaginary components of the wavefields, resulting in a complex synthetic data. To compare data synthesised between the two domain the Inverse Fourier transform is applied to the complex frequency domain seismic data trace by trace, to obtain its time domain representation, to compare it to time domain  $d_{syn}$ .



**Figure 3.11** Synthetic seismic data produced via the time domain time-marching method (left) and via inverse Fourier transform of the complex frequency domain wavefield data (right). It should be noted that the amplitude scale difference between the two data is due to the different absorbing boundary condition applied and the cause of the spurious edge reflection is due to the implementation of the Dirichlet boundary condition where the point outside of the domain of the model is set to zero.

### 3.5 Summary.

This chapter has discussed the methodology and crucial factors important for the forward modelling and tries to cover extensively the broad range of wave equations used for synthetic data generation and proposes its own original concept of the generalised viscoelastic wave equation within the porous media, where its theoretical motivation is that it is a simple and flexible wave equation that would contain as much information from the subsurface media as possible.

However, the proposed wave equation is formulated for a homogeneous media, but it should be noted that for heterogeneous media the product rule applied to the spatial derivatives with the viscoelastic and porous parameters may not be needed when utilising the more stable summation-by-parts method in finite difference (details in sub-chapter 4.2), where the information of the spatial derivatives of the viscoelastic and porous data, is replaced by an averaging factor of its co-efficient.

The topic of boundary conditions is also considered especially when the wave simulation reaches the edges of the models and that utilising all three (Absorbing Boundary Layer, Boundary Condition, and Free Surface Boundary Condition) would ensure accurate representation of wave propagation and prevent any spurious inward reflections. This chapter recommends using CPML as the absorbing boundary layer for the lower computational cost compared to PML, which requires require non-physical splitting of the seismic wave propagation, and hence more wavefields to be generated (Roden and Gedney, 2000, Komatitsch and Martin, 2007).

A very short introduction to source wavelets was also mentioned and that the formulation of the well-known Ricker wavelet is also explained. As of note due to such broad range of subjects and limitations the reader is advised to read upon  $Q$ -factor, Anisotropic modelling, and wavelet estimation.

Lastly this chapter goes through the step-by-step basis on the frequency domain modelling of wave propagation and describes the measures used in the MUMPs package (Amestoy et al, 2006) to solve the linear algebra, to attain the wavefield data in sub-chapter 3.4.

# Chapter 4: Topography

Conventional forward modelling assumes a flat surface and wave propagation is computed throughout a mesh made of quadrilateral cell representing the location of sub-surface medium properties. Realistically,  $d_{obs}$  are likely to be obtained from seismic data acquisition conducted on a series of variable elevation along the surfaces.

If these elevations are ignored, simulating wave propagation would generate erroneous  $d_{syn}$  that causes artificial cycle-skipping in FWI because data in  $d_{syn}$  are the result of recording waves reflected/refracted from incorrectly positioned structures (quantified as model parameters). However even if the wave simulation takes topography into account, it would still generate an incorrect  $d_{syn}$  that causes cycle-skipping in FWI, if it is performed over a computational domain mesh of equidistant nodes. This is because artificial scattering would occur when travelling waves encounter nodes representing model parameters whose computational domain location does not accurately represent its physical location. Performing FWI with this inaccuracy has created a reduction in resolution for the model being inverted (Bleibinhaus and Rondenay, 2009).

Therefore, a grid scheme that takes elevation into account, and the variably spaced model parameters are accurately positioned in the computational domain, so that artificial scattering would not occur and a truly representative  $d_{syn}$  is generated from forward modelling. This forward modelling must also consider an adaption of wave equation to the topography along with consideration of correctly placed sources of wave propagation as well as the backpropagation.

## 4.1 Body-fitted grid scheme

Body-fitted grid schemes are an operation of fitting computational grid points within an irregular shaped model, and are well established in computational fluid dynamics, and aerodynamics (Rane and Kovacevic, 2017, Mohebbi and Sellier, 2014). Spectral methods and finite-element method involving triangular grids and hexahedral meshes were suggested as to offer effective body fitting grids, however these methods have significant computational cost due to the many more additional grid points requiring higher memory requirements and time calculation (Zhang and Liu, 1999, Golubev et al, 2015). The curvilinear coordinate method utilises an optimized spatial operator and describes an irregular topography well, however this method requires additional derivative calculations, such as the chain rules used in tensorial formulation of wave equation (Komatitsch et al. 1996), which complicates wavefield modelling as it attempts to simulate wave propagation in irregular physical space onto computational domain cartesian model directly. The pseudo-orthogonal body-fitted grid scheme is a promising approach that offers well balanced trade-off between accurate body-grid generation and

propagation simulation with the computational costs (Rao and Wang, 2013). This method of grid generation can be easily implemented because it uses the finite-difference numerical method. In the pseudo-orthogonal body grid scheme, the model horizontal and vertical co-ordinates in the physical space  $(x,z)$  is mapped into the computational space  $(\zeta,\psi)$ . The relation between the two spaces is given by the rate of grid spacing changes in horizontal direction  $M(\zeta,\psi)$  and in vertical direction  $N(\zeta,\psi)$ . This relation is based upon the Poisson's equation, where the rate of grid spacing change is used to relate changes of computational co-ordinates with respect to the changes of the physical parameters.

$$\begin{aligned}\frac{\partial^2 \zeta}{\partial x^2} + \frac{\partial^2 \zeta}{\partial z^2} &= M(\zeta, \psi), \\ \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial z^2} &= N(\zeta, \psi),\end{aligned}\tag{4.1}$$

Since the Poisson's equation is an elliptic equation, the physical space co-ordinates corresponding to quadrilateral cell based co-ordinates in computational space  $x=x(\zeta,\psi)$  and  $z=z(\zeta,\psi)$  can be solved for iteratively with the elliptic partial differential equations. This is done only for the internal grids and the points along the peripheral boundaries are discussed afterwards.

$$\begin{aligned}\alpha_{BFG} \frac{\partial^2 x}{\partial \zeta^2} + \beta_{BFG} \frac{\partial^2 x}{\partial \psi^2} - 2\gamma_{BFG} \frac{\partial^2 x}{\partial \zeta \partial \psi} + J^2 \left( M \frac{\partial x}{\partial \zeta} + N \frac{\partial x}{\partial \psi} \right) &= 0, \\ \alpha_{BFG} \frac{\partial^2 z}{\partial \zeta^2} + \beta_{BFG} \frac{\partial^2 z}{\partial \psi^2} - 2\gamma_{BFG} \frac{\partial^2 z}{\partial \zeta \partial \psi} + J^2 \left( M \frac{\partial z}{\partial \zeta} + N \frac{\partial z}{\partial \psi} \right) &= 0,\end{aligned}\tag{4.2}$$

Where the coefficients for the second order horizontal  $\alpha_{BFG}$ , vertical  $\beta_{BFG}$ , derivative, and their first order counterpart  $J$ , are solved for each internal grid points.

$$\alpha_{BFG} = \dot{x}_\psi^2 + \dot{z}_\psi^2,\tag{4.3a}$$

$$\beta_{BFG} = \dot{x}_\zeta^2 + \dot{z}_\zeta^2,\tag{4.3b}$$

$$J = \frac{\partial(x,z)}{\partial(\zeta,\psi)} = \dot{x}_\zeta \dot{z}_\psi - \dot{z}_\zeta \dot{x}_\psi,\tag{4.3c}$$

Dot notation above physical space co-ordinates denotes discretised Poisson's equation of the derivative of physical spatial interval with respect to its corresponding computational counterpart and is determined with first order central difference scheme.

$$\dot{x}_\zeta = \frac{\partial x}{\partial \zeta}, \quad \dot{x}_\psi = \frac{\partial x}{\partial \psi}, \quad \dot{z}_\zeta = \frac{\partial z}{\partial \zeta}, \quad \dot{z}_\psi = \frac{\partial z}{\partial \psi},\tag{4.4}$$

All grid points must satisfy the grid orthogonality condition  $\gamma_{BFG}$ , to allow implementation of absorbing boundary condition for topographic waveform simulation, where grid lines should be perpendicular and parallel to points of intersection of the model. Grid quality  $q$  is used to measure levels of non-orthogonality and should be aimed to be kept as minimal as possible, and is the sum of all  $\gamma_{BFG}$ , where  $i$ , and  $j$  represent its horizontal and vertical co-ordinates.

$$\gamma_{BFG} = \dot{x}_\zeta \dot{x}_\psi + \dot{z}_\zeta \dot{z}_\psi = 0, \quad (4.5a)$$

$$q = \sum_{i,j} \gamma_{BFG}, \quad (4.5b)$$

Boundary point modification is needed to solve the non-orthogonal boundary grids, because if left neglected it would cause immense misalignment between the boundary and grid points (Patatonis and Atharassiadis, 1985), and become the source of new artificial scattering. Equation 4.6a relates to the orthogonality in equation 4.5a, and equations 4.6b and 4.6c signifies the smoothness in the  $\zeta$  and  $\psi$  directions, respectively. The system of equations 4.6a-c is an example of modifying the top layer boundary i.e. the surface, with grid point co-ordinates at  $i,j+1$  from  $x_{i,j+1}, z_{i,j+1}$  to  $\tilde{x}_{i,j+1}, \tilde{z}_{i,j+1}$ .

$$(x_{i+1,j} - x_{i-1,j})(\tilde{x}_{i,j+1} - x_{i,j}) + (z_{i+1,j} - z_{i-1,j})(\tilde{z}_{i,j+1} - z_{i,j}) = 0, \quad (4.6a)$$

$$(x_{i+1,j+1} - \tilde{x}_{i,j+1})(\tilde{z}_{i,j+1} - z_{i-1,j+1}) - (\tilde{x}_{i,j+1} - x_{i-1,j+1})(z_{i+1,j+1} - \tilde{z}_{i,j+1}) = 0, \quad (4.6b)$$

$$(x_{i,j+2} - \tilde{x}_{i,j+1})(\tilde{z}_{i,j+1} - z_{i,j}) - (\tilde{x}_{i,j+1} - x_{i,j})(z_{i,j+2} - \tilde{z}_{i,j+1}) = 0, \quad (4.6c)$$

The pseudo-orthogonal grid generation scheme was derived from the body-fitted grid scheme that is conventionally used to generate meshes for irregular shaped objects. However, for a model region with surfaces of various elevations, the nodes horizontal position can be fixed, and its grid rate change can be negligible  $M(\zeta, \psi) = 0$ . Then, the vertical grid rate change  $N(\zeta, \psi)$  can be determined, where within the model, the low velocity region near the surface have densely-packed grids, whereas for the higher velocity and deeper area are modelled with grids that are sparsely packed.

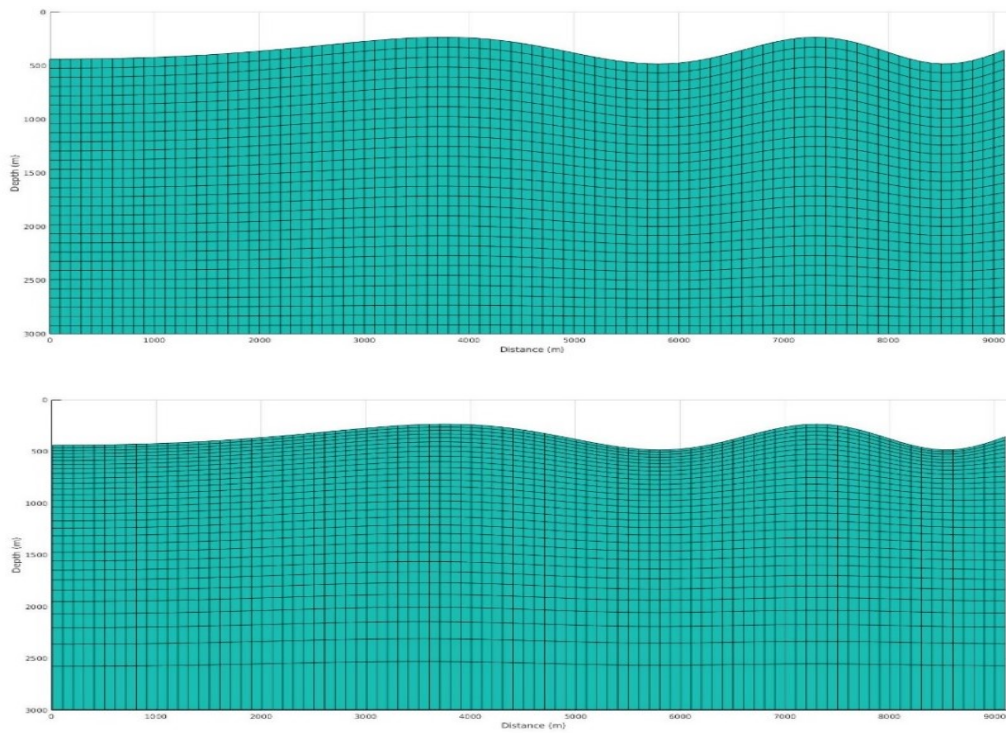
$$N(\zeta, \psi) = -\frac{2L(\alpha_{BFG} + \beta_{BFG})\Delta r}{J^2 \dot{z}_\psi (1+r)^2}, \quad (4.7)$$

$L = z_{i,j+1} - z_{i,j-1}$  is the distance between the two points closest to  $z_{i,j}$  and the grid ratio is calculated via  $r = (z_{i,j+1} - z_{i,j}) / (z_{i,j} - z_{i,j-1})$  the value for the change rate in grid size is  $\Delta r < 5\%$ .

Generating body-fitted grids must begin with setting boundaries with evenly spaced initial points and

then perform linear interpolation over those points to generate initial internal grids. The following iterative steps is performed:

1. Solve equation 4.2 to generate the orthogonal internal meshes.
2. Modify boundary points.
3. Measure grid quality  $q$ , if it's not sufficiently low enough, adjust  $M(\zeta, \psi)$  and  $N(\zeta, \psi)$ , and repeat the process by going to step 1.



**Figure 4.1** Mesh of grids for irregular topography generated by the author. Mesh before body-fitted grid scheme (top), and after applying the body-fitted grid scheme (bottom).

## 4.2 Wave equations for body-fitted grid mesh.

Irregular topography produces a loss of coherency of waves. Ignoring this can introduce artefacts in the inversion result especially at smaller wavelength. Multiples also lose their coherency due to strong topography, hence the corresponding artefacts are reduced and mitigates reverberation. However strong topography also produces additional scattering of body waves which can reduce resolution if not accounted for, and introduce Rayleigh waves, which along with other surface-related phases can introduce artifices to the synthetic data if the free surface effect (i.e. air) is ignored (Lee et al, 2009, Nuber et al, 2016, Bleibinhaus and Rondenay, 2009). Niño and Vides (2005) suggest applying wave equation datuming to perform static corrections on seismic data via extrapolation with a Kirchhoff integral applied to a datum to mimic flat surface propagation. Unfortunately, this further exacerbates the local minima problem as this distorts positions of reflection points in the seismic data.

