Development of 2D Full Waveform Inversion algorithm of seismic data

A Thesis

submitted to
Indian Institute of Science Education and Research (IISER) Pune in partial fulfillment of the requirements for the BS-MS Dual Degree Programme
by

Sujith M S

Indian Institute of Science Education and Research (IISER) Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411 008, INDIA.

December, 2022

Supervisor: Dr. Rahul Dehiya
TAC Member: Dr. Arun Singh

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान पुणे
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

IISER PUNE

Certificate

This is to certify that this dissertation entitled "Development of 2D Full Waveform Inversion algorithm of seismic data" towards the partial fulfillment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research (IISER), Pune represents the study/work carried out by Sujith M S at IISER Pune under the supervision of Dr. Rahul Dehiya, Assistant Professor, Department of Earth and Climate Science , during the academic year 2021-2022.

Dr. Rahul Dehiya

Committee:

Dr. Rahul Dehiya, IISER Pune

Dr. Arun Singh, IIT (ISM) Dhanbad

Declaration

I hereby declare that the research work presented in the report entitled "Development of 2D Full Waveform Inversion algorithm of seismic data" have been carried out by me at the Department of Earth and Climate Science, Indian Institute of Science Education and Research, Pune, under the supervision of Dr. Rahul Dehiya and the same has not been submitted elsewhere for any other degree.

Sujith M S

“Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.”
-John W. Tukey, 1962

Acknowledgments

I want to thank IISER Pune and Earth and Climate Science department for giving me the opportunity to undertake this study. The whole IISER community is always curious and eager to learn, which has always motivated me. I want to thank my supervisor Dr. Rahul Dehiya for his constant guidance and support throughout the whole period of the project. I started working with Dr. Dehiya in my third year. Through various projects, I have learned invaluable lessons and gained expertise in exploration and computational seismology principles. Every discussion with him was always insightful, informative, and enlightening, and it has helped me change my views on tackling a geophysical problem. His dedication to teaching and deep knowledge of geophysics is always a constant source of inspiration to me. I also want to thank the TAC member, Dr. Arun Singh, for his valuable suggestions and support regarding the inversion scheme using the NLCG method.

I would also like to thank Dr. Neena Joseph Mani, fifth-year project coordinator of the Earth and Climate Science department, and Dr. Girish Ratnaparkhi, Dean, Academics, for understanding my situation on the delay of the IPGP program and approving the extension of this project.

I would also like to acknowledge the computational resources and support provided by the PARAM Brahma facility under the National Supercomputing Mission, Government of India, at IISER Pune.

Abstract

Full Waveform Inversion (FWI) is an advanced and emerging seismic imaging technique that exploits the full wavefield information present in the data. FWI is an optimization problem where the misfit between the recorded and generated data is iteratively minimized. This project attempts to create a robust forward acoustic wave propagation engine with appropriate boundary conditions and a non-linear waveform inversion algorithm using the Gauss-Newton (GN) and Non-Linear Conjugate Gradient (NLCG) methods for the 2D seismic dataset. Both the acoustic wave propagation modeling and the waveform inversion are formulated in the frequency domain. The forward model is created using the fourth-order finite difference approximation of the scalar wave equation on staggered-grid. The boundaries are incorporated with Perfectly Matching Layers (PMLs) to suppress undesired reflections from edges. The forward modeling was posed as a sparse linear system and is solved for each frequency by LU decomposition. The seismogram in the time domain was retrieved by applying a Fourier transform to the generated data. The waveform inversion is implemented in a least-square sense, i.e., the objective function defined as L_2 norm of data residuals is minimized. A term containing the prior information about the model parameters is also incorporated in the objective function for stability. The inversion is done initially using the Gauss-Newton method, where the Hessian (the second-order derivative of the objective function) is approximated in terms of the first-order derivative of the forward operator. And later implemented using the Non-Linear Conjugate Gradient method, where the model is updated in the opposite of gradient direction. The forward modeling results establish the reliability and robustness of the created forward engine. We start with a simple two-layer model and compare the arrival time of the reflection event of generated seismogram with the calculated arrival time. A comparison of data generated using the algorithm and data generated using Madagascar is also provided. The waveform inversion is done at discreet frequencies; hence it is computationally less expensive as compared to the time domain approach. The inversion results on synthetic models with buried boxes at different depths with anomalous velocities are presented using both GN and NLCG methods. Both methods are also tested on a checkerboard velocity model for a rigorous investigation of the robustness of the algorithms.

Keywords: FWI, inverse theory, optimization, finite-difference methods, frequency domain modeling, acoustic wave propagation.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Context

Seismic data provide information on the physical parameters and detailed structure of the subsurface. Since the 1900s, modern seismology has witnessed a dramatic evolution in terms of seismic data acquisition and data processing techniques (Oldham, 1906). Another possible method of studying the subsurface geology is by drilling boreholes; however, due to the high cost and limited ability to provide information only at discrete locations make it useful for obtaining information at selected locations only. The seismic method offers a relatively rapid and cost-effective means of retrieving information on subsurface structures. For reliable imaging, the seismic data go through rigorous processing, however, in case of complex geology, the interpretation of the data may include a fair amount of uncertainties. The information derived from other geophysical surveying methods, including borehole data, can be integrated with seismic data to reduce these uncertainties. Seismic surveying methods can be divided into two major methods. First which makes use of natural source fields, and second, those that employ artificially generated energy to study the subsurface. Natural sources like earthquakes, ambient noise, could provide information on subsurface structures of exceptionally deeper depths, sometimes up to the Earth's core, and are logistically more simple to carry out than artificial active-source methods. However, the active-source methods are capable of producing a high-resolution image with fine details about the structures of the subsurface geology (Kearey et al., 2002) and have been a common survey method for hydrocarbon exploration since the late 20^{th} century.

Conventionally, the marine seismic data are acquired using a vessel that tows streamers and airgun arrays. The airguns produce an artificial acoustic wave, and the streamers comprise hydrophones that are capable of measuring the change in acoustic pressure inside the water column (Figure 1.1). On land, common sources are dynamite or Vibroseis, and geophones are used for field measurements and monitoring the land surface vibrations, usually with accelerometer sensors (Yilmaz, 2001). Ocean Bottom Nodes (OBNs) include these types of sensors in addition to a hydrophone (the combined set of sensors is sometimes referred to as the 4C sensor: 3 components geophone and a hydrophone). OBNs use battery-powered cableless receivers - although sometimes linked by a string to speed up the retrieval. Nodes have been designed to tackle the exploration in deep water areas beyond the 1km depth limit that most Ocean Bottom Cables (OBCs) have. By nature, nodes are also usable in complicated areas with obstacles (close to the platform, small islands, and rocks)

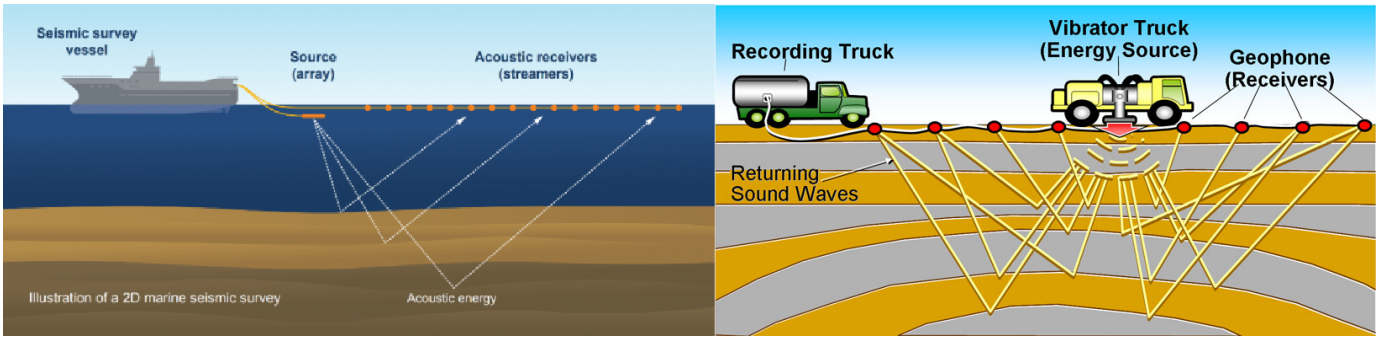

Figure 1.1: Illustration of 2D marine (left) and land (right) seismic survey methods; images taken from "https://www.nopsema.gov.au/offshore-industry/environmental-management/marine-seismic-surveys" and "http://geologylearn.blogspot.com/2015/06/marine-and-land-seismic-aquisition.html".

for which other types of sensors cannot be deployed without risks. The deployment of these nodes can be more flexible than the cables in OBC due to their smaller size and lower weight, and OBNs are significantly important for 4D time-lapse studies (Cantillo et al., 2010). Remotely Operated underwater Vehicles (ROV) are used to place these nodes at their predetermined locations accurately. OBNs are able to collect large offset and wide-azimuth data. The set of 4C sensors embedded in OBNs ensures proper measurements of ground movement in both vertical and horizontal directions, as well as pressure variations at the sea bottom.