In this thesis, the pseudo-orthogonal grid wave equation is used to describe the spatial derivatives of propagating waves, which could be applied using the finite difference method, thus it has low computational cost compared to other wave equation formulations. This method utilises a conversion term to relate the physical space to the computational space. These conversion terms are retrieved when setting the left-hand side of equation 4.3c to zero (Rao and Wang, 2013).

$$\zeta_x = \frac{\dot{z}_\psi}{J}, \quad \zeta_z = -\frac{\dot{x}_\psi}{J}, \quad \psi_x = -\frac{\dot{z}_\zeta}{J}, \quad \psi_z = \frac{\dot{x}_\zeta}{J}, \quad (4.8)$$

The first order physical space spatial derivatives consist of terms relating to computational derivatives.

$$\begin{aligned} \frac{\partial}{\partial x} &= \zeta_x \frac{\partial}{\partial \zeta} + \psi_x \frac{\partial}{\partial \psi}, \\ \frac{\partial}{\partial z} &= \zeta_z \frac{\partial}{\partial \zeta} + \psi_z \frac{\partial}{\partial \psi}, \end{aligned} \quad (4.9)$$

The second order spatial derivative involves expanding each term on the R.H.S. of equation 4.9 with the whole R.H.S. of the equation, thus resulting in the following five term formulation. The overhead dot above  $\zeta$  and  $\psi$  denotes a second order derivative with respect to its subscripts. Equation 4.10 denotes a computational domain second order derivative for a homogeneous media.

$$\begin{aligned} \frac{\partial^2}{\partial x^2} &= \zeta_x^2 \frac{\partial^2}{\partial \zeta^2} + \psi_x^2 \frac{\partial^2}{\partial \psi^2} + 2\zeta_x \psi_x \frac{\partial^2}{\partial \zeta \partial \psi} + \dot{\zeta}_x \frac{\partial}{\partial \zeta} + \dot{\psi}_x \frac{\partial}{\partial \psi}, \\ \frac{\partial^2}{\partial z^2} &= \zeta_z^2 \frac{\partial^2}{\partial \zeta^2} + \psi_z^2 \frac{\partial^2}{\partial \psi^2} + 2\zeta_z \psi_z \frac{\partial^2}{\partial \zeta \partial \psi} + \dot{\zeta}_z \frac{\partial}{\partial \zeta} + \dot{\psi}_z \frac{\partial}{\partial \psi}, \\ \frac{\partial^2}{\partial x \partial z} &= \zeta_x \zeta_z \frac{\partial^2}{\partial \zeta^2} + (\zeta_x \psi_z + \zeta_z \psi_x) \frac{\partial^2}{\partial \zeta \partial \psi} + \psi_x \psi_z \frac{\partial^2}{\partial \psi^2} + \dot{\zeta}_{xz} \frac{\partial}{\partial \zeta} + \dot{\psi}_{xz} \frac{\partial}{\partial \psi}, \end{aligned} \quad (4.10)$$

For inhomogeneous media, the first order spatial derivatives for the properties of the medium must be considered. For demonstration purposes the horizontal displacement wave equation for inhomogeneous elastic media (equation 3.5) is used, where spatial derivatives of wavefields and Lamé parameters are expanded in the computational domain and rearranged into the following.

$$\begin{aligned} \rho \frac{\partial^2 u}{\partial t^2} &= [(\lambda + 2\mu)\zeta_x^2 + \mu\zeta_z^2] \frac{\partial^2 u}{\partial \zeta^2} + [(\lambda + 2\mu)\psi_x^2 + \mu\psi_z^2] \frac{\partial^2 u}{\partial \psi^2} + 2[(\lambda + 2\mu)\zeta_x \psi_x + \mu\zeta_z \psi_z] \frac{\partial^2 u}{\partial \zeta \partial \psi} \\ &+ [(\lambda + \mu)\zeta_x \zeta_z] \frac{\partial^2 v}{\partial \zeta^2} + [(\lambda + 2\mu)\psi_x \psi_z] \frac{\partial^2 v}{\partial \psi^2} + [(\lambda + \mu)(\zeta_x \psi_z + \zeta_z \psi_x)] \frac{\partial^2 v}{\partial \zeta \partial \psi} \\ &+ \left[ \zeta_x^2 \frac{\partial(\lambda + 2\mu)}{\partial \zeta} + \zeta_z^2 \frac{\partial \mu}{\partial \zeta} + \zeta_x \psi_x \frac{\partial(\lambda + 2\mu)}{\partial \psi} + \zeta_z \psi_z \frac{\partial \mu}{\partial \psi} + \dot{\zeta}_x + \dot{\zeta}_z \right] \frac{\partial u}{\partial \zeta} \\ &+ \left[ \psi_x^2 \frac{\partial(\lambda + 2\mu)}{\partial \psi} + \psi_z^2 \frac{\partial \mu}{\partial \psi} + \zeta_x \psi_x \frac{\partial(\lambda + 2\mu)}{\partial \zeta} + \zeta_z \psi_z \frac{\partial \mu}{\partial \zeta} + \dot{\psi}_x + \dot{\psi}_z \right] \frac{\partial u}{\partial \psi} \\ &+ \left[ \zeta_x \zeta_z \left( \frac{\partial \lambda}{\partial \zeta} + \frac{\partial \mu}{\partial \zeta} \right) + \zeta_z \psi_x \frac{\partial \lambda}{\partial \psi} + \zeta_x \psi_z \frac{\partial \mu}{\partial \psi} + \dot{\zeta}_{xz} \right] \frac{\partial v}{\partial \zeta} + \left[ \psi_x \psi_z \left( \frac{\partial \lambda}{\partial \psi} + \frac{\partial \mu}{\partial \psi} \right) + \zeta_x \psi_z \frac{\partial \lambda}{\partial \zeta} + \zeta_z \psi_x \frac{\partial \mu}{\partial \zeta} + \dot{\psi}_{xz} \right] \frac{\partial v}{\partial \psi} \end{aligned} \quad (4.11)$$



### 4.3 Summation-by-parts formulation

However, the body-fitted grid method can produce mesh schemes with low precision and strong variation within cell size which causes instability, thus applying the above wave equation is restricted to meshes that have acute angles limited between  $67^\circ$ - $90^\circ$ , and the grid size can only be changed less than 5% (Rao and Wang, 2013), and not suitable in the presence of fluctuating interfaces beneath the surfaces, and regions with strong topography such as the ocean bottom (Rao and Wang, 2018).

Fortunately, this problem can be mitigated when the spatial derivatives are formulated with the Summation-by-parts (SBP) method, which has been used for quite some time, since its use for the first order derivative (Kreiss and Scherer, 1974, Strand, 1994) and second order derivative (Mattsson and Nordström, 2004). The SBP formulation was recently applied to more complicated equations such as the pseudo-acoustic wave equation (Rao and Wang, 2018), and elastic wave equation (Appelö and Petersson, 2009) as it provides strong stability for numerical simulation of finite difference. SBP method is proven stable because it mimics the effects of integration-by-parts by conserving the energy within the boundaries of each discretised step.

The first order spatial derivatives would be the same as that of equation 4.9, but the SBP second order derivatives, involves shifting the coefficients of the second derivative and multiplying it with the coefficients of the second derivative whilst omitting terms containing the overhead dot of  $\zeta$  and  $\psi$ .

$$\begin{aligned}\frac{\partial^2}{\partial x^2} &= \frac{\partial}{\partial \zeta} \left( \zeta_x^2 \frac{\partial}{\partial \zeta} \right) + \frac{\partial}{\partial \zeta} \left( \zeta_x \psi_x \frac{\partial}{\partial \psi} \right) + \frac{\partial}{\partial \psi} \left( \psi_x \zeta_x \frac{\partial}{\partial \zeta} \right) + \frac{\partial}{\partial \psi} \left( \psi_x^2 \frac{\partial}{\partial \psi} \right), \\ \frac{\partial^2}{\partial z^2} &= \frac{\partial}{\partial \zeta} \left( \zeta_z^2 \frac{\partial}{\partial \zeta} \right) + \frac{\partial}{\partial \zeta} \left( \zeta_z \psi_z \frac{\partial}{\partial \psi} \right) + \frac{\partial}{\partial \psi} \left( \psi_z \zeta_z \frac{\partial}{\partial \zeta} \right) + \frac{\partial}{\partial \psi} \left( \psi_z^2 \frac{\partial}{\partial \psi} \right), \\ \frac{\partial^2}{\partial x \partial z} &= \frac{\partial}{\partial \zeta} \left( \zeta_x \zeta_z \frac{\partial}{\partial \zeta} \right) + \frac{\partial}{\partial \zeta} \left( \zeta_x \psi_z \frac{\partial}{\partial \psi} \right) + \frac{\partial}{\partial \psi} \left( \zeta_z \psi_x \frac{\partial}{\partial \zeta} \right) + \frac{\partial}{\partial \psi} \left( \psi_x \psi_z \frac{\partial}{\partial \psi} \right),\end{aligned}\tag{4.12}$$

Spatial derivatives in equation 4.12 can be applied directly to elastic wave equation in homogeneous media by multiplying with Lamé parameters (like equation 3.3), but its application to inhomogeneous medium (equation 3.5) would result in an inconveniently long formula such as equation 4.11. Instead, the SBP can be applied to the original elastic wave equations i.e. equation 3.2, this thesis demonstrates this application with the horizontal displacement formula in 3.2.

$$\begin{aligned}\rho \frac{\partial^2 u}{\partial t^2} &= \frac{\partial}{\partial \zeta} \left[ ((\lambda + 2\mu) \zeta_x^2 + (\mu) \zeta_z^2) \frac{\partial u}{\partial \zeta} \right] + \frac{\partial}{\partial \zeta} \left[ ((\lambda + 2\mu) \zeta_x \psi_x + (\mu) \zeta_z \psi_z) \frac{\partial u}{\partial \psi} \right] + \frac{\partial}{\partial \zeta} \left[ (\lambda + \mu) \zeta_x \zeta_z \frac{\partial v}{\partial \zeta} \right] \\ &+ \frac{\partial}{\partial \zeta} \left[ ((\lambda) \zeta_x \psi_z + (\mu) \zeta_z \psi_x) \frac{\partial v}{\partial \psi} \right] + \frac{\partial}{\partial \psi} \left[ ((\lambda + 2\mu) \zeta_x \psi_x + (\mu) \zeta_z \psi_z) \frac{\partial u}{\partial \zeta} \right] + \frac{\partial}{\partial \psi} \left[ ((\lambda + 2\mu) \psi_x^2 + (\mu) \psi_z^2) \frac{\partial u}{\partial \psi} \right] \\ &+ \frac{\partial}{\partial \psi} \left[ ((\lambda) \zeta_z \psi_x + (\mu) \zeta_x \psi_z) \frac{\partial v}{\partial \zeta} \right] + \frac{\partial}{\partial \psi} \left[ (\lambda + \mu) \psi_x \psi_z \frac{\partial v}{\partial \psi} \right],\end{aligned}\tag{4.13a}$$

$$\begin{aligned}
\rho \frac{\partial^2 v}{\partial t^2} = & \frac{\partial}{\partial \zeta} \left[ ((\lambda + 2\mu)\zeta_z^2 + (\mu)\zeta_x^2) \frac{\partial v}{\partial \zeta} \right] + \frac{\partial}{\partial \zeta} \left[ ((\lambda + 2\mu)\zeta_z \psi_z + (\mu)\zeta_x \psi_x) \frac{\partial v}{\partial \psi} \right] + \frac{\partial}{\partial \zeta} \left[ (\lambda + \mu)\zeta_x \zeta_z \frac{\partial u}{\partial \zeta} \right] \\
& + \frac{\partial}{\partial \zeta} \left[ ((\lambda)\zeta_z \psi_x + (\mu)\zeta_x \psi_z) \frac{\partial u}{\partial \psi} \right] + \frac{\partial}{\partial \psi} \left[ ((\lambda + 2\mu)\zeta_z \psi_z + (\mu)\zeta_x \psi_x) \frac{\partial v}{\partial \zeta} \right] + \frac{\partial}{\partial \psi} \left[ ((\lambda + 2\mu)\psi_z^2 + (\mu)\psi_x^2) \frac{\partial v}{\partial \psi} \right] \\
& + \frac{\partial}{\partial \psi} \left[ ((\lambda)\zeta_x \psi_z + (\mu)\zeta_z \psi_x) \frac{\partial u}{\partial \zeta} \right] + \frac{\partial}{\partial \psi} \left[ (\lambda + \mu)\psi_x \psi_z \frac{\partial u}{\partial \psi} \right],
\end{aligned} \tag{4.13b}$$

SBP finite difference method is essentially a central finite difference method with boundary treatment upon each discretised step (Mattson, 2012). Before continuing, for clarity purposes, the author suggests that the formulations involving property  $(\lambda, \mu)$  and computational co-ordinate  $(\zeta, \psi)$  terms that multiply with the inner derivatives are taken out and denoted as  $H_I$ - $H_{I2}$  in equation 4.14.

$$\begin{aligned}
H_1 &= (\lambda + 2\mu)\zeta_x^2 + (\mu)\zeta_z^2 & H_7 &= (\lambda + \mu)\psi_x \psi_z \\
H_2 &= (\lambda + 2\mu)\zeta_x \psi_x + (\mu)\zeta_z \psi_z & H_8 &= (\lambda + 2\mu)\zeta_z^2 + (\mu)\zeta_x^2 \\
H_3 &= (\lambda + \mu)\zeta_x \zeta_z & H_9 &= (\lambda + 2\mu)\zeta_z \psi_z + (\mu)\zeta_x \psi_x \\
H_4 &= (\lambda)\zeta_x \psi_z + (\mu)\zeta_z \psi_x & H_{10} &= (\lambda)\zeta_z \psi_x + (\mu)\zeta_x \psi_z \\
H_5 &= (\lambda + 2\mu)\psi_x^2 + (\mu)\psi_z^2 & H_{11} &= (\lambda + 2\mu)\psi_z^2 + (\mu)\psi_x^2 \\
H_6 &= (\lambda)\zeta_z \psi_x + (\mu)\zeta_x \psi_z & H_{12} &= (\lambda)\zeta_x \psi_z + (\mu)\zeta_z \psi_x
\end{aligned} \tag{4.14}$$

Therefore, the inhomogeneous elastic wave equation in SBP formulation of equation 4.13a-b can be shortened as the equation 4.15a-b.

$$\begin{aligned}
\rho \frac{\partial^2 u}{\partial t^2} = & \frac{\partial}{\partial \zeta} \left[ H_1 \frac{\partial u}{\partial \zeta} \right] + \frac{\partial}{\partial \zeta} \left[ H_2 \frac{\partial u}{\partial \psi} \right] + \frac{\partial}{\partial \zeta} \left[ H_3 \frac{\partial v}{\partial \zeta} \right] + \frac{\partial}{\partial \zeta} \left[ H_4 \frac{\partial v}{\partial \psi} \right] \\
& + \frac{\partial}{\partial \psi} \left[ H_2 \frac{\partial u}{\partial \zeta} \right] + \frac{\partial}{\partial \psi} \left[ H_5 \frac{\partial u}{\partial \psi} \right] + \frac{\partial}{\partial \psi} \left[ H_6 \frac{\partial v}{\partial \zeta} \right] + \frac{\partial}{\partial \psi} \left[ H_7 \frac{\partial v}{\partial \psi} \right],
\end{aligned} \tag{4.15a}$$

$$\begin{aligned}
\rho \frac{\partial^2 v}{\partial t^2} = & \frac{\partial}{\partial \zeta} \left[ H_8 \frac{\partial v}{\partial \zeta} \right] + \frac{\partial}{\partial \zeta} \left[ H_9 \frac{\partial v}{\partial \psi} \right] + \frac{\partial}{\partial \zeta} \left[ H_3 \frac{\partial u}{\partial \zeta} \right] + \frac{\partial}{\partial \zeta} \left[ H_{10} \frac{\partial u}{\partial \psi} \right] \\
& + \frac{\partial}{\partial \psi} \left[ H_9 \frac{\partial v}{\partial \zeta} \right] + \frac{\partial}{\partial \psi} \left[ H_{11} \frac{\partial v}{\partial \psi} \right] + \frac{\partial}{\partial \psi} \left[ H_{12} \frac{\partial u}{\partial \zeta} \right] + \frac{\partial}{\partial \psi} \left[ H_7 \frac{\partial u}{\partial \psi} \right],
\end{aligned} \tag{4.15b}$$

The discretisation of the SBP finite difference method is essentially like that of conventional finite difference method but with the caveat that for the second order co-ordinate derivatives the model parameters and computational co-ordinates are averaged between two node points used in the spatial derivative. It does this by using the averaging operator  $E_{1/2}$  on the model parameters within the grid cells, which keeps the energy of the wave within each boundary steps.  $E_{1/2}$  is not applied to the mixed second order derivatives. Equation 4.15a-b are rewritten as follows.

$$\begin{aligned} \rho \frac{\partial^2 u}{\partial t^2} = & D_-^\zeta \left( E_{1/2}^\zeta (H_1) D_+^\zeta (u) \right) + D_0^\zeta \left( H_2 D_0^\psi (u) \right) + D_-^\zeta \left( E_{1/2}^\zeta (H_3) D_+^\zeta (v) \right) + D_0^\zeta \left( H_4 D_0^\psi (v) \right) \\ & + D_0^\psi \left( H_2 D_0^\zeta (u) \right) + D_-^\psi \left( E_{1/2}^\psi (H_5) D_+^\psi (u) \right) + D_0^\psi \left( H_6 D_0^\zeta (v) \right) + D_-^\psi \left( E_{1/2}^\psi (H_7) D_+^\psi (v) \right), \end{aligned} \quad (4.16a)$$

$$\begin{aligned} \rho \frac{\partial^2 v}{\partial t^2} = & D_-^\zeta \left( E_{1/2}^\zeta (H_8) D_+^\zeta (v) \right) + D_0^\zeta \left( H_9 D_0^\psi (v) \right) + D_-^\zeta \left( E_{1/2}^\zeta (H_3) D_+^\zeta (u) \right) + D_0^\zeta \left( H_{10} D_0^\psi (u) \right) \\ & + D_0^\psi \left( H_9 D_0^\zeta (v) \right) + D_-^\psi \left( E_{1/2}^\psi (H_{11}) D_+^\psi (v) \right) + D_0^\psi \left( H_{12} D_0^\zeta (u) \right) + D_-^\psi \left( E_{1/2}^\psi (H_7) D_+^\psi (u) \right), \end{aligned} \quad (4.16b)$$

For the second order derivative discretised in SBP formulation, involves a backward finite difference  $D_-$  applied to the forward finite difference  $D_+$  aligned with the averaging operator  $E_{1/2}$ . This results in the formulation of two averages of parameters, one for the forward difference, and another for the back difference.  $\zeta$  and  $\psi$  used as superscripts denotes the computational direction of which the finite difference and averaging operator is applied. Mixed second order derivatives applies central finite difference operator  $D_0$  in both directions, where it results in utilising nodes at the corner (figure 3.5).