Standard seismic data processing involves removing unnecessary noises and multiple reflections by keeping the primary reflections only. For removing these unwanted events, the data (from the time-offset domain) is transformed into other domains where they are spatially separated and can easily be removed. F-K transform, radon transform (Cao, 2006), and frequency filter are some of the common techniques used for noise removal. In the early days, the reflection data was stacked after NMO correction, where the events are made flat in each Common Depth Point (CDP) gathers, and the stack was migrated (post-stack migration (Bancroft, 2007)). Migration is a technique that maps the dipping reflectors to their actual subsurface positions and collapses diffraction events, thereby increasing the resolution (Sheriff and Geldart, 1987). Due to advancements in processing sequences and computing power, the denoised data is migrated before stacking (pre-stack migration (Dondurur, 2018)). Reverse Time Migration (RTM), Kirchhoff pre-stack depth (or time) migration (KPSDM), and Wave-Equation based Kirchhoff migration are some examples of pre-stack migration algorithms. In order to obtain a better image with effective focusing and a clear definition of events, the pre-stack migration algorithms require an accurate velocity model (Robein, 2010). In

earlier days, ray theory-based traveltime tomography was used for velocity model building. Traveltime tomography-based techniques using only the arrival time information of major events gave only approximate velocity models and sometimes failed for complex geological settings of the survey area.

1.2 Rise of Waveform Inversion

Over the past fifty years, the understanding of seismic imaging methodologies has rapidly evolved. Two major types of interaction between the medium and the seismic waves have been established for the model-building strategies. The first is a smooth interaction which occurs where the direction of propagation of the wave is affected with a minor deviation. This minor variation in the wavefields allows the waves to carry on to advance forward. This interaction, which causes slight variation in wave direction in a forward scattering mode, constitutes the transmission regime. Another type of interaction is a rough interaction where the direction of the propagation of the wave is completely modified by the properties of the medium. This type of interaction compels the waves to propagate in the same plane as the incident wave. This kind of interaction, in a backward scattering manner, constitutes the reflection regime. Other types of interaction, such as multiple scattering effects, which constitutes the diffusive regime, could also be used for the model building. However, these regimes require entirely different strategies and techniques (Wu and Aki, 1988). The transmission and reflection regimes have paved the way for two major strategies for subsurface imaging. The primary type of imaging is based on tomographic approaches in which smoothly varying information of the subsurface is reconstructed by taking into account slow variations of phases (or the traveltimes). Many approaches are considered depending on the different phases. First-arrival phases, refraction phases, reflection phases, and surface-wave phases are different possible phases used for these analyses and reconstruction. It was also observed that accounting for the combination of these different phases significantly enhances the quality of the reconstruction (Thurber and Ritsema, 2007). The secondary type of imaging is based on migration approaches. This can be implemented either in time or in space. This strategy is associated with Huygens' principle, which states that every point on the primary wavefront act as a source point for the secondary wavefront. Abrupt variations in the properties of the medium are retrieved through this strategy. The so-called least-square migration is one of the strategies that make use of the reflection regime and have improved the quality of the imaging techniques by accurately estimating the properties of the medium (Nemeth et al., 1999).

Towards the end of 20^{th} century, the waveform inversion scheme was posed as a local optimization

problem (Lailly, 1983; Tarantola, 1984), where the aim was to minimize the least-square misfits of recorded and predicted data. The minimization is done iteratively due to the highly non-linear nature of the problem. The above authors showed that the gradient of the misfit function through a perturbation model could be calculated by zero-lag correlation of incident wavefields from the source and the back-propagated residual wavefields from the receiver. It was observed that the reconstructed model after the first iteration of FWI shows a close resemblance to the image of the subsurface, migrated using the RTM algorithm. In FWI, the full-wavefield information content, including primary reflections, transmitted waves, supercritical reflections, and multiple scattered waves such as surface-related multiples, short-path multiples (Virieux and Operto, 2009) are exploited. Contrary to traveltime tomography, FWI provides a more accurate velocity model since it accounts for all the wave attributes, such as arrival time, amplitude, frequency, and phase.

The modeling of seismic data and waveform inversion can be performed either in the time domain or in the frequency domain (Plessix, 2007; Vigh and Starr, 2008). Hybrid approaches are also tried where the wave propagation is implemented in the time domain and the inversion in the frequency domain (Sirgue et al., 2008). Today most of the FWI algorithms are completely equipped in the time domain due to the good scalability, ease in parallelization, and moderate memory demand for large 3D datasets. For 2D datasets, frequency domain modeling and inversion are more effective and computationally less expensive (Pratt, 1999). A few implementations of 3D FWI in the frequency domain on synthetic and real data can be discovered in Plessix (2009); Virieux and Operto (2009); Ben-Hadj-Ali et al. (2008); Petrov and Newman (2014). Waveform inversion was quite computationally expensive, even for 2D surveys. Nevertheless, it has been successfully implemented in various studies by wave propagation using strategies such as finite difference techniques (Ikelle et al., 1988), finite element and spectral element methods (Bae et al., 2010), and extended ray theory (Sambridge and Drijkoningen, 1992). The insensitivity of seismic wavefields to intermediate wavelengths arises from short-offset acquisitions, and the iterative optimization process fails to retrieve the subsurface velocity structure accurately. The waveform inversion reconstructs the actual subsurface structure through such updates only when an adequately accurate starting model is provided and contains long-offset and transmission data to retrieve intermediate to long wavelength structures (Mora, 1987; Pratt and Worthington, 1990).

At this moment, FWI is still a deterministic approach, and nevertheless it is a powerful tool in modern seismic imaging methodologies. However, It faces some challenges and limitations such as the non-uniqueness of the model parameters estimation, the highly non-linear nature of the problem, convergence to local minima, high computational cost, lack of low-frequency signals and presence

of low-frequency noises in the recorded seismic data, and also the approximations we consider while creating the forward engine. The advancements in modern high-performance computing and multi-component wide-aperture and wide-azimuth acquisitions techniques, such as Ocean Bottom Nodes (multi- to rich-azimuth data), very long streamers (ultra-far offset data), mitigate these challenges to some extent and favors to obtain a robust and reliable subsurface image. Furthermore, FWI is becoming a mature strategy for high-resolution seismic imaging by the inversion of multi-parameters with elastic wave equation.

Over the past three decades, frequency-domain finite-difference (FDFD) methods have been ardently applied to model wave propagation. However, the challenges in solving the huge linear systems have limited its popularization and further advancements. Sparse LU decomposition strategy (George and Liu, 1981) was greatly favored for FDFD methods by allowing easy computation of multiple simultaneous sources for 2D problems. Nonetheless, the application for 3D problems is rare due to the huge memory requirement. One of the main benefits of FDFD methods in waveform inversion is the ease of sequentially using the low- to high-frequency components of the data, termed the spectral inversion strategy (Pratt and Worthington, 1990). Low-frequency data contains long wavelengths and it will be harder to skip a half cycle while matching the observed and modeled data. Then using these long wavelength updated models as the starting models for higher frequency components increases the possibility of convergence to the global minimum of the objective function. In the time domain, the waveform matching process is highly prone to cycle-skipping since low- and high-frequency components are inverted simultaneously. Arbitrary frequency-dependent attenuation mechanisms can be incorporated into FDFD wave propagation modeling at no extra computational cost by making the elastic parameters complex. In the same way, by the FDFD method, the inclusion of attenuation in the inversion is also made practical. The wave propagation modeling of multiple simultaneous sources by FDFD approaches is greatly simplified by using LU decomposition strategies. The memory requirement for this approach limits the use for 2D and 2.5D problems only, and for 3D problems, the frequency domain methods renounce this advantage, and iterative Krylov-type techniques are required. Despite these benefits, FDFD methods also suffer from a number of challenges that have limited their broad popularization. One of the main limitations of FDFD over TDFD (time-domain finite-difference) is its high sensitivity towards imperfect and poor boundary conditions since FDTD could easily remove those remnant energies by time windowing.