For demonstration purposes, the discretisation involves a generalised wavefield  $\varpi$  representing either horizontal  $u$ , or vertical  $v$ , displacement. The generalised spatial step  $h$  is utilised if the computational spatial step interval is uniform across horizontal  $\partial\zeta$ , and vertical direction  $\partial\psi$ , i.e.  $\partial\zeta = \partial\psi$ , or if the physical spatial step is the same i.e.  $\Delta x = \Delta z$ . In the demonstration for the averaging operator the generalised parameter  $H$  is used and could be describing any parameter of  $H_1$ - $H_{12}$ .

$$\begin{aligned} D_-^\zeta (\varpi) &= \frac{\varpi_{i,j} - \varpi_{i-1,j}}{h}, \quad D_+^\zeta (\varpi) = \frac{\varpi_{i+1,j} - \varpi_{i,j}}{h}, \quad D_-^\psi (\varpi) = \frac{\varpi_{i,j} - \varpi_{i,j-1}}{h}, \quad D_+^\psi (\varpi) = \frac{\varpi_{i,j+1} - \varpi_{i,j}}{h}, \\ D_0^\zeta (\varpi) &= \frac{\varpi_{i+1,j} - \varpi_{i-1,j}}{2h}, \quad D_0^\psi (\varpi) = \frac{\varpi_{i,j+1} - \varpi_{i,j-1}}{2h}, \quad E_{1/2}^\zeta (H) = \frac{H_{i+1,j} + H_{i,j}}{2}, \quad E_{1/2}^\psi (H_1) = \frac{H_{i,j+1} + H_{i,j}}{2}, \end{aligned} \quad (4.17)$$

Thus, with operators from equation 4.17 the second order derivative is expanded, where for second order derivatives the averaged parameters are found for both ways of the finite difference discretisation.

$$\begin{aligned} D_-^\zeta \left( E_{1/2}^\zeta (H) D_+^\zeta (\varpi) \right) &= \left( \frac{H_{i+1,j} + H_{i,j}}{2h^2} \right) \varpi_{i+1,j} + \left( \frac{H_{i-1,j} + H_{i,j}}{2h^2} \right) \varpi_{i-1,j} + \left( \frac{2H_{i,j} + H_{i+1,j} + H_{i-1,j}}{2h^2} \right) \varpi_{i,j}, \\ D_-^\psi \left( E_{1/2}^\psi (H) D_+^\psi (\varpi) \right) &= \left( \frac{H_{i,j+1} + H_{i,j}}{2h^2} \right) \varpi_{i,j+1} + \left( \frac{H_{i,j-1} + H_{i,j}}{2h^2} \right) \varpi_{i,j-1} + \left( \frac{2H_{i,j} + H_{i,j+1} + H_{i,j-1}}{2h^2} \right) \varpi_{i,j}, \\ D_0^\zeta \left( H D_0^\psi (\varpi) \right) &= \left( \frac{H_{i+1,j}}{4h^2} \right) \varpi_{i+1,j+1} + \left( \frac{H_{i-1,j}}{4h^2} \right) \varpi_{i-1,j-1} - \left( \frac{H_{i+1,j}}{4h^2} \right) \varpi_{i+1,j-1} - \left( \frac{H_{i-1,j}}{4h^2} \right) \varpi_{i-1,j+1}, \\ D_0^\psi \left( H D_0^\zeta (\varpi) \right) &= \left( \frac{H_{i,j+1}}{4h^2} \right) \varpi_{i+1,j+1} + \left( \frac{H_{i,j-1}}{4h^2} \right) \varpi_{i-1,j-1} - \left( \frac{H_{i,j+1}}{4h^2} \right) \varpi_{i-1,j+1} - \left( \frac{H_{i,j-1}}{4h^2} \right) \varpi_{i+1,j-1}, \end{aligned} \quad (4.18)$$

## 4.4 SBP: Absorbing Boundary Layer

Absorbing boundary layer are already a computational domain tool to mitigate spurious reflections from the artificial edge of the modelled region and are not a physical phenomenon. Thus, PML can be applied directly to the spatial derivatives of the computational co-ordinates in the same manner the forward modelling of a wave equation for a flat-surface model region (equation 3.24).

$$\frac{\partial}{\partial \zeta} \rightarrow \frac{1}{\dot{S}_\zeta} \frac{\partial}{\partial \zeta}, \quad \frac{\partial}{\partial \psi} \rightarrow \frac{1}{\dot{S}_\psi} \frac{\partial}{\partial \psi}, \quad (4.19)$$

Fortunately, the inward shifting of coefficients for the outer derivatives of the SBP formulation of second order derivatives can be applied the same way as that for the absorbing boundary layer.

$$\begin{aligned} \rho \frac{\partial^2 u}{\partial t^2} = & \frac{\partial}{\partial \zeta} \left[ \frac{H_1}{\dot{S}_\zeta^2} \frac{\partial u}{\partial \zeta} \right] + \frac{\partial}{\partial \zeta} \left[ \frac{H_2}{\dot{S}_\zeta \dot{S}_\psi} \frac{\partial u}{\partial \psi} \right] + \frac{\partial}{\partial \zeta} \left[ \frac{H_3}{\dot{S}_\zeta^2} \frac{\partial v}{\partial \zeta} \right] + \frac{\partial}{\partial \zeta} \left[ \frac{H_4}{\dot{S}_\zeta \dot{S}_\psi} \frac{\partial v}{\partial \psi} \right] \\ & + \frac{\partial}{\partial \psi} \left[ \frac{H_2}{\dot{S}_\zeta \dot{S}_\psi} \frac{\partial u}{\partial \zeta} \right] + \frac{\partial}{\partial \psi} \left[ \frac{H_5}{\dot{S}_\psi^2} \frac{\partial u}{\partial \psi} \right] + \frac{\partial}{\partial \psi} \left[ \frac{H_6}{\dot{S}_\zeta \dot{S}_\psi} \frac{\partial v}{\partial \zeta} \right] + \frac{\partial}{\partial \psi} \left[ \frac{H_7}{\dot{S}_\psi^2} \frac{\partial v}{\partial \psi} \right], \end{aligned} \quad (4.20a)$$

$$\begin{aligned} \rho \frac{\partial^2 v}{\partial t^2} = & \frac{\partial}{\partial \zeta} \left[ \frac{H_8}{\dot{S}_\zeta^2} \frac{\partial v}{\partial \zeta} \right] + \frac{\partial}{\partial \zeta} \left[ \frac{H_9}{\dot{S}_\zeta \dot{S}_\psi} \frac{\partial v}{\partial \psi} \right] + \frac{\partial}{\partial \zeta} \left[ \frac{H_3}{\dot{S}_\zeta^2} \frac{\partial u}{\partial \zeta} \right] + \frac{\partial}{\partial \zeta} \left[ \frac{H_{10}}{\dot{S}_\zeta \dot{S}_\psi} \frac{\partial u}{\partial \psi} \right] \\ & + \frac{\partial}{\partial \psi} \left[ \frac{H_9}{\dot{S}_\zeta \dot{S}_\psi} \frac{\partial v}{\partial \zeta} \right] + \frac{\partial}{\partial \psi} \left[ \frac{H_{11}}{\dot{S}_\psi^2} \frac{\partial v}{\partial \psi} \right] + \frac{\partial}{\partial \psi} \left[ \frac{H_{12}}{\dot{S}_\zeta \dot{S}_\psi} \frac{\partial u}{\partial \zeta} \right] + \frac{\partial}{\partial \psi} \left[ \frac{H_7}{\dot{S}_\psi^2} \frac{\partial u}{\partial \psi} \right], \end{aligned} \quad (4.20b)$$

The tools used for CPML as described in sub-chapter 3.3.2, is also applied in the computational domain modelling for the topographic region, the complex stretching factor, frequency attenuation and damping operator are applied in the computational co-ordinates  $\zeta$  and  $\psi$ . The formulation for the damping coefficient  $d_0$  is the same as that in equation 3.27b.

$$\dot{S}_\zeta = 1 + \frac{d_\zeta^*}{\dot{\alpha}_\zeta + i\omega}, \quad \dot{S}_\psi = 1 + \frac{d_\psi^*}{\dot{\alpha}_\psi + i\omega}, \quad (4.21)$$

$$\dot{\alpha}_\zeta = \dot{\alpha}_{\max} \left( 1 - \frac{|L_{pml} - L|}{L_{pml}} \right), \quad \dot{\alpha}_\psi = \dot{\alpha}_{\max} \left( 1 - \frac{|L_{pml} - L|}{L_{pml}} \right), \quad (4.22)$$

$$d_\zeta^* = d_0 \left( \frac{|L_{pml} - L|}{L_{pml}} \right)^2, \quad d_\psi^* = d_0 \left( \frac{|L_{pml} - L|}{L_{pml}} \right)^2, \quad (4.23)$$

Utilising the same generalised terms  $\varpi$  and  $H$  in the previous sub-chapter, the discretisation example of applying the PML to the wave equation 4.18 shows that the averaging operator is applied as well to the complex stretching factor  $\dot{S}$ .

$$\begin{aligned}
D_-^\zeta(E_{1/2}^\zeta(H)D_+^\zeta(\varpi)) &= \left(\frac{1}{2h^2}\left(\frac{H_{i+1,j}}{\dot{S}_{i+1,j}} + \frac{H_{i,j}}{\dot{S}_{i,j}}\right)\right)\varpi_{i+1,j} + \left(\frac{1}{2h^2}\left(\frac{H_{i-1,j}}{\dot{S}_{i-1,j}} + \frac{H_{i,j}}{\dot{S}_{i,j}}\right)\right)\varpi_{i-1,j} + \left(\frac{1}{2h^2}\left(\frac{H_{i+1,j}}{\dot{S}_{i+1,j}} + \frac{H_{i-1,j}}{\dot{S}_{i-1,j}} + \frac{2H_{i,j}}{\dot{S}_{i,j}}\right)\right)\varpi_{i,j}, \\
D_-^\psi(E_{1/2}^\psi(H)D_+^\psi(\varpi)) &= \left(\frac{1}{2h^2}\left(\frac{H_{i,j+1}}{\dot{S}_{i,j+1}} + \frac{H_{i,j}}{\dot{S}_{i,j}}\right)\right)\varpi_{i,j+1} + \left(\frac{1}{2h^2}\left(\frac{H_{i,j-1}}{\dot{S}_{i,j-1}} + \frac{H_{i,j}}{\dot{S}_{i,j}}\right)\right)\varpi_{i,j-1} + \left(\frac{1}{2h^2}\left(\frac{H_{i,j+1}}{\dot{S}_{i,j+1}} + \frac{H_{i,j-1}}{\dot{S}_{i,j-1}} + \frac{2H_{i,j}}{\dot{S}_{i,j}}\right)\right)\varpi_{i,j}, \\
D_0^\zeta(H D_0^\psi(\varpi)) &= \left(\frac{1}{4h^2}\frac{H_{i+1,j}}{\dot{S}_{i+1,j}}\right)\varpi_{i+1,j+1} + \left(\frac{1}{4h^2}\frac{H_{i-1,j}}{\dot{S}_{i-1,j}}\right)\varpi_{i-1,j-1} - \left(\frac{1}{4h^2}\frac{H_{i+1,j}}{\dot{S}_{i+1,j}}\right)\varpi_{i+1,j-1} - \left(\frac{1}{4h^2}\frac{H_{i-1,j}}{\dot{S}_{i-1,j}}\right)\varpi_{i-1,j+1}, \\
D_0^\psi(H D_0^\zeta(\varpi)) &= \left(\frac{1}{4h^2}\frac{H_{i,j+1}}{\dot{S}_{i,j+1}}\right)\varpi_{i+1,j+1} + \left(\frac{1}{4h^2}\frac{H_{i,j-1}}{\dot{S}_{i,j-1}}\right)\varpi_{i-1,j-1} - \left(\frac{1}{4h^2}\frac{H_{i,j+1}}{\dot{S}_{i,j+1}}\right)\varpi_{i-1,j+1} - \left(\frac{1}{4h^2}\frac{H_{i,j-1}}{\dot{S}_{i,j-1}}\right)\varpi_{i+1,j-1},
\end{aligned} \tag{4.24}$$

## 4.5 SBP: Other Boundary conditions, and Source

Likewise, to the computational co-ordinates' application of CPML onto the wave equation, the boundary condition is also performed using the computational co-ordinates, where upon utilising the Neumann boundary condition, the ghost points beyond the left, right, top and bottom edge are equivalent to the first internal node of their respective position.

$$\begin{aligned}
\frac{\partial \varpi}{\partial \zeta} &= \frac{\varpi_{1,j} - \varpi_{-1,j}}{2\partial \zeta} = 0 & \therefore \varpi_{1,j} &= \varpi_{-1,j}, \\
\frac{\partial \varpi}{\partial \zeta} &= \frac{\varpi_{n_\zeta+1,j} - \varpi_{n_\zeta-1,j}}{2\partial \zeta} = 0 & \therefore \varpi_{n_\zeta+1,j} &= \varpi_{n_\zeta-1,j}, \\
\frac{\partial \varpi}{\partial \psi} &= \frac{\varpi_{i,1} - \varpi_{i,-1}}{2\partial \psi} = 0 & \therefore \varpi_{i,1} &= \varpi_{i,-1}, \\
\frac{\partial \varpi}{\partial \psi} &= \frac{\varpi_{i,n_\psi+1} - \varpi_{i,n_\psi-1}}{2\partial \psi} = 0 & \therefore \varpi_{i,n_\psi+1} &= \varpi_{i,n_\psi-1},
\end{aligned} \tag{4.25}$$

The total horizontal co-ordinate points  $n_\zeta$  and total vertical co-ordinate points  $n_\psi$  are used to denote the node of the right and bottom edge, respectively. The node of the ghost points are then substituted with the nearest internal node, and can be directly applied to equation 4.24 of the CPML discretisation.

For the free surface boundary condition located at the top edge of the modelled region, equation 3.29 is used to describe the omission of vertical spatial derivative explained before in sub-chapter 3.3.4, where it results in the absence of vertical wavefield, and the purely horizontal displacement elastic wave equation 3.2.

$$\rho \frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial x} \left( \frac{\partial \sigma_{xx}}{\partial t} \right) \rightarrow \frac{\partial}{\partial x} \left( (\lambda + 2\mu) \frac{\partial u}{\partial x} \right), \tag{4.26}$$

The SBP formulation applied to the horizontal displacement wave equation on the top of the modelled region is thus simple compared to that of equation 4.13a.

$$\rho \frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial \zeta} \left[ (\lambda + 2\mu) \zeta_x^2 \frac{\partial u}{\partial \zeta} \right] + \frac{\partial}{\partial \zeta} \left[ (\lambda + 2\mu) \zeta_x \psi_x \frac{\partial u}{\partial \psi} \right] + \frac{\partial}{\partial \psi} \left[ (\lambda + 2\mu) \zeta_x \psi_x \frac{\partial u}{\partial \zeta} \right] + \frac{\partial}{\partial \zeta} \left[ (\lambda + 2\mu) \psi_x^2 \frac{\partial u}{\partial \zeta} \right], \quad (4.27)$$

On the top layer the CPML layer can be omitted, and the SBP discretisation of equation 4.27, is thus demonstrated below:

$$\begin{aligned} \frac{\partial}{\partial \zeta} \left[ (\lambda + 2\mu) \zeta_x^2 \frac{\partial u}{\partial \zeta} \right] &= D_-^\zeta \left( E_{1/2}^\zeta \left( (\lambda + 2\mu) \zeta_x^2 \right) D_+^\zeta (u) \right), \\ \frac{\partial}{\partial \zeta} \left[ (\lambda + 2\mu) \zeta_x \psi_x \frac{\partial u}{\partial \psi} \right] &= D_0^\zeta \left( \left( (\lambda + 2\mu) \zeta_x \psi_x \right) D_0^\psi (u) \right), \\ \frac{\partial}{\partial \psi} \left[ (\lambda + 2\mu) \zeta_x \psi_x \frac{\partial u}{\partial \zeta} \right] &= D_0^\psi \left( \left( (\lambda + 2\mu) \zeta_x \psi_x \right) D_0^\zeta (u) \right), \\ \frac{\partial}{\partial \zeta} \left[ (\lambda + 2\mu) \psi_x^2 \frac{\partial u}{\partial \zeta} \right] &= D_-^\psi \left( E_{1/2}^\psi \left( (\lambda + 2\mu) \psi_x^2 \right) D_+^\psi (u) \right), \end{aligned} \quad (4.28)$$

## 4.6 Summary

This chapter has provided a brief comment on the effects of topography and why it is necessary to take these into account. This chapter also outlined the body-fitted grid generation via the iterative use of Poisson equations (Rao and Wang, 2013), and the SBP formulation of deriving the computational domain spatial derivatives that was stable for use in finite-difference modelling of wave propagation (Rao and Wang, 2018).

The application of source for the computational co-ordinate forward modelling of topographic regions are applied the same way as that of equation 3.35 in sub-chapter 3.4. This is because the number of elements within the computational co-ordinate format of source field  $\mathbf{f}$ , is the same as that of the number of elements used in the source vector for conventional forward modelling.

Techniques of forward modelling and FWI for models with topography are the same as those for conventional flat-surface forward modelling and FWI, whether in the frequency domain in the form of linear algebra solver or a time-iterative method for the time domain. Thus, the singular notable difference is the generation of body grid scheme, and the interpretation of the wave equation to computational co-ordinates. However, despite the belief that the generation of  $d_{syn}$  for flat and non-flat surfaces are the same, inversion of certain features such as faults and vertical fractures may influence the result especially considering that these features should be represented in the computational co-ordinate for modelling the wave equation.

# Chapter 5: Advanced FWI

While the processes of the FWI been well established, there are still many concerns that causes it to fail, including the typical local minima problem. To reiterate, because FWI is essentially non-linear, meaning that solutions generated do not follow a single directional trend throughout the inversion iteration e.g. downwards or upwards, it builds troughs and peaks throughout the iterations. The lowest trough and the highest peak are respectively the global minima and global maxima of the solution within inversion that corresponds to the smallest and biggest differences between observed data and synthetic data when referring to data residuals. Smaller peaks and troughs are called the local maxima and minima, the aim of inversion is to generate the models that characterizes the physical properties of the subsurface media truthfully, i.e. global minima. The local minima problem occurs when the current solution is in the vicinity of a local minima where subsequent iterations generate solutions closer to the local minima. However, for the data misfit to reach the global minima, it must increase in value, which is prevented by the FWI, and hence the data misfit is ‘trapped’ in the local minima. One of the causes of the local minima is cycle-skipping, where mismatch in phases between the synthetic and observed data is greater than half-cycle (wavelength) or  $180^\circ$ , this implies a good match has been found between the two data but there are misaligned in time (Warner and Guasch, 2016). Hence cycle-skipping would cause erroneous updates in the model, there are many causes of cycle -skipping including but not limited to; highly inaccurate starting model, lack of low frequency data, and unaccounted artifacts that occur within the model such as fractures and formations. Another problem of FWI is amplitude discrepancy that can occur when amplitudes of the two data greatly differs that FWI would ultimately invert for a highly erroneous model. Causes of amplitude discrepancy includes the presence of noise in observed data that is non-existent in synthetic data, or inaccurate wavelet selections, these two can be usually remedied by denoising during seismic processing and utilising an appropriate wavelet.

Hence cycle-skipping continues to be the main issue when designing FWI, this chapter aims to inform of ongoing research of advanced forms of FWI such as Optimal Transport (Engquist and Yang, 2020), Adaptive Waveform Inversion -AWI (Warner and Guasch, 2016), Wavefield Reconstruction Inversion -WRI (Van Leeuwen and Herrmann, 2013), and Reflection Waveform Inversion -RWI (He et al., 2019).

## 5.1 Optimal Transport

In contrast to FWI when conducting seismic inversion where the residual is the metric based on which to update the model of physical properties, the optimal transport method takes the view of mapping the synthetic data to the observed data, where it aims to find the most efficient method of rearranging the synthetic data into the observed data. Originally introduced by Monge, 1781 with a ‘sand pile and hole’ setting, with this imagery for optimal transport in FWI there are multiple routes (maps) it takes to move

the sand from the sand-pile (synthetic data) to the hole (data set identical to the observed data), these maps incur different transportation costs (cost function) from one another, where the optimal map is the map with the lowest cost (Yang et al, 2018). It should be noted that the energy of the synthetic data given by the starting model should be the same as that of the observed data. The Wasserstein metric was introduced by Villani, 2003 to calculate the lowest cost, and optimal transport theory was first applied to the seismic inversion by Engquist and Froese, 2014, where general quadratic Wasserstein metric  $W_2$ , replaces misfit function.

$$W_2^2(f(x_e), g(y_e)) = \inf_{T \in M_T} \int_{X_e} |x_e - T(x_e)|^2 f(x_e) dx_e, \quad (5.1)$$

The transport plan  $T(x_e)$  represents the shift of the Euclidean position  $x_e$ , of an element within the function  $f(x_e)$ , to the Euclidean position  $y_e$ , of an element within the function  $g(y_e)$ , as shown in equation 5.2, where it is given that the position  $y_e$ , must be obtained from the transport plan  $T$ , being applied to position  $x_e$ . In the context of FWI functions  $f(x_e)$  and  $g(y_e)$  represent  $d_{syn}$  and  $d_{obs}$  respectively.  $X_e$  and  $Y_e$  represents positive real terms of  $x_e$  and  $y_e$  respectively. Map  $M_T$  is a set that contains all possible  $T(x_e)$ .

$$\begin{aligned} T : x_e &\rightarrow y_e, \\ (x_e, y_e) &\in X_e \times Y_e \quad \mathbb{R} \{ \cup +\infty \}, \\ T : X_e &\rightarrow Y_e, \end{aligned} \quad (5.2)$$

However, this method would be computationally expensive as it would require assessing the lowest transport cost out of innumerable transport plans or due to requiring an algorithm that has extensive computation (Benamou and Brenier, 2000, Bertsekas, 1992, Bosc, 2010). To deal with this issue, the Monge-Ampère equation was introduced to the optimal transport method to calculate the Wasserstein distance more efficiently (Knott and Smith, 1984, Brenier, 1991). Monge-Ampère equations are 2<sup>nd</sup> order partial differential equations of an unknown function  $u^*$ , with respect to variables  $x_e$  and  $y_e$ , where the determinant of the Hessian matrix of  $u^*$  is linear. Firstly, it must be noted that the relationship between  $f(x_e)$  and  $g(y_e)$  needs to be clarified, and based on equation 5.2 function  $g(y)$  is the same as  $g(T(x_e))$ , and that determinant of the gradient of  $T$  i.e. a Jacobian determinant describes the different local geometrical changes in the square matrix of  $T$  i.e. individual  $T$  for each position  $x_e$  to the corresponding position  $y_e$ .

$$f(x_e) = g(T(x_e)) \det(\nabla T(x_e)), \quad (5.3)$$

In the case of optimal transport  $u^*$  can be seen as a collection of  $x_e, y_e$  pairs, and that the gradients of the two points is equivalent to the transport plan  $T$ , such that  $T(x_e) = \nabla f_u(x_e)$ . Applying this knowledge to equation 5.3 and 5.2 yields the elliptic Monge-Ampère equation (5.4a) and the squared quadratic Wasserstein distance (5.4b) respectively.



$$\begin{aligned}
\det(\partial^2 u^*(x_e)) &= \frac{f(x_e)}{g(\nabla u^*(x_e))} + u_0^*, \quad x_e \in X_e \\
\nabla u^*(X_e) &= Y_e, \\
u^* &\text{ is convex.}
\end{aligned} \tag{5.4a}$$

$$W_2^2(f(x_e), g(y_e)) = \int_{X_e} |x_e - \nabla u^*(x_e)|^2 f(x_e) dx_e, \tag{5.4b}$$

Where  $\partial^2$  represents the second order derivative of  $u^*$ . The solution  $u^*$  contains additional information where the vector  $\nabla u^*(x_e) - x_e$  reveals which parts of the  $f(x_e)$  and  $g(y_e)$  are connected under the optimal transport map (Engquist and Froese, 2014). The L.H.S. of first formula in equation 5.4a i.e.  $\det(\partial^2 u^*(x_e))$  is the determinant of the Hessian matrix for  $u^*$ .  $u_0^*$  equates to difference between the determinant and the function division in first formula of equation 5.4a to ensure equivalency in the equation.

$$\det(\partial^2 u^*(x_e)) = \frac{\partial^2 u^*}{\partial x_e^2} \frac{\partial^2 u^*}{\partial y_e^2} - \left( \frac{\partial^2 u^*}{\partial y_e \partial x_e} \right)^2, \tag{5.5}$$

Methods to solve the first equation in 5.4a and eventually attain  $\nabla u^*(x_e)$  can be like that of the forward modelling method, such as the explicit iteration method, projection method, and the Newton method (Benamou et al, 2014). To strictly enforce the squared quadratic Wasserstein distance to follow a convex function, that is for it to have no more than one minimum i.e. contain only global minimum, it was suggested to utilise an almost-monotone finite difference modelling (FDM) formulation of the Monge-Ampère equation which is discretised in a mesh-like grid co-ordinate system (Froese, 2012, Yang et al, 2018). The almost monotone formulation utilises the max math notation  $\max\{\}$ , to enforce the function to be either entirely non-increasing or entirely non-decreasing. FDM formulation involves calculation of the almost monotone approximation  $M_F[u^*]$ , using the 2<sup>nd</sup> order approximation of the Monge-Ampère equation  $M_N[u^*]$ , which is essentially the first formula in 5.4a and monotone approximation  $M_M[u^*]$ . The monotone approximation is the minimum between approximation of FDM Monge-Ampère equation aligned to grid axes  $MA_1[u^*]$ , and those aligned to stencil corners  $MA_2[u^*]$ .

$$\begin{aligned}
MA_1[u^*] &= \max \left\{ \frac{\partial^2 u^*}{\partial x_1^2}, \delta^* \right\} \max \left\{ \frac{\partial^2 u^*}{\partial x_2^2}, \delta^* \right\} + \min \left\{ \frac{\partial^2 u^*}{\partial x_1^2}, \delta^* \right\} + \min \left\{ \frac{\partial^2 u^*}{\partial x_2^2}, \delta^* \right\} \\
&\quad - f / g \left( \frac{\partial u^*}{\partial x_1}, \frac{\partial u^*}{\partial x_2} \right) - u_0^*,
\end{aligned} \tag{5.6a}$$

$$MA_2[u^*] = \max \left\{ \frac{\partial^2 u^*}{\partial l^2}, \delta^* \right\} \max \left\{ \frac{\partial^2 u^*}{\partial l_\perp^2}, \delta^* \right\} + \min \left\{ \frac{\partial^2 u^*}{\partial l^2}, \delta^* \right\} + \min \left\{ \frac{\partial^2 u^*}{\partial l_\perp^2}, \delta^* \right\} - f / g \left( \frac{1}{\sqrt{2}} \left( \frac{\partial u^*}{\partial l} + \frac{\partial u^*}{\partial l_\perp} \right), \frac{1}{\sqrt{2}} \left( \frac{\partial u^*}{\partial l} - \frac{\partial u^*}{\partial l_\perp} \right) \right) - u_0^*, \quad (5.6b)$$

$$M_M[u^*] = -\min \{ MA_1[u^*], MA_2[u^*] \}, \quad (5.6c)$$

$$M_N[u^*] = - \left( \frac{\partial^2 u^*}{\partial x_1^2} \frac{\partial^2 u^*}{\partial x_2^2} - \left( \frac{\partial^2 u^*}{\partial x_1 \partial x_2} \right)^2 \right) - f / g \left( \frac{\partial u^*}{\partial x_1}, \frac{\partial u^*}{\partial x_2} \right) - u_0^* = 0, \quad (5.6d)$$

$$M_F[u^*] = M_M[u^*] + \varepsilon S_f \left( \frac{M_N[u^*] - M_M[u^*]}{\varepsilon} \right), \quad (5.6e)$$

The mesh-spatial step  $\partial x_1$  and  $\partial x_2$  denotes approximation along the horizontal and vertical directions respectively and represents the resolution of the grid. The diagonal mesh-spatial step  $\partial l$  and  $\partial l_\perp$  denotes approximation along negative and positive slope, respectively. The value for the function  $g$  is chosen via calculating its co-ordinate system  $g(,)$ , small parameter  $\delta^*$ , is used to as a limit to keep the 2<sup>nd</sup> order derivative away from zero, and  $K$  is the Lipschitz constant of  $y_e$  of  $f(x_e)/g(y_e)$ .  $\varepsilon$  is a small parameter to that ensures operated value  $a$ , is in the range of which filter  $S_f$ , can be applicable and is used to scale the filtered results to their original order of magnitude. Discretisation of the FDM formulation involves the spatial differentiation of the function  $u^*_{ij}$ .

$$\delta^* = \frac{K \Delta x_e}{2}, \quad (5.7a)$$

$$S_f(a) = \begin{cases} a & |a| \leq 1 \\ 0 & |a| \geq 2 \\ -a + 2 & 1 \leq a \leq 2 \\ -a - 2 & -2 \leq a \leq -1 \end{cases}, \quad (5.7b)$$

$$\begin{aligned}
\frac{\partial u^*}{\partial x_1} &= \frac{1}{2dx_e} (u_{i+1,j}^* - u_{i-1,j}^*), \\
\frac{\partial u^*}{\partial x_2} &= \frac{1}{2dx_e} (u_{i,j+1}^* - u_{i,j-1}^*), \\
\frac{\partial^2 u^*}{\partial x_1^2} &= \frac{1}{dx_e^2} (u_{i+1,j}^* + u_{i-1,j}^* - 2u_{i,j}^*), \\
\frac{\partial^2 u^*}{\partial x_2^2} &= \frac{1}{dx_e^2} (u_{i,j+1}^* + u_{i,j-1}^* - 2u_{i,j}^*), \\
\frac{\partial^2 u^*}{\partial x_1 \partial x_2} &= \frac{1}{4dx_e^2} (u_{i+1,j+1}^* + u_{i-1,j-1}^* - u_{i+1,j-1}^* - u_{i-1,j+1}^*), \\
\frac{\partial u^*}{\partial l} &= \frac{1}{2\sqrt{2}dx_e} (u_{i+1,j+1}^* - u_{i-1,j-1}^*), \\
\frac{\partial^2 u^*}{\partial l^2} &= \frac{1}{2dx_e^2} (u_{i+1,j+1}^* + u_{i-1,j-1}^* - 2u_{i,j}^*), \\
\frac{\partial^2 u^*}{\partial l_{\perp}^2} &= \frac{1}{2dx_e^2} (u_{i+1,j-1}^* + u_{i-1,j+1}^* - 2u_{i,j}^*),
\end{aligned} \tag{5.7c}$$

Applying equations 5.7a-c to equations 5.6a-e means the finite difference approximation of the Monge-Ampère equation can be solved iteratively or with Newton's Method to retrieve the discrete solution  $u_h$ , (Yang et al, 2018). Thus, Wasserstein distance is the summation of the shifts in  $x_e$  in the form of first order derivative of  $u_h$  with respect to  $x_e$ , i.e.  $\partial_x u_h$  for each data point  $j$  over the whole data dimension  $n$ .

$$W_2^2(f(x_e), g(y_e)) \approx \sum_{j=1}^n (x_j - \partial_{x_j} u_h)^T \text{diag}(f(x_e)) (x_j - \partial_{x_j} u_h), \tag{5.8}$$

To compute the adjoint source, the (formal) Jacobian  $\nabla M_F[u^*]$  of the scheme is required for the linearization of the Monge-Ampère equation, which is obtained when solving equations 5.6a-e via Newton's method (Froese, 2012, Engquist et al, 2016). Once linearized, equation 5.9a can be used in the form of  $Ax^* = b$ , where sparse matrix  $A$  contains  $\nabla M_F[u^*]$ , vector  $b$  contains  $\delta f(x_e)$  denoting perturbation of function  $f(x_e)$ , both are used to retrieve vector  $x^*$  containing  $\delta u^*$ , denoting perturbation of  $u^*$ . The solution of equation 5.9a is then used to solve the gradient of the optimal transport objective function, as well as the potential function  $u_p$ , which must meet the conditions of discretised equation 5.4a in equation 5.9c.

$$\nabla M_F[u^*] \delta u^* = \delta f(x_e), \tag{5.9a}$$

$$\nabla F_{OT}(m^k) = \sum_{j=1}^n \left[ -2\delta u^* + \text{diag}(x_j - \partial_{x_j} u_p) \right] (x_j - \partial_{x_j} u_p), \tag{5.9b}$$

$$M[u_p] = 0, \tag{5.9c}$$

The strength of this method is that it combines the most desirable properties of travel-time differences and  $L_2$ -norm in comparing seismic signals, can be easily applied to higher dimensions, and was found to be more accurate and it takes less computational time for cycle-skipped case, but is just as fast as the standard  $L_2$ -based FWI for non-cycle skipped case (Yang et al, 2018).