One of the vital elements for a robust FWI result is an efficient and powerful forward engine that can account for all the phenomena while the wave interacts with the subsurface/ medium. Frequency

domain forward modeling is particularly favored over time domain for multiple simultaneous source experiments because of its computational efficiency (Pratt and Worthington, 1990) due to the finite frequency effect. The critical step in FDFD modeling responsible for its computational efficiency is solving the massive sparse linear system. The FDFD approach gained popularity for 2D acoustic and elastic problems with centered finite difference discretization on a single grid using second-order spatial derivative approximations (Pratt and Worthington, 1990). This led to the five-point stencils for the acoustic and nine-point stencils for the elastic wave equations. This scheme demands at least ten gridpoints be included within the shortest wavelength for accurate modeling and numerical stability. This requirement of the minimum number of gridpoints was reduced by having a higher number of stencil points. The fourth-order FD approximations on staggered-grid led to thirteen-point stencils and required only four gridpoints per shortest wavelength.

Through this study, we try to develop a robust forward wave propagation engine in the frequency domain with proper boundary conditions. Elastic wave equation could give more accurate results but demands much higher computational cost due to the influence of multiple parameters such as P- and S-wave velocity, mass density, anisotropy, attenuation, etc. The existence of these multi-parameters increases the ill-posedness of the local optimization process due to the increased degrees of freedom and the variable sensitivity of these different parameters (Virieux and Operto, 2009). Therefore, we are constraining the acoustic approximation of the wave equation and considering a single parameter, P-wave velocity, for the inversion. The inversion algorithm is implemented in the frequency domain using the Gauss-Newton (GN) method (Fletcher, 1986) and the Non-Linear Conjugate Gradient (NLCG) method (Fletcher and Reeves, 1964). The Gauss-Newton method is a variant of the Newton method for determining the minimum of a non-linear system, and it corresponds to a more accurate optimization strategy by utilizing the Hessian (the second derivative of the objective function) information (Wang and Dong, 2021). The GN method does not actually calculate the second derivative explicitly but approximates it in terms of the Fréchet derivative of the forward operator, which is computationally less expensive to assess as compared to full Newton methods. NLCG is a further development of the linear conjugate gradient method to solve non-linear systems. In NLCG, the model update is performed in the opposite direction of the gradient, and the amount of update is estimated using the line search methods. NLCG exhibits a faster convergence than the steepest descent methods but slower than GN. Even though GN has a faster convergence rate than NLCG, NLCG requires lesser computational resource per iteration than GN.

2 ‖ Material and Methods

Before the rise of Full Waveform Inversion, the velocity information of the subsurface was retrieved by the techniques such as traveltime tomography, wave equation tomography, and waveform tomography, which are ray-based methods or methods using approximations of the wave equation. In traveltime tomography, the first arrival time of major events from source to receiver was picked, and the difference between the observed and modeled traveltime was minimized. Wave equation tomography is based on the cross-correlation of recorded data with the synthetic data (Wang et al., 2014). The cross-correlation measures the similarity (shift in traveltime) between recorded and synthetic waveforms, and the misfit is minimized by maximizing the cross-correlation. And in waveform tomography, only phase, by normalizing the amplitude, is accounted for in the inversion and is helpful in detecting shadow zones (low-velocity gradient) or regions that cause triplications (high-velocity gradient). Unlike these methods, in FWI, the difference between the recorded and simulated data (sample by sample, i.e., both traveltime and amplitude) is minimized iteratively to get an image of the subsurface that theoretically could represent the subsurface accurately if non-uniqueness is appropriately handled. FWI does not require the manual picking of events, and it is a data-driven strategy. Consequently, no subjective choice of user or bias of event picking algorithm is involved in the subsurface velocity model building. FWI takes into account the complete information available in the seismic data, exploiting traveltime and waveform attributes such as amplitude, phase, and frequency content. All kinds of events present in the data, such as primary reflections, transmitted waves, and surface multiples, could be accounted for in the inversion process. Ideally, FWI output should be the exact representation of subsurface structures and parameters. But due to the limitations such as approximate forward simulation introduced to optimize the high computational cost and issue of cycle skipping, currently, FWI is mainly used as a velocity model building strategy for the pre-stack depth migration (Borisov and Singh, 2015). The final subsurface image is retrieved after stacking the migrated seismic data.

Low-frequency signals are of high significance for FWI to overcome the solution being trapped in the local minima. The quality of usable low-frequency components present in the data increases the chance of matching the modeled data with the observed data within a half wavelength. The initial model also determines the robustness of the FWI output. If the initial model is far away from the proximity of the global minimum, FWI will output the wrong subsurface image by converging to one of the many local minima of the misfit function. This happens because of the wrapping-around behavior of phase in the frequency domain or the cycle-skipping issue(figure 2.1) in the

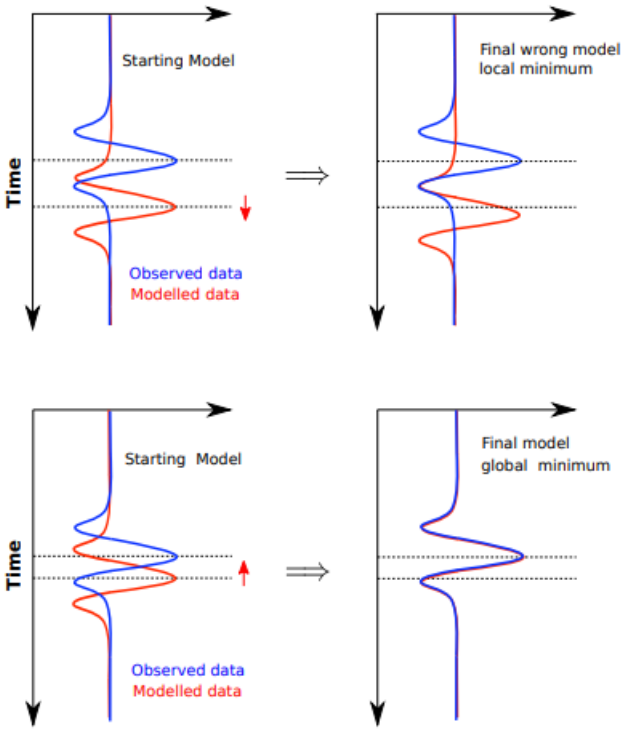

Figure 2.1: Representation of the issue of cycle-skipping in FWI after Anagaw (2014).

time domain (Bunks et al., 1995; Sirgue and Pratt, 2004). Cycle-skipping in FWI originates from the oscillatory nature of seismic data. It occurs when the arrival time of the modeled waveform differs from that of the observed waveform by more than half a period. If the initial model is not sufficiently accurate, it is possible to face the cycle-skipping ambiguity. In the attempt to fit the waveforms, the half-cycle shifted waveforms wrongly match and lead to one of the local minima during the optimization process. This is significant in the case of least-square misfit functions (L_2 norm of the data residual). Therefore, it is required to rely on conventional methods such as traveltime tomography for constructing the starting models.

The near-surface effects, such as complex and irregular surface topography, and surface wave propagation through the highly heterogeneous near-surface on land, contribute to the high non-linear behavior of the optimization problem. It is also required to have specific acquisition geometry for better and accurate results, in particular, long offset data because of the presence of deepest penetrating transmitted waves. The solution of ill-posed problems suffers from non-uniqueness, instability, or non-existence; the inverse theory provides the means to solve ill-posed problems. Conversely, if a problem has a stable unique solution, it is termed a well-posed problem. In the ill-posed waveform

inversion problem, a unique model cannot be estimated because of the existence of multiple models which are capable of producing data that could possibly match with the observed data or provide minimal data residual value. Prior information regarding the model parameters, preconditioning of data, and regularization methods are essential for a rewarding FWI scheme (Tarantola, 1987; Virieux and Operto, 2009). An accurate representation of the source wavelet is also significant for modeling the waveforms.

Lailly (1983) and Tarantola (1984) introduced the concept of waveform inversion in the time domain, and subsequently, it was developed in the frequency domain by Pratt (1999). Each domain has its own benefits and constraints. In wide-aperture seismic data, the redundancy of the wavenumber allows the inversion of model parameters using discreet frequency values in the frequency domain. This makes it computationally less expensive than the time domain approach and is preferred for small-scale problems (Sirgue and Pratt, 2004). It also has the ability to incorporate multiple simultaneous sources efficiently. The inversion is executed in a sequential manner starting from the low- to high-frequencies. This sequential inversion can be implemented with a single frequency value or multiple frequencies. The output models obtained from the inversion using low-frequency components are used as the starting model for subsequent inversions with high-frequency components. Using the low-frequency, in the beginning, will deliver long wavelength updates to the model. The consequent inversions with higher frequencies bring short wavelength updates and finer details of the subsurface structures. This multiscale spectral inversion approach overcomes the issue of the cycle-skipping (in time domain) or wrapping around of phase (in frequency domain) due to the reduced probability of skipping a half-cycle for low-frequency components (Bunks et al., 1995). If all the frequencies are used in the inversion, then the frequency domain approach is identical to the time domain approach. Despite these advantages of the frequency domain approach, the ease and practicability in implementation and parallelization of the algorithm for massive 3D datasets, the time domain approach is most favored and broadly popularized for waveform inversion (Warner et al., 2013). Time domain-based wave modeling and inversion allows easy removal of spurious and remnant energies, as well as muting of specific part of the data by time-windowing, and it also provides higher flexibility for data preconditioning. The multiscale spectral inversion strategy equivalent to the frequency domain can also be implemented at an increased cost in the time domain by band-pass (or low-bass) filtering the time domain data based on different frequency bandwidths.

FWI is a highly non-linear, non-unique, and ill-posed inverse problem. In order to accurately retrieve the subsurface structure, any available information should be used while the optimization process. Geophysical methods such as well logs, stratigraphic data, gravimetric methods, elec-

tromagnetic prospecting, etc could provide information about the subsurface. Incorporating these pieces of information in FWI could result in a robust reconstruction (Bell et al., 1978). It is possible to integrate such prior information as a preconditioner for data gradient (Fomel and Claerbout, 2003). Or it is also possible to jointly update for different model parameters using other types of data along with seismic data (Asnaashari et al., 2013). Both these approaches improve convergence and linearity and provide a reliable output through the completely data-driven inversion strategy.

As mentioned earlier, implementing the inversion in the frequency domain provides an unaliased image using a very limited number of discreet frequencies (Pratt and Worthington, 1988). Previously Wu and Toksöz (1987) presented that in the wavenumber domain, even a single frequency provides a significant amount of information about the subsurface. These observations demonstrate the low computational cost for inversion in the frequency domain since the computational cost is directly proportional to the number of discreet frequencies used. However, Freudenreich and Singh (2000) investigated the efficiency of frequency domain over time domain inversion and concluded that frequency domain inversion fails in the presence of a limited offset range and claimed that there exists no strategy to choose the discreet frequencies appropriately. Sirgue and Pratt (2004) revisited the frequency domain waveform inversion more rigorously and came up with a strategy for selecting discreet frequencies for an efficient waveform inversion and imaging. The frequencies can be selected using the following expression,

$$f_{n+1} = f_n \left\{ \sqrt{1 + R_{max}^2} \right\} \tag{2.1}$$

where f_n is the previously selected frequency value, and R_{max} is the ratio of maximum half-offset to the depth of interest for imaging.

The easiness in the implementation of preconditioning strategies either on the data or the gradient varies in different domains. Time windowing with residuals (Shipp and Singh, 2002) is a popular preconditioning strategy; however, data residual is necessary to be represented in the time domain. Nonetheless, it is possible to employ such a strategy in the frequency domain by including complex values for frequencies (Mallick and Frazer, 1987). Conversely, in the frequency domain, it is straightforward to implement the multiscale spectral inversion strategy at no additional computational cost. Low-frequency components of the data show high linearity with respect to model perturbations than higher frequencies, and this sequential strategy dramatically improves the chance of reaching the global minimum while the optimization process (Sirgue and Pratt, 2004). Such a methodology is possible in the time domain by low-pass filtering the data (Bunks et al., 1995),

however, with an additional cost. The fact that the propagation of individual discreet frequency is computationally cheaper than full-band time domain and therefore FDFD approach is considered to be computationally superior and efficient (Marfurt, 1984).

Another advantage of frequency domain forward modeling is its ability to incorporate multiple simultaneous sources. Due to the advancements in seismic data acquisition methods, simultaneous shooting (firing more than one source at a time or with a slight delay) is common in present-day acquisition designs. The simultaneous source shooting considerably reduces the survey duration and improves the illumination by increasing the source density (Abma et al., 2013). Most conventional data processing steps are designed to process the non-overlapping source records. Therefore, the separation of these blended source records into their equivalent non-overlapping source record is a critical step prior to classical processing sequences. Deblending is the seismic processing step for separating the overlapping simultaneous shot records into each non-overlapping single shot record. However, frequency domain modeling does not require the shots to be non-overlapping. By implementing the modeling and inversion in the frequency domain, the deblending process, which is a computationally heavy task, can be skipped, and the raw shot records can be used for inversion.

Full Waveform Inversion is an advanced data-driven tool for extracting pieces of information about the subsurface. This potentially high-resolution comes with a very high cost and strict requirements to mitigate the issue of cycle skipping, non-uniqueness, and high non-linearity (Virieux et al., 2017). FWI is evolving towards incorporating the elastic wave equation in the forward engine because of the increased focus on the amplitude information and also inversion up to higher frequencies to get very-high-resolution image of the subsurface. However, this may require multi-parameter estimation in which the Hessian matrix has a significant role and also encounter increased non-linearity. With the advancements in massive computational capabilities and modern seismic acquisition methodologies, we believe there could be much more innovations to be made in the field of FWI so that it can serve as a single-independent technique for retrieving the accurate high-resolution images of the subsurface.

2.1 The Forward Problem

Since the optimization process is minimizing the difference between the forward modeled and the recorded data, the modeling of the seismic wave propagation has a significant contribution to the robustness of the inversion result. i.e., the modeled data is trying to match with observed data sample by sample, and therefore it is crucial to generate the data as realistically as possible. The

forward modeling is done using fourth-order finite difference discretization of first-order hyperbolic formulation (Virieux, 1984) of the acoustic wave equation (scalar wave equation) on staggered-grids (Hustedt et al., 2004). The 13-point stencils used for the discretization are shown in figure 2.2. It is essential to introduce non-reflecting artificial absorbing boundary layers in the forward engine to simulate an unbounded domain. A sponge-like Perfectly Matching Layer (PML) boundary condition (Berenger, 1994) is included in the model to absorb undesired reflection energies from the edges of the physical domain. The PML boundary is not considered at the top edge in order to simulate a realistic free surface boundary condition, as it will allow reflections from the top edge (surface-related multiples).

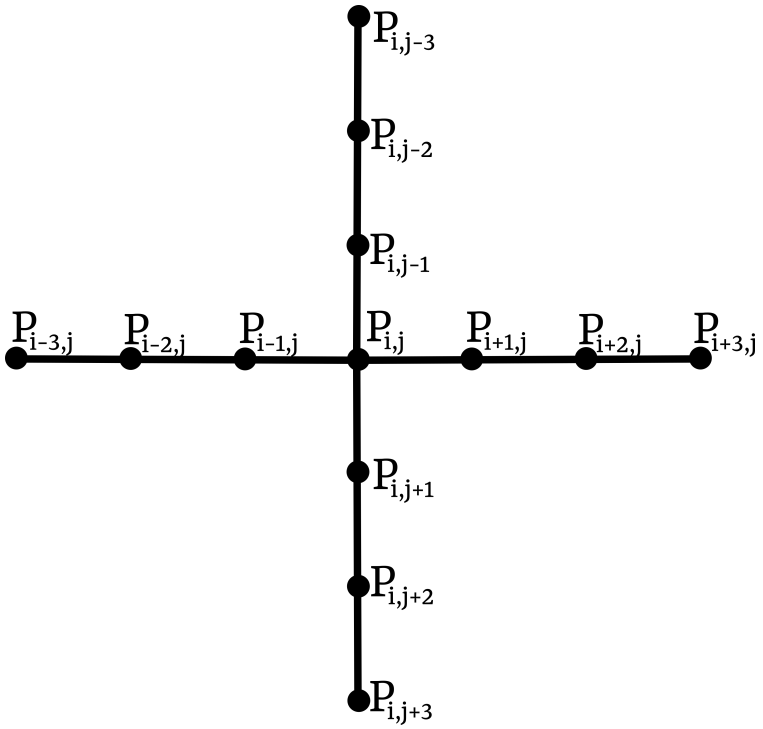

Figure 2.2: Illustration of 13-point stencils used for the finite difference discretization of the scalar wave equation.