However, there are challenges and requirements that makes the Optimal transport less attractive to implement. The Wasserstein metric requires that both  $f(x_e)$  and  $g(y_e)$  to be strictly positive when computed directly, whereas most seismic signals are not, adding a positive constant to shift the two functions to be positive distorts the optimal transport map, utilising the envelope would produce inaccurate results (Engquist and Froese, 2014). Comparing the positive and negative parts of the two functions can yield accurate results but it would essentially double the time to calculate  $W_2$ . The other problem is the requirement of conservation of mass where every information that is in  $f(x_e)$  must also be in  $g(y_e)$ , that is done with the use of the scaling, however despite with scaling this does not reflect the reality that  $f(x_e)$  and  $g(y_e)$  does not always contain the same amount of information. Problems also arises from the solution including a requirement that the set  $Y_e$  to be a convex set, which is achieved by pre-processing the data, however the pre-processing may introduce artificial transport. Another difficulty is that it requires  $f(x_e)/g(y_e)$  to be Lipschitz continuous in the in the  $y_e$  variable and so requires a careful selection of the pre-processing parameter and a regularising kernel (Engquist and Froese, 2014).

## 5.2 Adaptive Waveform Inversion

The motivation for adaptive waveform inversion (AWI) was firstly to ensure a good match between  $d_{syn}$  and  $d_{obs}$  despite starting and true model differences and secondly to expand dimensionality of the unknown model space found to be common in all inversions that overcome cycle-skipping (Warner and Guasch, 2014, 2016). AWI meets the motivations with the use of Wiener filter  $\mathbf{w}$ , a matching convolutional filter that scans the element series of  $d_{syn}$  vector trace  $\mathbf{p}$  shifted by time and the  $d_{obs}$  vector trace  $\mathbf{d}$ , in order to find the best arrangements that reduces the misfit. Therefore AWI “adapts” waveforms of  $\mathbf{p}$  to that of  $\mathbf{d}$  and takes a different approach to data misfit in  $L_2$ -norm FWI, where instead it seeks to minimize the disparity between the filters applied and that of the ideal identity filter (Warner and Guasch, 2014). An identity filter is achieved when the updating model and true earth model matches, where it does not adjust  $d_{syn}$  but allows it to pass through the inversion as the output. This misfit function like the deconvolution objective aims to achieve unity, where  $d_{syn}$  and  $d_{obs}$  are identical, as opposed to zero in conventional  $L_2$ -norm FWI, that signifies absence of data misfit. AWI differs with deconvolution objective function, in that its misfit involves normalization of  $\mathbf{w}$ , where its co-efficient are determined when applied to matrix  $\mathbf{P}$ , if it can reduce the following function  $\phi_{L2-AWI}$ .  $\mathbf{P}$  is the synthetic

Toeplitz matrix where each column contains the same vector  $\mathbf{p}$ , but they differ in that they shift an element; thus each diagonal band represents a single element from  $d_{syn}$  (Warner and Guasch, 2016).

$$\phi_{L2-AWI} = \frac{1}{2} \|\mathbf{P}\mathbf{w} - \mathbf{d}\|^2, \quad (5.10)$$

When  $\mathbf{w}$  is applied to  $\mathbf{P}$  it weighs various arrangements of  $\mathbf{p}$  to the entirety of  $\mathbf{d}$  rather than individual events, and hence because of its overall coverage, equation 5.10 doesn't have multiple local minima but a single global minimum, avoiding the cycle-skipping issues. However, there are two variants of AWI determined by the choice of wiener filters, the conceptually straightforward forward wiener filter  $\mathbf{w}_{for}$ , and the mathematically simple reverse wiener filter  $\mathbf{w}_{rev}$ , represents transformation of  $\mathbf{p}$  to  $\mathbf{d}$ , and  $\mathbf{d}$  to  $\mathbf{p}$  respectively.  $\mathbf{D}$  is the Toeplitz matrix of  $\mathbf{d}$ .

$$\mathbf{w}_{for} = \frac{\mathbf{P}^T \mathbf{d}}{\mathbf{P}^T \mathbf{P}}, \quad (5.11a)$$

$$\mathbf{w}_{rev} = \frac{\mathbf{D}^T \mathbf{p}}{\mathbf{D}^T \mathbf{D}}, \quad (5.11b)$$

Equation 5.11b is like deconvolution objective function equation 2.5 but differs from equation 5.11a where the denominator in the latter involves autocorrelation of  $d_{syn}$  instead of  $d_{obs}$ . Their cross-correlation are alike, however the wiener filter omits the stabilising factor.

$$\phi_{AWI-for} = \frac{1}{2} \frac{\|\mathbf{T}\mathbf{w}_{for}\|^2}{\|\mathbf{w}_{for}\|^2}, \quad (5.12a)$$

$$\phi_{AWI-rev} = \frac{1}{2} \frac{\|\mathbf{T}\mathbf{w}_{rev}\|^2}{\|\mathbf{w}_{rev}\|^2}, \quad (5.12b)$$

The diagonal weighing matrix  $\mathbf{T}$  penalises coefficients of  $\mathbf{w}$  depending upon the magnitude of the time lags (time domain). When the wiener filter is eventually reduced to an identity filter it becomes a unit-amplitude delta function at zero lag, thus AWI objective functions  $\phi_{AWI-for}$  and  $\phi_{AWI-rev}$  attempts to force  $\mathbf{w}$  towards a delta function at zero lag.

$$\frac{\partial \phi_{AWI-for}}{\partial \mathbf{m}} = \mathbf{W}^T \mathbf{P} (\mathbf{P}^T \mathbf{P})^{-1} \left( \frac{2(\phi_{AWI-for}) \mathbf{I} - \mathbf{T}^2}{\mathbf{w}_{for}^T \mathbf{w}_{for}} \right) \mathbf{w}_{for}, \quad (5.13a)$$

$$\frac{\partial \phi_{AWI-rev}}{\partial \mathbf{m}} = \mathbf{D}(\mathbf{D}^T \mathbf{D})^{-1} \left( \frac{\mathbf{T}^2 - 2(\phi_{AWI-rev}) \mathbf{I}}{\mathbf{w}_{rev}^T \mathbf{w}_{rev}} \right) \mathbf{w}_{rev}, \quad (5.13b)$$

Since AWI can use the adjoint source (equation 5.13a and 5.13b) to perform backpropagation, it has the advantage of being able to be performed in sequence with gradients method and optimisation algorithms like that of FWI, making it straightforward to apply.  $\mathbf{I}$  is the identity matrix.

Testing results have demonstrated that AWI is immune to cycle skipping and deals successfully with data differences of one and a half cycles, while FWI can only deal with differences below half a cycle (Guasch et al., 2019). AWI was found to recover models slightly better than FWI when there are limitations in seismic acquisition such as smaller bandwidth, and greater shot-receiver aperture. The reason is that Wiener coefficients at large time lags effected by arrival-time mismatches due to missing reflectors, are penalized in AWI. Hence, AWI can marginally recover deeper regions in models due to its increased sensitivity to reflections at intermediate and long wavelengths. AWI is effective in the presence of incorrect source wavelet, as the convolution of Wiener filters with the source wavelet used to generate  $d_{syn}$  adapts the earth model so that the latter evolves towards the source wavelet that generated  $d_{obs}$  (Warner and Guash, 2016).

Whenever good matching occurs between  $d_{obs}$  and  $d_{syn}$ , phase locking can be used to produce higher fidelity and resolution within the result model. Phase locking essentially fixes an amplitude in one function to the other amplitude in the other function based on phase similarities. Unfortunately, due to unexplained differences in  $d_{obs}$ , phase locking can lead to erroneous results in AWI, it was suggested to start inversion with AWI and change to FWI after several iterations (Warner and Guash, 2016).

One potential problem in AWI is that its higher sensitivity to receiver spacing and finite sources can produce fine-scale acquisition footprints (artefacts) in inverted model. This problem could be resolved with regularisation, however depending on survey geometries in seismic acquisition traces for each  $d_{syn}$  generated in each AWI iterations may need to be closely spaced, thus increasing computational costs (Guasch et al., 2019). Despite AWI seeming imperviousness to the absence of low frequency data, AWI may be bolstered with its inclusion, as low frequency data which helps to recover model details in deeper regions. To find any substantial problems AWI may face, it is prudent to explore various potential survey geometries that may arise in seismic acquisition, and test AWI against other sources of local minima that are as potentially devastating as cycle-skipping.

### 5.3 Wavefield Reconstruction Inversion

Wavefield Reconstruction Inversion (WRI) was proposed to solve the problem of local minima that arise from the non-linear relationship between model parameters and data misfit found in FWI (Van Leeuwen and Herrmann, 2013). The motivation for WRI is to apply ‘full-search space’ approach in the optimisation problem, that involves finding the subsurface model and wavefields that minimizes data misfit, in contrast FWI is a ‘reduced space’ approach because it only utilises the wavefields.

WRI differs to other methods that expand the search space such as ‘contrast-source formulation’ and ‘all-at-once method’, in that it does not require significant storage space and wavefield updates (Van der Berg and Kleinman, 1997, Haber et al, 2000). This is because WRI follows a two steps inversion, where firstly a data-constrained wave equation is solved for each source and a given trade-off parameter  $w$ , and lastly the model is updated with the ‘reconstructed’ wavefields from their corresponding sources. The data-constrained wave equation is drawn from the observation that  $d_{obs}$  is a collection of ‘true’ wavefields recorded at the receiver and is used to ‘conform’ wavefields to the wave equation of the model at the current iteration  $k$ , and for  $d_{obs}$ . This two-step inversion generates an ‘equivalent-medium averaged’ model estimates, where its spatial wavelength content is related to information about the source e.g., bandwidth. Data-constrained wave equation is applied as though it’s a regularisation term but differs where their model estimates term can be considered as an ‘interpolator’ and ‘extrapolator’ respectively (Van Leeuwen and Herrmann, 2013).

The WRI objective function  $\phi_{WRI}$ , builds upon the  $L_2$ -norm objective function with the addition of the data-constrained wave equation where it finds the difference between the sparse wave equation matrix  $\mathbf{S}$  applied to the wavefield vector  $\mathbf{u}$ , and the synthetic data source vector  $\mathbf{f}_{syn}$ .  $\phi_{WRI}$  is calculated for each receiver  $r$ , and source  $s$ , throughout all receivers  $NR$ , and all sources  $NS$ .

$$\phi_{WRI}^k = \sum_{s=1}^{NS} \sum_{r=1}^{NR} \left( \frac{1}{2} \|d_{syn} - d_{obs}\|_2^2 + \frac{w^2}{2} \|\mathbf{S}(m^k) \mathbf{u} - \mathbf{f}_{syn}\|_2^2 \right), \quad (5.14)$$

$\phi_{WRI}$  is minimized with respect to both model parameters  $m^k$ , and wavefields  $\mathbf{u}$ , to perform the two-step inversion and avoid storing wavefields the ‘augmented wave equation’ (equation 5.15) is formulated when the derivative of  $\phi_{WRI}$  with respect to  $\mathbf{u}$  equates to zero. In forward modelling, extraction operator  $P$ , is used to extract data  $\mathbf{u}$  at the receivers’ locations to generate  $d_{syn}$ . As stated before, based on the knowledge that  $d_{obs}$  is derived from applying  $P$  onto the ‘true’ wavefield  $\mathbf{u}$ , the linear algebra expression  $Ax^*=b$ , can be used to derive  $\mathbf{u}$ , where  $P$  is represented as sparse matrix  $A$ , and  $d_{obs}$  is represented as vector  $b$ , the vector  $x^*$  for  $\mathbf{u}$  can be solve for.

$$\begin{pmatrix} P \\ w\mathbf{S}(m^k) \end{pmatrix} \mathbf{u} = \begin{pmatrix} d_{obs} \\ w\mathbf{f}_{obs} \end{pmatrix}, \quad (5.15)$$

Then  $\mathbf{u}$  can be applied by the regularized version of the wave equation  $w\mathbf{S}$ , of model parameters  $m$  at current iteration  $k$ , to achieve the regularised version of the observed data source vector  $w\mathbf{f}_{obs}$ . The gradient for WRI is then derived as the derivative of the  $\phi_{WRI}$  with respect to the model, whilst  $\mathbf{u}$  is fixed, hence WRI does not require backpropagation of the adjoint source.

$$\nabla F_{WRI}(\mathbf{m}^k) = \frac{\partial \phi_{WRI}}{\partial \mathbf{m}} = \sum_{s=1}^{NS} \sum_{r=1}^{NR} \omega^2 \text{diag}(\mathbf{u}_{s,r})^* (\mathbf{S}_{s,r}(\mathbf{m}^k) \mathbf{u} - \mathbf{f}_{syn}), \quad (5.16)$$

The WRI gradient involves correlation of  $\mathbf{u}$  solved at equation 5.15 with the wave equation residuals, it can be immediately applied as an update. Accumulations of these correlation terms over the receivers and sources eliminates the need to store wavefields, reducing the memory storage cost of WRI while taking advantage of the full search space, and help reduce the steps taken for inversion.

Another strength of WRI is that because it tries to minimise the misfit over both models and wavefields i.e. larger search space, it causes each step to be mostly linear, meaning it is less prone to local minima arising from model differences.

In addition to relaxation of data fitting, such as that of the least squares' formulation in  $L_2$ -FWI, the WRI also relaxes the physics of the wave equation in the least-squares form. The two relaxations give the advantage of reducing influences of spurious data matching i.e. incorrect amplitude misfit due to phase mismatch, and erroneous data arising out of incorrect application of wave equation applied to inaccurate models. This removes certain sources of local minima, allowing WRI to be less sensitive to imprecise starting models and move towards global minima (Van Leeuwen and Herrmann, 2013).

For conventional starting model WRI was found to be no more effective in recovering the true model than FWI, despite the two-step efficiency and resistance to certain local minima (Symes, 2020). It was emphasised that WRI may only be sufficient to mitigate some local minima but not all (Van Leeuwen and Herrmann, 2013).

Mere smoothness of objective functions can impose restriction on wave equation operators, where the data does not accurately represent wavefields generated by a propagating wave, thus it impedes stable and reliable recovery of model parameters, this predicament was not found in FWI (Symes, 2020).

## 5.4 Reflection Waveform Inversion

Reflection Waveform Inversion (RWI) aims to recover background velocities throughout the whole model including at deeper depths utilising reflection waves with both short and long offsets, whereas FWI relies on refraction and turning waves to produce background velocities (Yao et al, 2020). Background velocity recovery with refraction and turning waves are restricted by seismic acquisition



geometries, where the depth travelled by these waves travel is limited. The High resolution of near surface interfaces in the inverted model can be attributed to the reflection waves in short offset, however utilising reflection waves at longer offsets would retrieve background velocities at lower depths (Yao et al, 2020).

A fundamental step in RWI is the separation of migration and tomography components of the model update, where the former represents refraction and turning waves, is stronger than the latter (reflection waves) in FWI.

There are three approaches to achieve separation of migration and tomography components, up/down wavefield separation, scattering angle filter, and born modelling (Yao et al, 2020). Wavefield separation involves applying a criterion to a wavefield in frequency-wavenumber domain. In the scattering angle filter approach, the two components are distinguished by the scattering angle generated at the point of meeting between the wavefields. The born modelling approach involves a couple modelling for incident and scattering wavefields.

A problem for the latter two approach is the computational costs, especially for the born modelling in frequency domain, and of the three methods of the scattering angle filter approach some of them are unsuitable for anisotropic models.

Of the three approach, this thesis recommends the up/down wavefield separation as it is believed to be the simplest to implement as it only involves a single additional spatial Fourier transform, except for certain circumstances discussed later in this sub-chapter.

The gradient itself is split into the migration component  $\nabla F_{mig}(m^k)$ , and the tomography component  $\nabla F_{tomo}(m^k)$ , which in turn are formed from the split wavefield components of the forward source wavefield and the backpropagated residual wavefield. Demonstrating in equation 5.17a and 5.17b with the  $L_2$ -norm gradient (equation 2.31), the gradient component  $\nabla F_{mig}(m^k)$  and  $\nabla F_{tomo}(m^k)$  are formed with summation of split wavefields of opposing and same directions, respectively. The wavefield is split according to vertical directions, i.e. both forward wavefield and backpropagated wavefield is split for upwards  $u_{fu}(k_s, \omega)$ ,  $u_{bu}(k_s, \omega)$  respectively and downwards  $u_{fd}(k_s, \omega)$ ,  $u_{bd}(k_s, \omega)$  respectively.

$$\nabla F_{mig}(m) = \frac{2\omega^2}{V^3} \sum_{\omega} \sum_s (u_{fu}(k_s, \omega)u_{bd}(k_r, \omega) + u_{fd}(k_s, \omega)u_{bu}(k_r, \omega)), \quad (5.17a)$$

$$\nabla F_{tomo}(m) = \frac{2\omega^2}{V^3} \sum_{\omega} \sum_s (u_{fu}(k_s, \omega)u_{bu}(k_r, \omega) + u_{fd}(k_s, \omega)u_{bd}(k_r, \omega)), \quad (5.17b)$$

To obtain the split wavefield Fourier transform is applied along the spatial domain  $x$ , to return the spatial angular frequency  $k$ . Criteria to distinguish upwards  $u_+$  (5.18a) and downwards  $u_-$  (5.18b) wavefields generated at each spatial angular frequency  $\omega$ , are based on  $k$  along depth  $z$ , (Hu and McMechan, 1987).

$$u_+(k_z, \omega) = \begin{cases} u(k_z, \omega) & k_z \geq 0 \\ 0 & k_z < 0 \end{cases} \quad (5.18a)$$

$$u_-(k_z, \omega) = \begin{cases} 0 & k_z \geq 0 \\ u(k_z, \omega) & k_z < 0 \end{cases} \quad (5.18b)$$

Basis for equation 5.18a-b is based upon assuming upward wavefield indicated by positive wavenumber and downward wavefield indicated by negative wavenumber, corresponding to the reflection and incident waves, respectively. Unfortunately, the presence of large dipping angles can cause wavefield separation to fail (Liu et al, 2011), where tomography and migration components can contain elements from each other, and lead to highly erroneous velocity updating in RWI (Wang et al, 2018). One of the solutions to the large dipping angle problem is to split the wave equation along horizontal and vertical axis, followed by applying equation 5.17 to obtain their respective upward and downward wavefields (Irabor and Warner, 2016), however like conventional PML absorbing boundary layer this involves a non-physical splitting of the wave equation, and in frequency domain would require extra computation of wavefield. To avoid the non-physical splitting, the discrete Hilbert transform can be applied to allow decomposition in the horizontal direction (Li et al, 2019).

It can be noted that the same equation 5.17 for RWI components decomposition can also be applied for reverse time migration (RTM) to remove background noise, with the exceptions that it applied complex conjugate to the source's upwards and downwards wavefield, emphasises on the migration component and treats the tomography component as background noises (Liu et al, 2011).

The procedure for RWI overlaps that of FWI in that they both can use the same objective functions and optimisation method, except for an additional step of wavefield decomposition in the gradient. FWI treat migration and tomography component with equal weight, but the aim of RWI is to recover a smooth background model. A weighting co-efficient  $w$  can be applied in the gradient for RWI to decide which of the two gradient components should take precedence in RWI, values suggested can be found from Tang et al, (2013). It is prudent to assign high values for  $w$  initially to recover the smooth background velocity, but as the inversion iteration progresses it should gradually diminish to recover high frequency features in the inverted model.

$$\nabla F(m^k) = \nabla F_{mig}(m^k) + (w) \nabla F_{tomo}(m^k), \quad (5.19)$$

Whilst RWI does not specifically calculate the misfit function it could be applied in conjunction with the other forms of FWI.

## 5.5 Summary

In this chapter we have covered the topics of four promising forms of waveform inversion of optimal transport, adaptive waveform inversion, wavefield reconstruction inversion and reflection waveform inversion. Each of these waveform inversions can be considered as having their own ‘philosophy’ of successfully inverting seismic data. Optimal transport inversion can be considered as ‘rearranging’ information of the synthetic data to be the same as that of the observed seismic data. Adaptive waveform inversion with its immunity to cycle-skipping can be thought as ‘molding and cropping’ the synthetic data towards the observed data. Wavefield reconstruction inversion attempts to derive misfits with the inclusion of wavefield differences, by ‘rebuilding’ the wavefields that would have formed the observed data set. Reflection waveform inversion attempts to invert the reflection wave data specifically, which involves gradient ‘splitting’ into tomography and migration components which themselves are formed of downward and upward wavefields.

Key advantages of optimal transport include generating a more accurate model solution for cycle-skipped case in less computational time, as well as its applicability in higher dimensions (Yang et al, 2018). A main concern for optimal transport is that the energy within  $d_{syn}$  generated from the starting model must equal to  $d_{obs}$ , which could mean that the initial model input must be fairly accurate (Engquist and Froese, 2014). In the case of cycle skipping AWI would be better than optimal transport due to its immunity to this phenomena (Warner and Guash, 2016). The sensitivity of AWI allows it to marginally recover deeper regions in models but also leaves it dependent on receiver spacing and vulnerable to acquisition footprints in the solution model. This could be resolved with generating more closely spaced  $d_{syn}$ , at a price of increased computational costs (Guasch et al., 2019). WRI has lower memory storage costs as it does not need to store wavefields can lead to lower computational costs, this is due to its use of the full search space approach, which allows it perform inversion in fewer but more linear steps (Van Leeuwen and Herrmann, 2013). The relaxation of data fitting and the physics of the wave equation allows WRI to be less sensitive to inaccurate starting models. However, for conventional starting model WRI was found to be no more effective than FWI and smoothing of misfit functions may result in incorrect  $d_{syn}$ , which impedes the recovery of model parameters, this predicament was not found in FWI (Symes, 2020). RWI attempts to circumvent the problem associated with limited geometric acquisition that affects AWI by utilising reflection waves to recover background velocity model (Yao et al, 2020), which in turn can help generate a more accurate  $d_{syn}$ . Of the three approaches of RWI, the born modelling and scattering angle filter have high computational costs, whereas the wavefield separation approach requires Hilbert transform to avoid non-physical splitting and careful selection of weighting co-efficient (Yao et al, 2020, Li et al, 2019). Where each method has their own merits and limitations, the authors of these advanced FWI method do not claim these can supplant FWI and instead suggested to perform these methods with FWI in an iterative scheme.

# Conclusion

A brief guide to FWI has been largely covered in this thesis, and the simplest implementation method has been outlined in chapter 2, where objective functions perform a central role in seismic inversion. Examples of objective functions given include the conventional least-squares norm, cross-correlation, deconvolution, and envelope objective functions, where they consider different features of the waveforms of the seismic data, with the aim of reaching the global minima. However, it should be noted that there is still ongoing research into other forms of misfit functions in practice and theory, and in the future research these will be looked into.

The motivation for ‘joint deconvoluted envelope and phase residual misfit function’ was to combine the features of the many misfit function mentioned in chapter 2, and exploit advantages of these features, for the amplitudes and phases of the seismic data to honour the definition of FWI. Waveform envelope is used to account for the low frequency that would be absent in conventional FWI and retrieve the smooth background velocity model,  $L_2$ -norm of the phases tries to tackle the cycle skipping issues. The deconvolution features allow the confinement of energy to zero-time lag and aims to achieve unity as a means to measure matching data matching. However, despite the theoretical specialisation of the proposed objective function, it was likewise suggested, as other misfit-based FWI, to apply conventional  $L_2$ -norm FWI following inversion with proposed misfit functions. Chapter 2 briefly covers regularisation and outlines the steps and methods throughout FWI optimisation methods to generate the model gradient, search direction, and step-length, necessary for the model update.

The importance of the forward modelling for a successfully inversion cannot be overstated enough. Chapter 3 outlined the various wave equation that is used in synthetic data generation, from elastic and acoustic to including viscous and porous properties, and briefly mentions  $Q$ -factor, anisotropy, and seismic wavelets, all of which covers a broad range of topics that should be further studied for future research. It is believed that a wave equation that can consider as many properties as possible would result in information-rich synthetic data that is closer to observed data compared to more conventional wave equations. Thus, the generalised viscoelastic wave equation in porous media was formulated. However, it has yet to be tested. Discretisation is essential for modelling travelling waves and the finite difference method is recommended, as it is the simplest to understand and implement. To prevent spurious information in synthetic data that arises from artificial reflections for computational simulation of wave propagation, three conditions implemented at the periphery of the model were identified. Absorbing boundary layers surround the model helps to diminish outward bound waves from its edges, and boundary condition applied at the outer edges of computational modelling domain impedes inward-bound weakened waves. Free surface boundary condition characterises wave displacement behaviour on the surface of parameter models. Implementing all three conditions is recommended to avoid

obtaining wrong data due to artificial travelling waves, though it may increase computational costs especially when the applied model is large. The MUMPS free software package was recommended to solve the linear algebra  $Ax^* = b$  for modelling wave propagation in frequency domain. It should be noted that there are other available solvers that can carry out the same task such as Multfront and UMFPACK. It is prudent to conduct a comparison between these solvers in carrying out forward modelling to determine which is the least computationally efficient and accurate.

Chapter 4 takes into consideration the possibility of seismic data being obtained from regions with topography and identifies that the most computationally efficient method is the body-fitted grid scheme. This scheme can be executed iteratively with the finite difference method to generate mesh grids that fits within non-uniform and quadrilateral shaped modelling domain. This chapter is inspired by previous work (Rao and Wang, 2013,2018) and is believed to have potential for further expansion of the topic. However computational co-ordinates must be considered when discretising wave equations within the mesh and implementing the proposed wave equation of this thesis within the body-fitted computational domain is considered for future studies.

Chapter 5 describes four of the more advanced forms of FWI that takes a different approach to the misfit led FWI, where each of them are specialised to tackle unique problems within FWI. There were no suggestions as for the advanced FWI to replace conventional FWI, in fact it should be suggested that an inversion scheme that implements both would prove to be effective for an efficient recovery of the the true model.

All things considered it should be agreed that the range of topics in seismic inversion is vast, and that a thesis would do no justice in explaining all, and that new ideas such as the ones proposed by the author is a small but welcoming drop in the barrel of the field of geophysics.

# References

- Aghamiry, H.S., Gholami, A., Operto, S., (2020) Full Waveform Inversion with Adaptive Regularization. arXiv:2001.09846.
- Aghazade, K., Amini, N., Asnaashari, A., (2015) Tikhonov Regularization of 1D Waveform Inversion in Frequency-Domain. *The 17th Iranian Geophysical Conference–May, 2015*.
- Aki, K., & Richards, P. (2002) Quantitative seismology (second edition): University Science Books.
- Alkhalifa, T., Choi, Y., (2014) From tomography to full-waveform inversion with a single objective function, *Geophysics*, 79(2), pp R55-R61.
- Al Muhaidib, A. M., Sen, M. K., Toksoz, M. N., (2012) Integration of geology, rock physics, logs, and prestack seismic data for reservoir porosity estimation. *AAPG bulletin*, 96(7), 1235-1251.
- Alterman, Z. & Karal, F.C. (1968) Propagation of elastic waves in layered media by finite-difference methods, *Bulletin of the Seismological Society of America*, 58(1), 367-398.
- American Association of Petroleum Geology – AAPG (2019) *Seismic interpretation*. Available from: [https://wiki.aapg.org/Seismic\\_interpretation#:~:text=Simply%20defined%2C%20seismic%20interpretation%20is,from%20the%20processed%20seismic%20record.&text=The%20interpretation%20process%20can%20be,structural%2C%20stratigraphic%2C%20and%20lithologic](https://wiki.aapg.org/Seismic_interpretation#:~:text=Simply%20defined%2C%20seismic%20interpretation%20is,from%20the%20processed%20seismic%20record.&text=The%20interpretation%20process%20can%20be,structural%2C%20stratigraphic%2C%20and%20lithologic). [Accessed 19/03/2021].
- Amestoy, P.R., Guermouche, A., L'Excellent, J.-Y., and Pralet, S. (2006) Hybrid scheduling for the parallel solution of linear systems, *Parallel Computing*, 32, 136-156.
- Andrey, B., Vladimir, G., Ilya, T., (2000) Estimation of fracture parameters from reflection seismic data: I. HTI model due to a single fracture set. *Geophysics*, 65, 1788–1802.
- Appelö, D., Petersson, N.A., (2009) A Stable Finite Difference Method for the Elastic Wave Equation on Complex Geometries with Free Surface, *Communications in Computational Physics*, 5 (1), 84-107.
- Backus, G.E., Gilbert, F., (1968) The resolving power of gross Earth data. *Geophysical Journal of the Royal Astronomical society*, 16, 169-205.
- Bale, R.A., (1999) Elastic wave-equation migration of HTI seismic data. *SEG Technical Program Expanded Abstracts*, 23(1), 997.
- Bale, R.A., Margrave, G.F., (2005) Elastic wave-equation migration for isotropic and HTI media. *2005 CSEG National Convention*.
- Bath, M., (1974) Spectral analysis in geophysics: Developments in Solid Earth Geophysics, *Elsevier Science Publishing Company*, 7.
- Bednar, J.B., Shin, C., & Pyun, S., (2007) Comparison of waveform inversion, part 2: phase approach, *Geophysical Prospecting*, 55, 465-475.
- Benamou, J.D., Brenier, Y., (2000) A computational fluid mechanics solution to the Monge Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393.
- Benamou, J.D., Froese, B.D., Oberman, A.M., (2014) Numerical solution of the Optimal Transportation problem using the Monge-Ampère equation, *Journal of Computational Physics*, 260(1), 107-126.
- Berenger, J.P., (1994) A perfectly matched layer for the absorption of electromagnetic waves, *Journal of Computational Physics*, 114(2), 185-200.
- Bernardino, A., Santos-Victor, J., (2005) A Real-Time Gabor Primal Sketch for Visual Attention, *Iberian Conference on Pattern Recognition and Image Analysis*, 335-342.
- Bertsekas, D.P., (1992) Auction algorithms for network flow problems: a tutorial introduction, *Computational Optimization and Applications*, 1(1), 7–66.
- Biot, M. A., - (1956a), Theory of propagation of elastic waves in a fluidsaturated porous solid. 1. Low-frequency range, *Journal of the Acoustical Society of America*, 28, 168–178.

- (1956b), Theory of propagation of elastic waves in a fluidsaturated porous solid. 1. Higher-frequency range, *Journal of the Acoustical Society of America*, 28, 179–191.
- (1962a), Mechanics deformation and acoustic propagation in porous media, *Journal of Applied Physics*, 33, 1482–1498.
- (1962b), Generalized theory of acoustic propagation in porous dissipative media, *Journal of Applied Physics*, 34, 1254–1264.
- Bishop, T.N., Bube, K.P., Cutler, R.T., Langan, R.T., Love, P., Resnick, J.R., Shuey, R.T., Spindler, D.A., and Wyld, H.W., (1985) Tomographic determination of velocity and depth in laterally varying media, *Geophysics*, 50(6), 903-923.
- Bleibinhaus, F., and Rondenay, S., (2009) Effects of surface scattering in fill-waveform inversion, *Geophysics*, 60 (6), WCC69-WCC77.
- Blias, E., (2011) Q-factor estimation through optimization approach to near-offset VSP data. *SEG 2011 annual meeting*.
- Bloot, R., Schleicher, J., Santos, L.T., (2013) On the elastic wave equation in weakly anisotropic VTI media. *Geophysical Journal International*, 192(3), 1144-1155.
- Bording, R.P., Gersztenkorn, A., Lines, L.R., Scales, J.A., Treitel, S., (1987) Application of seismic travel-time tomography. *Geophysical Journal of the Royal Astronomical Society*, 90(2), 285-303.
- Bosc, D., (2010) Numerical approximation of optimal transport maps, *SSRN Electronic Journal*.
- Bozdağ, E., Trampert, J., Tromp, J., (2011) Misfit functions for full waveform inversion based on instantaneous phase and envelope measurement. *Geophysical Journal International*, **185**, 845-870.
- Brenier, Y., (1991) Polar factorization and monotone rearrangement of vector-valued functions, *Communications on Pure and Applied Mathematics*, 44, 375–417.
- Brossier, R., Operto, S., Virieux, J., (2009) Robust elastic frequency-domain full-waveform inversion using the L1 norm, *Geophysical Research Letters*, 36, L20310.
- Bunks, C., Saleck, F. M., Zaleski, S., and Chavent, G., (1995) Multiscale seismic waveform inversion, *Geophysics*, 60, 1457–1473.
- Carcione, J.M., Kosloff, D., Kosloff, R., (1988) Wave propagation simulation in a linear visco-acoustic medium. *Geophysical Journal International*, 93, 393–401.
- Carcione, J.M., (1996) Wave propagation in anisotropic, saturated porous media: plane-wave theory and numerical simulation. *The Journal of the Acoustical Society of America*, 99(5), 2655–2666.
- Carcione, J.M., Cavallini, F., Mainardi, F., Hanyga, A., (2002) Time-domain modeling of constant-Q seismic waves using fractional derivatives, *Pure Applied Geophysical*, 159, 1719-1736.
- Carcione, J.M., (2010) A generalization of the Fourier pseudospectral method, *Geophysics*, 75, A53–56.
- Chapman, C.H., Hobro, J.W., Robertsson, J.O., (2014) Correcting an acoustic wavefield for elastic effects. *Geophysical Journal International*, 197, 1196–1214.
- Charlier, J.P., Crowet, F., (1985) Wave equations in linear viscoelastic materials. *The Journal of the Acoustical Society of America*, **79**, 895.
- Chavent, G., Goodson, R.E. (ed.), and Polis, M. (ed.), (1974) Identification of function parameters in partial differential equations. *Joint Automatic Control Conference, Austin, Texas, USA by the American Society of Mechanical Engineer*.
- Cheng, P., and Margrave, G. F., (2012a) A match-filter method for Q estimation. *SEG expanded abstract, SEG 2012 annual meeting*.
- Cheng, P., Margrave, G.F., (2012b) Estimation of Q: a comparison of different computational methods. *CREWES Research report*, 24.
- Claerbout, J., (1971) Toward a unified theory of reflector mapping, *Geophysics*, 36, 467–481.
- Claerbout, J., (1976) Fundamentals of geophysical data processing, *McGraw-Hill Book Co., Inc*.