The system of hyperbolic 2D wave equations in the time domain is represented as (Hustedt et al.,

2004):

$$
\begin{aligned}
&\frac{\partial \Psi_x(x,z,t)}{\partial t} + \gamma_x(x)\Psi_x(x,z,t) = \kappa(x,z)\frac{\partial \mathscr{P}(x,z,t)}{\partial x} + \mathscr{S}(x,z,t) \\
&\frac{\partial \Psi_z(x,z,t)}{\partial t} + \gamma_z(z)\Psi_x(x,z,t) = \kappa(x,z)\frac{\partial \mathscr{Q}(x,z,t)}{\partial z} \\
&\frac{\partial \mathscr{P}(x,z,t)}{\partial t} + \gamma_x(x)\mathscr{P}(x,z,t) = \frac{1}{\rho(x,z)}\left(\frac{\partial \Psi_x(x,z,t)}{\partial x} + \frac{\partial \Psi_z(x,z,t)}{\partial x}\right) \\
&\frac{\partial \mathscr{Q}(x,z,t)}{\partial t} + \gamma_z(z)\mathscr{Q}(x,z,t) = \frac{1}{\rho(x,z)}\left(\frac{\partial \Psi_x(x,z,t)}{\partial z} + \frac{\partial \Psi_z(x,z,t)}{\partial z}\right)
\end{aligned}
\tag{2.2}
$$

where $\mathscr{P}(x,z,t)\ and\ \mathscr{Q}(x,z,t)$ represent the components of the particle velocity vector of pressure wavefield in the Cartesian grid. The pressure wavefield $\Psi(x,z,t)$ is represented in terms of two components such that,

$$\Psi(x,z,t) \;=\; \Psi_x(x,z,t) \;+\; \Psi_z(x,z,t)$$

$\mathscr{S}(x,z,t)$ represents the term corresponding to the acoustic source. $\rho(x,z)$ is the mass density (inverse of density is called the buoyancy, $b(x,z)$), and $\kappa(x,z)$ is the bulk modulus. The mass density needs to be interpolated between gridpoints horizontally and vertically. A harmonic averaging of density is performed for the interpolation. $\gamma_x\ and\ \gamma_z$ define the damping behavior of PML boundary layers. These functions are zero inside the physical domain and non-zero only in the boundary layers. γ is defined as $\gamma_x(x) \;=\; C_{PML}cos(\frac{\pi}{2}\frac{x}{L})$ (Hustedt et al., 2004), where L represents the width of the boundary layer, and x is the local coordinate whose origin is located at the outer edges of the physical domain. The value of C_{PML} is estimated such that the reflected energy from the boundaries is minimum. The above system of equations is transformed to Fourier domain in order to represent the wave equation in the frequency domain, and a new function $\xi_x(x)$ is defined such that $\xi_x(x) \;=\; 1+\frac{i\gamma_x}{\omega}$ and $\xi_z(z) \;=\; 1+\frac{i\gamma_z}{\omega}$;

$$
\begin{aligned}
&-i\omega\frac{\xi_x(x)}{\kappa(x,z)}\Psi_x(x,z,\omega) = \frac{\partial \mathscr{P}(x,z,\omega)}{\partial x} + \mathscr{S}(x,z,\omega) \\
&-i\omega\frac{\xi_z(z)}{\kappa(x,z)}\Psi_z(x,z,\omega) = \frac{\partial \mathscr{Q}(x,z,\omega)}{\partial z} \\
&-i\omega\mathscr{P}(x,z,\omega) = \frac{b(x,z)}{\xi_x(x)}\frac{\partial \Psi(x,z,\omega)}{\partial x} \\
&-i\omega\mathscr{Q}(x,z,\omega) = \frac{b(x,z)}{\xi_z(z)}\frac{\partial \Psi(x,z,\omega)}{\partial z}
\end{aligned}
\tag{2.3}
$$

The above system is discretized using fourth-order centered FD approximation on staggered-grids

such that;

$$\begin{aligned}\left[\frac{\partial \mathscr{P}}{\partial x}\right]_{i,j} &\approx \frac{1}{\Delta_x}\left(\frac{9}{8}(\mathscr{P}_{i+\frac{1}{2},j}-\mathscr{P}_{i-\frac{1}{2},j})-\frac{1}{24}(\mathscr{P}_{i+\frac{3}{2},j}-\mathscr{P}_{i-\frac{3}{2},j})\right)\\ \left[\frac{\partial \mathscr{Q}}{\partial z}\right]_{i,j} &\approx \frac{1}{\Delta_z}\left(\frac{9}{8}(\mathscr{Q}_{i,j+\frac{1}{2}}-\mathscr{Q}_{i,j-\frac{1}{2}})-\frac{1}{24}(\mathscr{Q}_{i,j+\frac{3}{2}}-\mathscr{Q}_{i,j-\frac{3}{2}})\right)\end{aligned} \tag{2.4}$$

Using the above two systems of equations, we obtain the following equation consisting of 13-point stencils;

$$\begin{aligned}\frac{-\omega^2}{\kappa_{i,j}}\Psi_{i,j} = \frac{1}{\Delta_x^2\xi_i}&\left\{\begin{array}{l}\frac{9}{8}\left[\begin{array}{l}\frac{b_{i+\frac{1}{2},j}}{\xi_{i+\frac{1}{2}}}\left(\frac{9}{8}(\Psi_{i+1,j}-\Psi_{i,j})-\frac{1}{24}(\Psi_{i+2,j}-\Psi_{i-1,j})\right)\\ -\frac{b_{i-\frac{1}{2},j}}{\xi_{i-\frac{1}{2}}}\left(\frac{9}{8}(\Psi_{i,j}-\Psi_{i-1,j})-\frac{1}{24}(\Psi_{i+1,j}-\Psi_{i-2,j})\right)\end{array}\right]\\ -\frac{1}{24}\left[\begin{array}{l}\frac{b_{i+\frac{3}{2},j}}{\xi_{i+\frac{3}{2}}}\left(\frac{9}{8}(\Psi_{i+2,j}-\Psi_{i+1,j})-\frac{1}{24}(\Psi_{i+3,j}-\Psi_{i,j})\right)\\ -\frac{b_{i-\frac{3}{2},j}}{\xi_{i-\frac{3}{2}}}\left(\frac{9}{8}(\Psi_{i-1,j}-\Psi_{i-2,j})-\frac{1}{24}(\Psi_{i,j}-\Psi_{i-3,j})\right)\end{array}\right]\end{array}\right\}\\ +\frac{1}{\Delta_z^2\xi_j}&\left\{\begin{array}{l}\frac{9}{8}\left[\begin{array}{l}\frac{b_{i,j+\frac{1}{2}}}{\xi_{j+\frac{1}{2}}}\left(\frac{9}{8}(\Psi_{i,j+1}-\Psi_{i,j})-\frac{1}{24}(\Psi_{i,j+2}-\Psi_{i,j-1})\right)\\ -\frac{b_{i,j-\frac{1}{2}}}{\xi_{j-\frac{1}{2}}}\left(\frac{9}{8}(\Psi_{i,j}-\Psi_{i,j-1})-\frac{1}{24}(\Psi_{i,j+1}-\Psi_{i,j-2})\right)\end{array}\right]\\ -\frac{1}{24}\left[\begin{array}{l}\frac{b_{i,j+\frac{3}{2}}}{\xi_{j+\frac{3}{2}}}\left(\frac{9}{8}(\Psi_{i,j+2}-\Psi_{i,j+1})-\frac{1}{24}(\Psi_{i,j+3}-\Psi_{i,j})\right)\\ -\frac{b_{i,j-\frac{3}{2}}}{\xi_{j-\frac{3}{2}}}\left(\frac{9}{8}(\Psi_{i,j-1}-\Psi_{i,j-2})-\frac{1}{24}(\Psi_{i,j}-\Psi_{i,j-3})\right)\end{array}\right]\end{array}\right\}+\mathscr{S}_{i,j}\end{aligned} \tag{2.5}$$

This discrete frequency domain wave equation can be represented as a linear system of the form;

$$\mathscr{M}\psi = s \tag{2.6}$$

where $\mathscr{M}$ is the impedance matrix, s is the source vector and ψ is the pressure wavefield. Because of the boundary conditions, the elements of the impedance matrix are complex in nature. And the elements depend on model parameters (and frequency), and most importantly, $\mathscr{M}$ is sparse.