- Clayton, R. & Engquist, B. (1977) Absorbing boundary conditions for acoustic and elastic wave equations, *Bulletin of Seismological Society of America*, 67(6), 1529-1540.
- Clement, F., Chavent, G., and Gomez, S., (2001) Migration-based travel-time waveform inversion of 2-D simple structures: a synthetic example, *Geophysics*, 66(3), 845–860.
- Cobo, Y., Carlos, C.M., Chi, S.H., (2018) Improving model resolution with FWI for imaging and interpretation in a Gulf of Mexico dataset. *2018 SEG International Exposition and Annual Meeting*, doi: <https://doi.org/10.1190/segam2018-2997675.1>
- Collino, F. & Tsogka, C. (2001) Application of the Perfectly Matched Absorbing Layer Model to the Linear Elastodynamic Problem in Anisotropic Heterogeneous Media, *Geophysics*, 66, 294-307.
- Day, S. M., Minster, J.B., (1984) Numerical simulation of attenuated wavefields using a Pade approximant method. *Geophysical Journal International*, 78, 105–118.
- Deal, M., (2014) Full Wavefield Inversion: A Paradigm Change in the Imaging and Interpretation of Seismic Data. *Paper presented at the 21st World Petroleum Congress*.
- Deng, F., McMechan, G.A., (2007) True-amplitude prestack depth migration. *Geophysics*, 72, S155–S166.
- Dvorkin, J., Nur, A., (1993) Dynamic poroelasticity: a unified model with the squirt and the Biot mechanisms, *Geophysics*, 58(4), 524–533.
- Ellefsen, K.J., (2009) A comparison of phase inversion and traveltimes tomography for processing near-surface refraction traveltimes, *Geophysics*, 74(6), WCB11–WCB24.
- Emmerich, H., Korn, M., (1987). Incorporation of attenuation into time-domain computations of seismic wavefields: *Geophysics*, 52, 1252-1264.
- Engquist, B. & Majda, A. (1977) Absorbing boundary condition for the numerical simulation of the waves, *Mathematics of Computation*, 31 (139), 629-651.
- Engquist, B., Froese, B. D., (2014), Application of the Wasserstein metric to seismic signals, *Communications in Mathematical Sciences*, 12, 979–988.
- Engquist, B., Froese, B.D., & Yang, Y., (2016), Optimal transport for seismic full waveform inversion, *Communications in Mathematical Sciences*, 14, 2309–2330.
- Engquist, B., Yang, Y.N., (2020) Optimal Transport Based Seismic Inversion: Beyond Cycle Skipping, arXiv preprint arXiv:2002.00031.
- Esser, E., Guasch, L., Van Leeuwen, T., Aravkin, A.Y., Herrmann, F.J., (2018) Total Variation Regularization Strategies in Full-Waveform Inversion, *SIAM Journal on Imaging Sciences*, 11(1), 376-406.
- Estrela, V.V., Magalhães, H.A., Saotome, O., (2016) Total Variation Applications in Computer Vision In: Kamila, N.K. (ed.) *Handbook of Research on Emerging Perspectives in Intelligent Pattern Recognition, Analysis, and Image Processing*, India, IGI Global, 41-64.
- Fellah, Z.E.A., Mohamed Fellah, M., Depollier, C., Ogam, E., Mitri, F.G., (2018) Wave Propagation in Porous Materials. In: Reyhanoglu, M. (ed.) *Computational and Experimental Studies of Acoustic Waves*. North Carolina, America, IntechOpen, pp.99-116.
- Froese, B. D., (2012) A numerical method for the elliptic Monge-Ampère equation with transport boundary conditions. *SIAM Journal on Scientific Computing*, 34, A1432–A1459, doi: 10.1137/110822372.
- Fu, L., Guo, B.W., Schuster, G.T., (2018) Multiscale phase inversion of seismic data, *Geophysics*, 83(2), R159-R171.
- Futterman, W. I., (1962) Dispersive body waves. *Journal of Geophysical Research*, 67, 5279-5291.
- Gao, K., Fu, S.B., Chung, E.T., (2018) A high-order multiscale finite-element method for time-domain acoustic-wave modelling. *Journal of Computational Physics*, Volume 360, p. 120-136.
- Gao, K., Huang, L.J., (2019) Acoustic- and elastic-waveform inversion with total generalized p-variation regularization. *Geophysical Journal International*, 218(2), 933-957.
- Gao, Y.J., Song, H.J., Zhang, J.H., and Yao, Z.X., (2017) Comparison of artificial absorbing boundaries for acoustic wave equation modelling, *Exploration Geophysics*, 48(1), 76–93.



- Gassmann, F., (1951) Elastic waves through a packing of spheres, *Geophysics*, 16(4), 673-685.
- Golubev, V.I., Petrov, I.B., Khokhlov, N.i., Shul'ts, K.I., (2015) Numerical Computation of Wave Propagation in Fractured Media by Applying the Grid Characteristic Method on Hexahedral Meshes. *Computational Mathematics and Mathematical Physics*, 55(3), 512–522.
- Green, C.H., (1979), John Clarence Karcher, 1894–1978, Father of the Reflection Seismograph, *Geophysics*, 178.
- Guasch, L., Warner, M., Ravaut, C., (2019) Adaptive waveform inversion: Practice, *Geophysics*, 84(3), R447-R461.
- Guiton, A., (2012) Blocky regularization schemes for full waveform inversion, *Geophysical Prospecting*, 60, 870–884.
- Guo, P., McMechan, G.A., Ren, L., (2019) Modeling the viscoelastic effects in P-waves with modified viscoacoustic wave propagation, *Geophysics*, 84(6), 381-394.
- Haase, A. B., Stewart, R. R., (2004), Attenuation (Q) from VSP and log data: Ross Lake, Saskatchewan, *2004 CSEG National Convention*.
- Haber, E., Ascher, U.M., Oldenburg, D., (2000), On optimization techniques for solving nonlinear inverse problems, *Inverse Problems*, 16(5), 1263–1280.
- Hadamard, J., (1902). Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*. pp. 49–52.
- Hager, W. W., Zhang, H., (2006) A survey of nonlinear conjugate gradient methods. *Pacific Journal of Optimization*, 2, 35–58.
- He, B., Liu, Y.K., Lu, H.Y., Shi, H.X., Yi, J., (2019) Reflection waveform inversion with variable density, *Journal of Applied Geophysics*, <https://doi.org/10.1016/j.jappgeo.2019.103827>
- Hestenes, M.R., Stiefel, E., (1952) Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49, 409-436.
- Hodgetts, D. and Howell, J.A., (2000) Synthetic seismic modelling of a large-scale geological cross-section from the Book Cliffs, *Petroleum Geoscience*, 6, 221–229.
- Hofmann, B., Mathé, P., (2017) Tikhonov regularization with oversmoothing penalty for non-linear ill-posed problems in Hilbert scales. *Inverse Problems*, 34(1), DOI: 10.1088/1361-6420/aa9b59.
- Hu, L.Z., McMechan, G.A., (1987) Wavefield transformations of vertical seismic profiles. *Geophysics*, 52(3), 307–21.
- Huang, J.P., Li, C., Wang, R.R., (2015) Plane-wave least-squares reverse time migration for rugged topography. *Journal of Earth Science*, 26(4), 471–480.
- Huang, Y.J., Zhu, G.M., Liu, C.Y., (2011) An Approximate Acoustic Wave Equation for VTI Media. *Chinese Journal of Geophysics*, 54(4), 599-607.
- Jannsen, D., Voss, J., Theilen, F., (1985) Comparison of methods to determine Q in shallow marine sediments from vertical reflection seismograms. *Geophysical Prospecting*, 33, 479-497, 1985.
- Jaynes, E.T., (1968) Prior probabilities. *IEEE Transactions on System Science and Cybernetics*, 4, 227-241.
- Johnson, D., Koplik, J., and Dashen, R., (1987) Theory of dynamic permeability and tortuosity in fluid-saturated porous media. *Journal of Fluid Mechanics*, 176, 379–402.
- Kelly, K.R., Ward, R.W., Treitel, S., Alford, R.M., (1976) Synthetic Seismograms: A finite-difference approach, *Geophysics*, 41(1), 2-27.
- Kim, M., Cho, H., Min, D.J., Shin, C., (2011) Comparison of frequency-selection strategies for 2D frequency domain acoustic waveform inversion, *Pure and Applied Geophysics*, 168, 1715–1727.
- Kjartansson, E., (1979) Constant-Q wave propagation and attenuation. *Journal of Geophysical Research*, 84, 4737-4748.
- Knott, M., Smith, C.S., (1984) On the optimal mapping of distributions. *Journal of Optimization Theory and Applications*, 43(1), 39–49.
- Kohno, H., Bathe, K.J., Wright, J.C., (2010) A finite element procedure for multiscale wave equations with application to plasma waves. *Computers and Structures*, 88, 87–94.

- Komatitsch, D., Coute, F., Mora, P., (1996) Tensorial formulation of the wave equation for modelling curved interfaces, *Geophys. J. Int.*, 127, 156–168.
- Komatitsch, D., Martin, R., (2007) An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation. *Geophysics*, 72, SM155–SM167.
- Kreiss, H.O., Scherer, G., (1974) Finite element and finite difference methods for hyperbolic partial differential equations, *Mathematical aspects of finite elements in partial differential equations. Symposium Proceedings.* 33, 195–212.
- Kuzuoglu, M., and Mittra, R., (1996) Frequency dependence of the constitutive parameters of causal perfectly matched anisotropic absorbers, *IEEE Microwave and Guided Wave Letters*, 6(12), 447–449.
- Lailly, P., (1983) The seismic inverse problem as a sequence of before stack migrations: Conference on Inverse Scattering, Theory and Application, *Society for Industrial and Applied Mathematics*, Expanded Abstracts, 206–220.
- Lee, S.J., Komatitsch, D., Huang, B.S., and Tromp, J., (2009) Effects of Topography on Seismic-Wave Propagation: An Example from Northern Taiwan. *Bulletin of the Seismological Society of America*, 99(1), 314–325.
- Li, F.P., Gao, J.H., Gao, Z.Q., Jiang, X.D., Sun, W.B., (2019) A causal imaging condition for reverse time migration using the Discrete Hilbert transform and its efficient implementation on GPU. *Journal of Geophysics and Engineering*, 16(5), 894–912.
- Li, Y., Gu, H.M., (2019) Full waveform inversion for velocity and density with rock physical relationship constraints. *Journal of Applied Geophysics*, 167, 106–117.
- Lin, Y., Huang, L., (2015). Acoustic-and elastic-waveform inversion using a modified total-variation regularization scheme. *Geophysical Journal International*, 200(1), 489–502.
- Lines, L., Vashegani, F., (2012) Estimating heavy oil viscosity from crosswell seismic data, *Journal of Seismic Exploration*, 21(3), 247–266.
- Lions, J., (1972) Nonhomogeneous boundary value problems and applications. *Springer Verlag, Berlin*.
- Liu, X., Gao, F., Zhang, Y., Rao, Y., Wang, Y.H., (2018) Seismic resolution enhancement in shale-oil reservoirs, *Society of Exploration Geophysicists*, 83(5), 1SO-Z29.
- Liu, X., Greenhalgh, S., (2018) Numerical modeling of poro-viscoelastic wave propagation in effective Biot media using a mixed grid finite-difference frequency-space domain approach. *SEG Technical Program Expanded Abstracts 2018*, 3813–3817.
- Liu, X., Ning, J., Zhang, G., (2009) Cauchy sparse constrained Bayesian estimation based seismic blind deconvolution frame and algorithm. *Geophysical Prospecting Petroleum*, 48, 459–64.
- Liu, C.M., Wang, D.L., Wang, T., Feng, F., Wang, Y.G., (2017) Multichannel sparse deconvolution of seismic data with shearlet–Cauchy constrained inversion. *Journal of Geophysics and Engineering*, 14(5), Pages 1275–1282.
- Liu, L.S., Wu, Y.H., Sun, F.L., (2018) Blow-up of solutions for a nonlinear viscoelastic wave equation with initial data at arbitrary energy level. *Applicable Analysis: An International Journal*. 98(12), 2308–2327.
- Liu, J.W., Yong, W.A., (2016) Stability analysis of the Biot/squirt models for wave propagation in saturated porous media, *Geophysical Journal International*, 204, 535–543.
- Liu, F., Zhang, G., Morton, S.A., Leveille, J.P., (2011) An effective imaging condition for reverse-time migration using wavefield decomposition. *Geophysics*, 76(1), S29–39.
- Liu, Z.Y., Zhang, J., (2017) Joint Traveltime and Waveform Envelope Inversion for Near-surface Imaging. *Pure and Applied Geophysics* 174 (3), 1269–1289.
- Luo, S., Sava, A., (2011) deconvolution-based objective function for wave-equation inversion. *SEG Technical Program Expanded Abstracts*, 30(1), 2788–2792.
- Luo, Y., Ma, Y., Wu, Y., Liu, H.W., Cao, L., (2016) Full-traveltime inversion, *Geophysics*, 81(5), R261–R274.
- Mahieux, C.A., (2006) Environmental Impact on Micromechanical and Macromechanical Calculations. *Environmental Degradation of Industrial Composites*, 175–232.
- Martin, G.S., (2004) *The Marmousi2 model, elastic synthetic data, and an analysis of imaging and AVO in a structurally complex environment*. Masters’ thesis. University of Houston.

- Martinez, J.R., Chemingui, N., Crawley, S., Zou, Z.H., Valenciano, A., and Klochikhina, E., (2016) A robust FWI gradient for high-resolution velocity model building. *SEG Technical Program Expanded Abstracts*: 1258-1262.
- Mattsson, K., Nordström, J., (2004) Summation by parts operators for finite difference approximations of second derivatives. *Journal of Computational Physics*, **199**, 503–540, <https://doi.org/10.1016/j.jcp.2004.03.001>.
- Mattsson, K. (2012) Summation by parts operators for finite difference approximations of second derivatives with variable coefficients. *Journal of Scientific Computing*, **51**, 650–682, <https://doi.org/10.1007/s10915-011-9568-1>.
- Martin, R., Bodet, L., Tournat, V., Rejiba, F., (2019) Seismic wave propagation in nonlinear viscoelastic media using the auxiliary differential equation method. *Geophysical Journal International*, *Oxford University Press*, 216 (1), pp.453-469.
- Mehta, C.H., Goel, B.S., Bhatta, D.D., Radhakrishnan, S., (1991) Minimum phase wavelet by ARMA factorization, *IEEE Transactions on Signal Processing*, 39(2), 512-515.
- Mohebbi, F., Sellier, M., (2014) Aerodynamic Optimal Shape Design Based on Body-Fitted Grid Generation. *Mathematical Problems in Engineering*, doi: 10.1155/2014/505372.
- Monge, G., (1781) Mémoire sur la théorie des déblais et de remblais. *Histoire de l'académie royale des sciences de paris: Avec les Mémoires de Mathématique et de Physique pour la mme année*, 666–704.
- Nandi, P., (2018) An approximate tomographic operator for RTM-based WEMVA in the common-offset domain. *SEG Technical Program Expanded Abstracts*, 5238-5242.
- Nandi, P., Albertin, U.K., (2020) Wave-equation migration velocity analysis in the common-offset domain to avoid cycle-skipping for FWI. *Journal of Applied Geophysics*, 180, Available from: <https://doi.org/10.1016/j.jappgeo.2020.104115>.
- Neep, J. P., Sams, M. S., Worthington, M. H., and O'Hara-Dhand, K. A., (1996) Measurement of seismic attenuation from high-resolution crosshole data. *Geophysics*, 61, 1175-1188.
- Nemeth, T., Wu, C., Schuster G.T., (1999) Least-squares migration of incomplete reflection data. *Geophysics*, 64(1), 208–221.
- Niño, C.L.B., and Vides, L.A.M., (2005) Wave equation datuming to correct topography effect on foothill seismic data. *Earth Sciences Research Journal*, 9(2), 132-138.
- Nocedal, J., (1980) Updating Quasi-Newton Matrices with Limited Storage, *Mathematics of Computation*, 35(151), 773–782.
- Nuber, A., Manukyan, E., Maurer, H., (2016) Ground topography effects on near-surface elastic full waveform inversion. *Geophysical Journal International*, 207, 67-71.
- Palacz, M., (2018) Spectral Methods for Modelling of Wave Propagation in Structures in Terms of Damage Detection—A Review. *Applied Science*, 8(7), 1124-1149.
- Patantonis, D.E., Atharassiadis, N.A., (1985) A numerical procedure for the generation of orthogonal body-fitted coordinate systems with direct determination of grid points on the boundary, *International Journal for Numerical Methods in Fluids*, 5, 245–255.
- Phillips, S.J., Dudík, M., (2008) Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation. *Ecography*, 31, 161–175.
- Phillips, J.S., Elith, J., (2013) On estimating probability of presence from use-availability or presence-background data. *Ecology*, 94(6), 1409-1419.
- Picottia, S., Carcione, J.M., Rubino, J.G., Santos, J.E., Cavallinia, F., (2010) A viscoelastic representation of wave attenuation in porous media. *Computers and Geosciences*, 36(1), 44-53.
- Plessix, R.-E., (2006) A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophysical Journal International*, 167, 495–503.
- Pratt, R.G., (1990) Frequency-domain elastic wave modelling by finite differences: a tool for crosshole seismic imaging, *Geophysics*, 55, 626– 632.
- Pratt, R. G., and Worthington M. H., (1990) Inverse theory applied to multi-source cross-hole tomography, Part 1: Acoustic wave-equation method, *Geophysical Prospecting*, 38, 287–310.