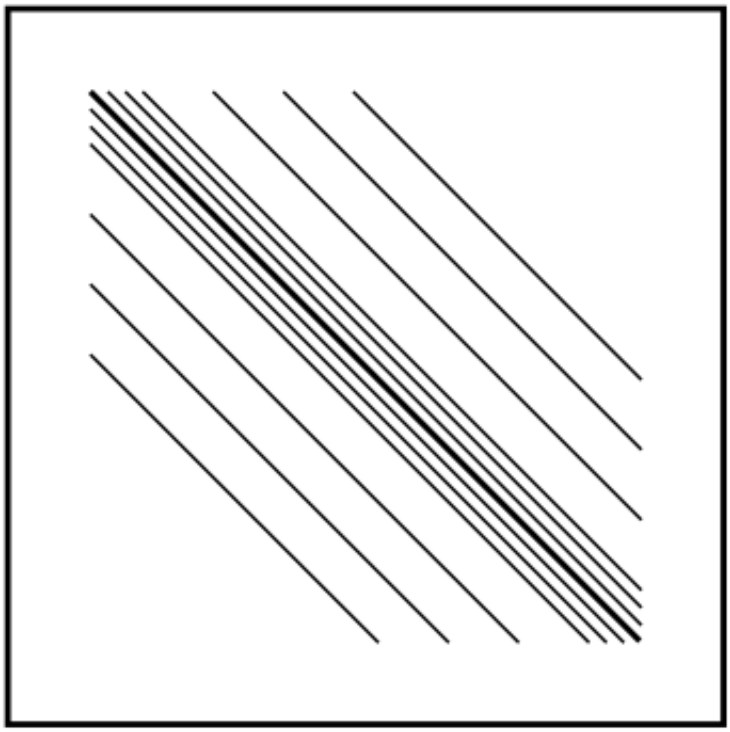

Figure 2.3: Structure of the complex-valued $\mathscr{M}$ matrix after Hustedt et al. (2004).

Multiple sources can be easily incorporated into the s vector. For each discrete frequency value, the matrix $\mathscr{M}$ is constructed and solved to obtain corresponding wavefields. The dimension of $\mathscr{M}$ matrix is $(nx \times nz) \times (nx \times nz)$, where nx and nz are the number of gridpoints in x- and z-directions respectively ($(nx \times nz)^2$ elements in total). In the case of the 13-point stencil, $(13 \times nx \times nz)$ elements are the only non-zero elements. These non-zero elements are distributed along the main diagonal and six adjacent parallel bands, and the bandwidth is $6 \times nz$ as shown in the figure 2.3. The matrix characteristics such as bandwidth and the number of non-zero elements decide the efficiency and performance of the modeling algorithm (Hustedt et al., 2004). The details of the 13-point stencils are given in Appendix A

To keep the computation practical, it is necessary to limit the extent of the domain considered for modeling. This is accomplished by placing Perfectly Matching Layers (PMLs) at the truncation boundaries. Ideally, the truncation boundaries are reflectionless, and they completely absorb the energy. In many numerical simulations, Absorbing Boundary Conditions (ABCs) are extensively adopted to generate an unbounded domain. The main purpose of these ABCs is to absorb outgoing waves from the edges and to generate an unbounded physical domain. Several ABCs have been

tested in the past for its usage in finite difference wave modeling. The ABC proposed by Cerjan et al. (1985) contains lossy material layers at the edges to suppress the reflected wavefields. These layers are observed to be absorbing the energy imperfectly, and undesired spurious energy comes from the edges; they degrade the numerical accuracy of the results. So it is important to implement the perfectly absorbing boundary condition in the forward modeling. Theoretically, an ABC should absorb all energy arriving at the edges irrespective of the frequency and incident angles. However, in practical problems, the performance of ABCs vitiates after some ranges of frequencies and incident angles. The Perfectly Matching Layer (PML) boundary condition proposed by Berenger (1994) overcome these issue to a greater extent. PML is regarded to be one of the finest techniques for ABC implementation currently available because of its superior absorption over a wide range of incident angles and its insensitivity towards frequency. In this study, we implemented the PML boundary condition to the scalar wave equation. The term ξ in Eq. 2.5 governs the boundary condition. The PMLs are considered as an extension of the outermost cell of the physical domain, meaning the wave velocity and mass density of the PMLs are equal to that of the out-most cell. This approach is studied and implemented to minimize the numerical reflection error and to perfectly absorb the energy at the interface between the PML domain and the physical domain.

2.2 The Inverse Problem

FWI is performed to reconstruct the subsurface model that could generate synthetic data which well matches the recorded seismic data. The similarity between the synthetic and observed data is measured quantitatively by a misfit/ penalty function, and the prime objective is to minimize that function. Although FWI is known for its ability to provide accurate high-resolution images, its practical applications on large-scale datasets have been limited by many challenges. The high non-linear behavior of the problem and the non-existence of the perfect optimization scheme to account for the non-linear nature is one of the main challenges. Numerous iterative inversion strategies such as Non-Linear Conjugate Gradient (NLCG) (Hestenes and Stiefel, 1952), steepest-descent, gradient-descent which are based on the gradient methods, or methods such as full Newton (FN), quasi-Newton, limited-memory BFGS (l-BFGS) (Nocedal, 1980), and Gauss-Newton (GN) methods which are based on Newton optimization scheme have been implemented for waveform inversion (Herrmann et al., 2011). For large-scale problems, gradient-based methods are more popular due to their cheaper demand of computational resources for simulating forward/adjoint response on each iteration. However, they generally have a slow convergence rate and have difficulties in properly scaling the update in the inverted model in comparison to Newton's methods. The convergence

can be improved by scaling the gradient using the diagonal elements of pseudo-Hessian (Shin et al., 2001). Conversely, Newton-based methods exhibit a faster rate of convergence. Nevertheless, the rapid convergence comes with an additional cost. In the full-Newton method, the second derivative of the misfit function, which is the Hessian matrix, is explicitly estimated and inverted. Whereas quasi-Newton methods are a cost-effective alternative to the full-Newton methods. For instance, in the l-BFGS method, without explicitly computing or storing the Hessian in the memory, it approximates the inverse of the full-Hessian matrix from a finite number of gradients and updates the model (Nocedal, 1980). The choice of model parameterization significantly affects the behavior of the Hessian matrix and thereby influence the rate of convergence and the accuracy of the inversion result.

2.2.1 Gauss-Newton Method

The inversion is driven by optimizing the misfit function defined in the least-square sense. The least-square misfit function is defined as;

$$\phi_d(m) \doteq \frac{1}{2}\left[d^{obs} - \mathscr{F}(m)\right]^T W_d^T W_d \left[d^{obs} - \mathscr{F}(m)\right] \tag{2.7}$$

where $\mathscr{F}(m)$ denotes the modeled data, d^{obs} is the observed data, superscript ^{T} represents conjugate transpose of the matrix and W_d denotes the data weighting matrix of dimension $N_d \times N_d$ (where N_d is the number of data points). Generally, W_d is a diagonal matrix with elements equal to the reciprocal of standard deviation (or data amplitudes scaled by the error floor). The weighting matrix is introduced to give equal weights to the data points and prevents the result from being biased towards the high amplitude data samples and, at the same time, penalize the erroneous data points. Although the main motive is to minimize the above function, one may have to redefine the objective function to stabilize the process (Bell et al., 1978). In order to do so, a regularisation term is added to the above function, defined as,

$$\phi_m(m, m^{ref}) \doteq \left(m - m^{ref}\right)^T W_m^T W_m \left(m - m^{ref}\right), \tag{2.8}$$

where m^{ref} contains the prior information about the model, W_m defines the model smoothness, which can be a finite difference approximation of Laplacian operator (∇^2) or an identity matrix or a sparse operator with different weights for each diagonal.

The final objective function can be defined as the sum of the above two functions as;

$$\phi(m) \doteq \phi_d(m) + \lambda\phi_m(m, m^{ref}) \tag{2.9}$$

where λ is a scalar which determines the weight of regularisation term in the overall objective function.

To optimize the above objective function, there are several algorithms with varying rates of convergence and computational costs, such as steepest-descent, gradient-descent, conjugate gradient methods, full-Newton's method, Gauss-Newton method, quasi-Newton method, l-BFGS, etc. In Newton methods, the exact Hessian matrix, the second derivative of the misfit function, is given as;

$$\nabla_m^2\phi(m) \equiv \nabla_m\mathscr{F}(m)^T\nabla_m\mathscr{F}(m) - \nabla_m^2\mathscr{F}(m)^T(d - \mathscr{F}(m)) \tag{2.10}$$

In Gauss-Newton (GN) method, the first-order derivative of the forward operator, Jacobian, is used to approximate Hessian (by neglecting the second-order term of RHS). Calculating the exact Hessian is often computationally intractable. Therefore GN method is considered computationally more practical and has a faster convergence rate compared to gradient-based schemes. In this study, we used the GN method to optimize the objective function.

By minimizing the Eq. 2.9 and discarding higher-order derivative terms, we obtain the linear system as;

$$\left[J^T W_d^T W_d J + \lambda W_m^T W_m\right]\Delta m_i = \left\{J^T W_d^T W_d\,(d - \mathscr{F}(m_i)) - \lambda W_m^T W_m\left(m - m^{ref}\right)\right\} \tag{2.11}$$

where J is the Jacobian matrix ($\nabla_m\mathscr{F}(m)$), Δm_i denotes the model perturbation at the i^{th} iteration given by,

$$\Delta m_i = m_i - m_{i-1} \tag{2.12}$$

The Eq. 2.11 can be rewritten to directly estimate the updated model, m_{i+1} (the model parameters corresponding to the $(i+1)^{th}$ iteration) by substituting Eq. 2.12 in Eq. 2.11 and we obtain;

$$m_{i+1} = \left[J^T W_d^T W_d J + \lambda W_m^T W_m\right]^{-1}\left[J^T W_d^T W_d\,(\Delta r_i + J m_i) + \lambda W_m^T W_m m^{ref}\right] \tag{2.13}$$

where $\Delta r_i = d - \mathscr{F}(m_i)$ is the the data residual corresponding to the i^{th} iteration.

Since the impedance matrix is complex-valued, its Fréchet derivative is also automatically complex-valued. While optimization, it is a common practice to decompose the real and imaginary values of the Jacobian. This strategy of decomposition of complex elements is also applied on the frequency domain data samples and is regarded as independent observations.

Jacobian Matrix Formulation

In the GN method, the construction of the Jacobian matrix takes most of the runtime. The Jacobian is defined as the Fréchet derivative of the forward operator with respect to the model parameters. The elements of the matrix are defined as;

$$J_{i,j} = \frac{\partial \mathscr{F}_i(m)}{\partial m_j} \tag{2.14}$$

where i goes from 1 to N_d and j goes from 1 to N_p.

It is not necessarily required that the receiver positions match with the gridpoints, and in order to interpolate the recordings at the gridpoints, ψ to the actual observation points, we define a transformation matrix T, such that the observations d^{pred} is defined as;

$$d^{pred} = T\psi \tag{2.15}$$

Using Eq. 2.6, Eq. 2.14, Eq. 2.15, the Jacobian is estimated using the following matrix operation;

$$J = T\mathscr{M}^{-1}G \tag{2.16}$$

where G is given as;

$$G = \left[\left(-\frac{\partial \mathscr{M}}{\partial m_1}\psi\right)\left(-\frac{\partial \mathscr{M}}{\partial m_2}\psi\right)\ldots\left(-\frac{\partial \mathscr{M}}{\partial m_{N_p}}\psi\right)\right] \tag{2.17}$$

It was noted that estimating J^T is computationally cheaper than estimating J directly because esti-

mation of J requires N_p calls (Eq. 2.16) , so J^T can be estimated as;

$$J^T = G^T(\mathscr{M}^{-1})^T T^T \tag{2.18}$$

And for solving Eq. 2.18, it requires only N_d calls and generally $N_d < N_p$.

2.2.2 Non-Linear Conjugate Gradient Method

Non-Linear Conjugate Gradient (NLCG) is a Krylov subspace-based optimization strategy for determining the minimum of a non-linear system. The GN method updates the model using the Hessian, (approximated using the Jacobian). For large 3D problems, GN demands massive computational resource. Important and reliable alternative for large problems are l-BFGS and NLCG. l-BFGS is a quassi-newton method. In NLCG, the minimum of the misfit is estimated using the gradient of the misfit for update direction and line search method for step length.

Hestenes and Stiefel (1952) proposed the linear conjugate gradient method as an iterative optimization strategy for solving linear systems. The forward operator of the linear system should be a positive definite coefficient matrix. Fletcher and Reeves (1964) proposed the NLCG method, and it was one of the very first techniques to solve large-scale non-linear systems. Many modified versions of the method have been introduced over time. The prime benefit of this method is the low memory requirement as they store no matrices and have a faster convergence rate than steepest descent (Nocedal and Wright, 2006).

The gradient of the misfit function, $(\nabla_m \phi(m))$, represents the direction of the maximum increase. So the opposite (steepest descent) direction indicates the direction towards minima and is given by;

$$h = -\nabla_m \phi(m) \tag{2.19}$$

The model parameters are updated in the steepest descent direction such that, $m_i = m_{i-1} + \alpha_i h_i$ The value of α is determined by the line search approach where the descent direction is estimated and then estimates a step length that decides the amount of model perturbation in the descent direction. The value of α_i can be estimated by the Taylor series expansion of $\phi(m_{i+1}) = \phi(m_i + \alpha_i h_i)$ as follows (Yang et al., 2015);

$$\phi(m_i + \alpha_i h_i) = \phi(m_i) + \alpha_i \nabla_m \phi(m_i)^T h_i + \frac{\alpha^2}{2} h_i^T \nabla_m^2 \phi(m_i) h_i \tag{2.20}$$

The derivative of the above expression with respect to α_i gives;

$$\nabla_m \phi(m_i)^T h + \alpha h_i^T \nabla_m^2 \phi(m_i) h_i = 0 \tag{2.21}$$

The value of α_i can be estimated from the above equation as follows;

$$\alpha_i = \frac{(J_i h_i)^T \Delta r_i}{(J_i h_i)^T (J_i h_i)} \tag{2.22}$$

After the first iteration, the descent direction is updated as;

$$h_i = -\nabla_m \phi(m_i) + \beta_i h_{i-1} \tag{2.23}$$

To estimate the value of β, Fletcher and Reeves (1964) proposed a method known as Fletcher–Reeves method. They extended the method from linear systems to non-linear systems. The value of β through the Fletcher–Reeves method is given as;

$$\beta_i = \frac{\nabla_m \phi(m_i)^T \nabla_m \phi(m_i)}{\nabla_m \phi(m_{i-1})^T \nabla_m \phi(m_{i-1})} \tag{2.24}$$

There are many variants of the Fletcher–Reeves method, such as Polak–Ribière, Hestenes-Stiefel, and Dai–Yuan methods. In this study, we use Polak–Ribière formula (Polak and Ribiere, 1969) described as;

$$\beta_i = \frac{\nabla_m \phi(m_i)^T \left\{\nabla_m \phi(m_i) - \nabla_m \phi(m_{i-1})\right\}}{\nabla_m \phi(m_{i-1})^T \nabla_m \phi(m_{i-1})} \tag{2.25}$$

It was observed that the Fletcher-Reeves method converges if the initial model is sufficiently near to the global minimum, whereas the Polak–Ribière method could possibly enter in an infinite cycle without converging in rare situations. However, it has a better convergence rate than Fletcher–Reeves approach (Shewchuk, 1994).

3 Results and Discussion

3.1 Forward modeling

The forward modeling is done by solving the Eq. 2.6 using the UMFPACK solver for each frequency. This frequency domain data is transformed into the time domain to retrieve the seismogram. For generating the data, a source with the signature of Ricker wavelet (Ricker, 1953) with a dominant frequency of 3Hz is considered. The following figure represents the amplitude spectrum of the wavelet.