- Pratt, G., Shin, C., and Hicks, G., (1998) Gauss-newton and full newton methods in frequency-space seismic waveform inversion, *Geophysical Journal International*, 133(2), 341–362.
- Pratt, R. G., (1999) Seismic waveform inversion in the frequency domain, part 1: theory and verification in a physical scale model, *Geophysics*, 64, 888–901.
- Pride, S.R., Berryman, J.G. and Harris, J.M., (2004) Seismic attenuation due to wave-induced flow, *Journal of Geophysical Research*, 109, B01201, doi:10.1029/2003JB002639.
- Pyun, S.J., Shin, C.S., Bednar, J. B., (2007) Comparison of waveform inversion, part 3: amplitude approach, *Geophysical Prospecting*, 55(4), 477–485.
- Raikes, S.A., White, R.E., (1984) Measurements of earth attenuation from downhole and surface seismic recordings. *Geophysical Prospecting*, 32, 892–919.
- Ramirez, A.C., Lewis, W., (2010) Regularization and full-waveform inversion: A two-step approach. *SEG Technical Program Expanded Abstracts*.
- Rane, S., Kovacevic, A., (2017) Application of numerical grid generation for improved CFD analysis of multiphase screw machines. *IOP Conference Series: Materials Science and Engineering*, 232, doi:10.1088/1757-899X/232/1/012017
- Rao, Y., and Wang, Y.H., (2013) Seismic waveform simulation with pseudo-orthogonal grids for irregular topographic models, *Geophysical Journal International*, 194, 1778–1788.
- Rao, Y., and Wang, Y.H., (2018) Seismic waveform simulation for models with fluctuating interfaces, *Scientific Reports*, 8(1):3098. Available from 10.1038/s41598-018-20992-z
- Ricker, N., (1944) Wavelet functions and their polynomials, *Geophysics*, 9(3), 287–409.
- Roden, J.A., Gedney, S.D., (2000) Convolution PML (CPML): An efficient FDTD implementation of the CFS–PML for arbitrary media, *Microwave and Optical Technology Letters*, 27(5), 334–339.
- Rudin, L.I., Osher, S., Fatemi, E., (1992) Nonlinear total variation based noise removal algorithms. *Physica D*, 60, 259–268.
- Santos, J.E., Ravazzoli, C.L., Gauzellino, P.M., Carcione, J.M., Cavallinia, F., (2004) Simulation of waves in poro-viscoelastic rocks saturated by immiscible fluids: numerical evidence of a second slow wave. *Journal of computational acoustics*, 12(1), 1–21.
- Sava, P., and Biondi, B., (1999) Wave-equation migration velocity analysis: 69th Annual International Meeting. Society of Exploration Geophysicists, Expanded Abstracts, 1723–1726.
- , (2004a) Wave-equation migration velocity analysis, I: Theory, *Geophysical Prospecting*, 52(6), 593–606.
- , (2004b) Wave-equation migration velocity analysis — II: Subsalt imaging examples, *Geophysical Prospecting*, 52, 607–623.
- Schuster, G.T., (1993) Least-squares cross-well migration. *SEG Technical Program Expanded Abstracts*. 110–113.
- Scott-Blair, G.W., (1947) The role of psychophysics in rheology. *Journal of Colloid and Interface Science*, 2, 21–32.
- Seidl, R., Rank, E., (2017) Full waveform inversion for ultrasonic flaw identification, *AIP Conference Proceedings*, 1806, 090013.
- Sharma, M.D., and Gogna, M.L., (1991) Seismic wave propagation in a viscoelastic porous solid saturated by viscous liquid, *Pure and applied geophysics*, 135, 383–400.
- Shen, P., Symes, W., and Stolk, C. C., (2003) Differential semblance velocity analysis by wave-equation migration, *73rd Annual International Meeting SEG Expanded Abstracts*, 2132–2135.
- Shin, C., and Min. D., (2006) Waveform inversion using logarithmic wavefield, *Geophysics*, 71(3) R31–R42.
- Shin, C., Pyun. S., and Bednar. J.B., (2007) Comparison of waveform inversion, part 1: conventional wavefield vs logarithmic wavefield, *Geophysical Prospecting*, 55, 449–464.
- Shin, C., and Ha, W., (2008) A comparison between the behaviour of objective functions for waveform inversion in the frequency and Laplace domains, *Geophysics*, 73, no. 5, VE119–VE133.
- Shuey, R.T., (1985) A simplification of the Zoeppritz equations. *Geophysics*, 50 (9), 609–614.

- Sirgue, L., and Pratt, R. G., (2004) Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies, *Geophysics*, 69, 231–248.
- Smit, W. and de Vries, H., (1970) Rheological models containing fractional derivatives. *Rheologica Acta*, 9, 525–534.
- Society of exploration geophysicists – SEG (2018) *Seismic stratigraphy*. Available from: [https://wiki.seg.org/wiki/Seismic\\_stratigraphy](https://wiki.seg.org/wiki/Seismic_stratigraphy). [Accessed 19/03/2021].
- Society of Exploration Geophysicists (2020) *AGL Elastic Marmousi*. Available from: [https://wiki.seg.org/wiki/AGL\\_Elastic\\_Marmousi](https://wiki.seg.org/wiki/AGL_Elastic_Marmousi) [Accessed 28/03/2021].
- Song, H.T., Zhong, C.K., (2010) Blow-up of solutions of a nonlinear viscoelastic wave equation. *Nonlinear Analysis: Real World Applications*, 11(5), 3877–3883.
- Soufyane, A., Feng, B.F., (2020) Optimal decay rates of a nonlinear time-delayed viscoelastic wave equation. *Differential and Integral Equations*, 33(1), 43–65.
- Stanford Exploration Project. (2002) Migration Velocity Analysis: Introduction. Available from: [http://sepwww.stanford.edu/public/docs/sep112/paul3/paper\\_html/node1.html](http://sepwww.stanford.edu/public/docs/sep112/paul3/paper_html/node1.html) [02/03/2021]
- Strand, B. (1994) Summation by parts for finite difference approximations for  $d/dx$ . *Journal of Computational Physics*, 110, 47–67, <https://doi.org/10.1006/jcph.1994.1005>.
- Stück, H., Koch, R., Siegesmund, S., (2013) Petrographical and petrophysical properties of sandstones: statistical analysis as an approach to predict material behaviour and construction suitability. *Environmental Earth Science*, 69, 1299–1332.
- Sun, X., Tang, X., Cheng, C.H., Frazer, L.N., (2000) P- and S- wave attenuation logs from monopole sonic data. *Geophysics*, 65, 755–765.
- Suzuki, A., Tsubota, Y., Wu, W.J., Yamanaka, K., Terada, T., Kawabata, K., (2019) Full waveform inversion for ultrasound computed tomography with high-sensitivity scan method, *Proc. SPIE 10955, Medical Imaging 2019: Ultrasonic Imaging and Tomography*, 109550A.
- Symes, W.W., (2020) Wavefield Reconstruction Inversion: an example, *arXiv:2003.14181*.
- Tan, S.R., Huang, L.J., (2014) An efficient finite-difference method with high-order accuracy in both time and space domains for modelling scalar-wave propagation. *Geophysical Journal International*, 197(2), 1250–1267.
- Tang, Y., Lee, S., Baumstein, A., Hinkley, D., (2013) Tomographically enhanced full waveform inversion: SEG Technical Program Expanded Abstracts. *Society of Exploration Geophysicists*, 1037–1041,
- Tanimoto, T., Hadziioannou, C., Igel, H., Wasserman, J., Schreiber, U., Gebauer, A., (2015) Estimate of Rayleigh-to-Love wave ratio in the secondary microseism by colocated ring laser and seismograph. *American Geophysical Union*, 42(8), 2650–2655.
- Tarantola, A., (1984) Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, 49, 1259–1266.
- Tarantola, A., (1986) A strategy for nonlinear elastic inversion of seismic reflection data, *Geophysics*, 51, 1893–1903.
- Tarantola, A., (1987) Inverse Problem Theory: Methods for Data Fitting and Parameter Estimation, *Elsevier, Amsterdam*.
- Tikhonov, A.N., (1935) Théorèmes d'unicité pour l'équation de la chaleur. *Matematicheskij Sbornik*, 42, 199–216.
- Tikhonov, A. N., Arsenin, V.Y., (1977) Solution of Ill-posed Problems, *Washington: Winston & Sons*. ISBN 0-470-99124-0.
- Tikhonov, A.N., Goncharsky, A.V., Stepanov, V.V., Yagoda, A.G., (1995) Numerical Methods for the solution of Ill-posed problems, *Dordrecht*.
- Tonn, R., (1991) The determination of seismic quality factor Q from VSP data: A comparison of different computational methods. *Geophysical Prospective*, 39, 1–27.
- Trinks, I., Singh, S.C., Chapman, C.H., Barton, P.J., Bosch, M. & Cherrett, A., (2005) Adaptive traveltime tomography of densely sampled seismic data, *Geophysics Journal International*, 160(3), 925–938.
- University of California San Diego (2018) Earthquake Seismology. Available from: <https://igppweb.ucsd.edu/~gabi/sio15/lectures/Lecture07.html> [Accessed: 25/03/2021].

- Valenciano, A.A., Chemingui, N., Whitmore, D., and Brandsberg-Dahl, S., (2011) Wave equation migration with attenuation and anisotropy compensation, *81st Annual International Meeting of Society of Exploration Geophysicists*, 232–236. (Expanded Abstracts)
- Van den Berg, P.M., Kleinman, R.E., (1997), A contrast source inversion method, *Inverse Problems*, 13(6), 1607–1620.
- Van Leeuwen, T., and Mulder, W. A., (2010) A correlation-based misfit criterion for wave-equation traveltime tomography, *Geophysical Journal International*, 182, 3, 1383–1394.
- Van Leeuwen, T., Herrmann, F.J., (2013) Mitigating local minima in full-waveform inversion by expanding the search space, *Geophysical Journal International*, 195(1), 661–667.
- Venkateshan, S.P., Swaminathan, P., (2014) Chapter 12 – Introduction to PDEs. *Computational Methods in Engineering*. Madras, India, Academic Press, pp.523-527.
- Vigh, D., Starr, E. W., (2008) Comparisons for waveform inversion, time domain or frequency domain, *78<sup>th</sup> Annual International Meeting SEG, Expanded Abstracts*, 1890–1894.
- Villani, C., (2003) *Topics in Optimal Transportation*. Lyon, France, 58, American Mathematical Society.
- Virieux, J., (1986) P-SV wave propagation in heterogeneous media: velocity stress finite-difference method, *Geophysics*, 51, 889–901.
- Virieux, J., Operto, S., (2009) An Overview of Full Waveform Inversion in Exploration Geophysics, *Geophysics*, 74(6), PP. WC1-WC26.
- Wang, Y.H., (2008) Seismic Inverse Q-Filtering. *Wiley-Blackwell*.
- Wang, Y.H., (2016) Generalized viscoelastic wave equation. *Geophysical Journal International*, 204, 1216-1221.
- Wang, Y.H., (2016) Seismic Inversion: Theory and Applications, London, *Wiley-Blackwell*.
- Wang, H.Y., Singh, S.C., Calandra, H., (2014) Integrated inversion using combined wave-equation tomography and full waveform inversion, *Geophysical Journal International*, 198, 430-466.
- Warner, M., Ratcliffe, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N., Vinje, V., Štekl, I., Guasch, L., Win, C., Conroy, G., and Bertrand, A., (2013) Anisotropic 3D full-waveform inversion. *Geophysics*, 78(2), R59-R80.
- Warner, M. R., Guasch, L., (2014) Adaptive waveform inversion: Theory, *76th Annual International Conference and Exhibition, EAGE, Extended Abstracts*, doi: 10.3997/2214-4609.20141092.
- Warner, M., Guasch, L., (2016) Adaptive waveform inversion: Theory, *Geophysics*, 81(6), R429-R445.
- Wood, A.B., (1941) A textbook of sound, *G. Bell and Sons Ltd*, London.
- Wu, R.S., Luo, J.R., Wu, B.Y., (2014) Seismic envelope inversion and modulation signal model, *Geophysics* 79(3) PP. WA13-WA24.
- Wu, R.S., Luo, J.R., (2015) Seismic envelope inversion: reduction of local minima and noise resistance *Geophysical Prospecting*, 63, 597-614.
- Xu, W.C., Li, Z.C., Jiao, W., Yan, Z., (2015) A pure viscoacoustic equation for vti media applied in anisotropic rtm. *Journal of Geophysics and Engineering*, 12(6), 969-977.
- Yang, C.Y., Li, X.Y., Wang, Y., (2015) An analysis of 3D anisotropic–viscoelastic forward modeling and dissipation. *Journal of Geophysics and Engineering*, 12(6), 1036–1048.
- Yang, Z.Q., Liu, Y., Ren, Z.M., (2014) Comparisons of visco-acoustic wave equations, *Journal of Geophysics and Engineering*, 11(2), 1-7.
- Yang, Y.N., Engquist, B., Sun, J.Z., Hamfeldt, B.F., (2018) Application of optimal transport and quadratic Wasserstein metric to full-waveform inversion, *Geophysics*, 83(1), R43-R62.
- Yang, Q.J., Malcolm, A., (2019) Single parameter full waveform inversion in fluid-saturated porous media. SEG Technical Program Expanded Abstracts: 1450-1454.

- Yang, Q.J., Mao, W.J., (2017) Simulation of seismic wave propagation in 2-D poroelastic media using weighted-averaging finite difference stencils in the frequency–space domain. *Geophysical Journal International*, 208(1), 148–161.
- Yao, G., Wu, D., Wang, S.X., (2020) A review on reflection-waveform inversion. *Petroleum Science*, 17,334–351.
- Zennir, K., Alodhaibi, S.S., (2020) A Novel Decay Rate for a Coupled System of Nonlinear Viscoelastic Wave Equations. *Mathematics*, 8(2), 203.
- Zhang, J. & Liu, T., (1999) P-SV-wave propagation in heterogeneous media: grid method. *Geophysical Journal International*, 136, 431–438.
- Zhang, S.Z., Schuster, G., and Luo, Yi., (2012) Angle-domain Migration Velocity Analysis using Wave-equation Reflection Traveltime Inversion. *SEG Technical Program Expanded Abstracts*, 1-6.
- Zhang, Y., Zhang, P., Zhang, H., (2010) Compensating for visco-acoustic effects in reverse-time migration. *80th Annual International Meeting of Society of Exploration Geophysicists*, 3160–3164. (Expanded Abstracts)
- Zhang, Y., Zhang, H.Z., Zhang, G.Q., (2011) A stable TTI reverse time migration and its implementation, *Geophysics*, 76(3), WA3-WA11.
- Zhang, W.S., Zheng, H., (2019) A multiscale method for wave propagation in 3D heterogeneous poroelastic media. *Geophysics*, 84(4), T237-T257.
- Zhou, W., Brossier, R., Operto, S., Virieux, J., (2015) Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation. *Geophysical Journal International*, 202(3), 1535–1554.
- Zhou, X.C., Xu, Y.Z., Lin, J.G., (2017) Wavelet estimation in varying coefficient models for censored dependent data, *Statistics and Probability Letters*, 122, 179-189.
- Zoeppritz, K., (1919) On reflection and transmission of seismic waves by surfaces of discontinuity, *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse*, 66–84.