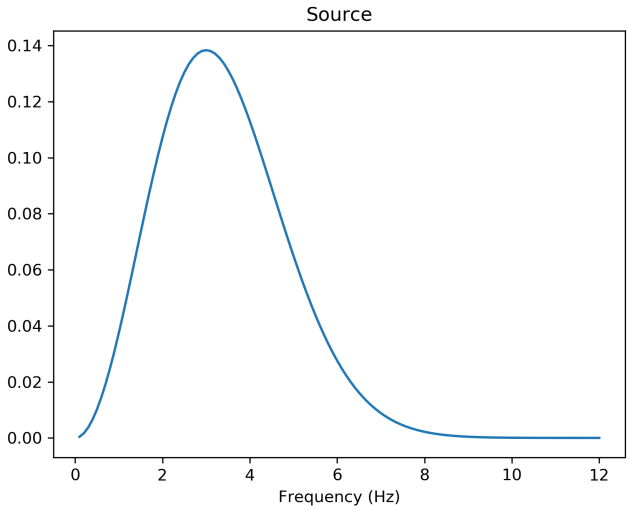

Figure 3.1: Ricker wavelet as the source wavelet with a peak frequency of 3Hz in the frequency domain.

As a preliminary test to verify the reliability of seismograms created using the forward scheme, we considered a simple two-layer model; the first layer with a velocity of $1500ms^{-1}$ and the second layer with a velocity of $1600ms^{-1}$ with the interface at a depth of 1000m. Data was generated form this model in the frequency domain and then transformed into the time domain to obtain the seismogram. In the seismogram, we expect only a single reflection event corresponding to the interface of the two layers. Since we know the depth of interface, and offset, we can calculate the arrival time for the event at each receiver using the equation given by;

$$t = \sqrt{t_0^2 + \frac{x_{off}^2}{v_1^2}}, \tag{3.1}$$

where t_0 is the zero-offset two-way traveltime, x_{off} is the offset, and v_1 is the velocity of the first layer. The above equation is the arrival expression for the reflection event from a horizontal reflector for two-layer medium. This expression is used in seismic data processing for normal moveout (NMO) correction, which is done before CMP stacking in order to make events arrive at the same time.

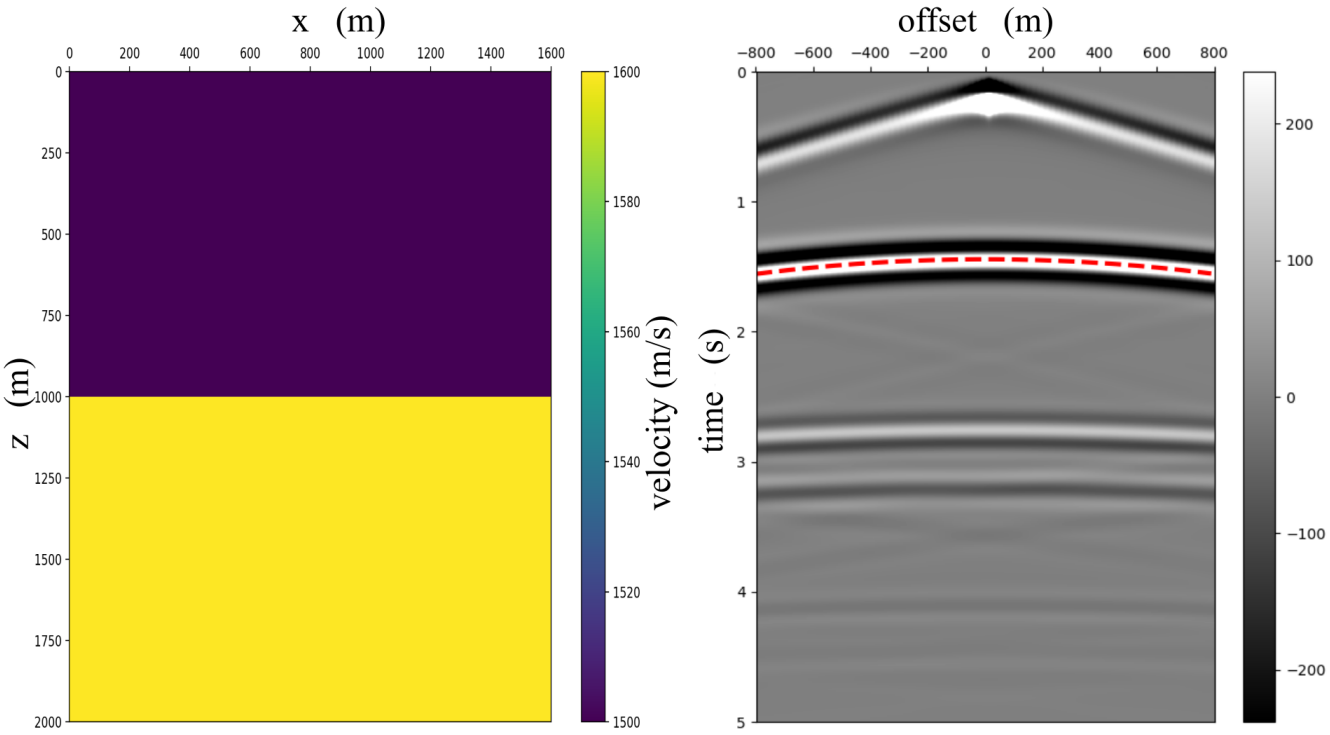

Figure 3.2: Simple two layer velocity model (left) and overlay of the estimated arrival time (red dash line) of reflection event on the generated seismogram (right). Since the source wavelet is not a zero-phase wavelet a shift of 0.180s is applied to the estimated arrival time.

In figure 3.2, the left panel corresponds to the two-layer velocity model, and the colorbar represents the velocity values inside the domain. The right panel shows an overlay of the seismogram with the arrival time (red dashed line) of the reflection event calculated using the above equation. Since the source wavelet is not a zero-phase wavelet, a time shift of 0.180s (180ms) is applied and it is evident that the calculated arrival time matches precisely with the reflection event in the seismogram. The seismogram shown here is not the raw record, but an amplitude correction is applied to visualize each events.

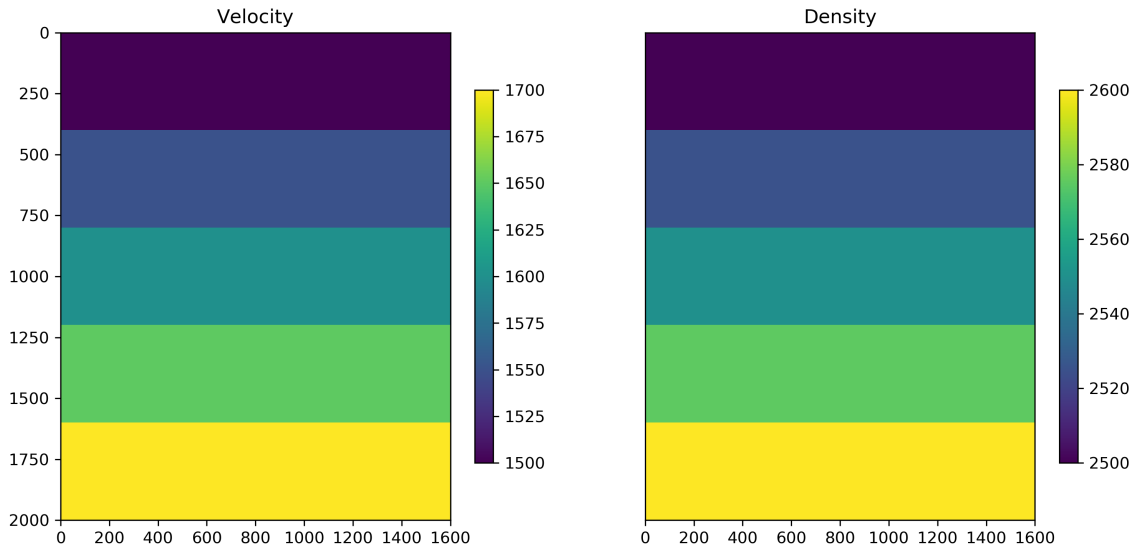

Figure 3.3: Layered velocity and density (model 1) used to create the forward simulation. The velocity of the top layer is $1500 ms^{-1}$ and it increases by $25 ms^{-1}$ with each layer. The density of top layer is $2500 kgm^{-3}$ and increment is $50 kgm^{-3}$.

For the next test, a five-layer model, as shown in figure 3.3 with each layer of 20 gridpoints in z-direction is considered. The velocity varies from 1500 to 1700 ms^{-1} with an increment of $50 ms^{-1}$ and density ranges from 2500 to 2600 kgm^{-3} with an increment of $25 kgm^{-3}$. Since this study is done using fourth-order FD approximation on staggered-grid (resulting in a 13-point stencil), in order to get a stable result, there should be at least four gridpoints per shortest wavelength. The minimum wavelength is 125 m (minimum velocity=$1500 ms^{-1}$, maximum frequency=12Hz), so we considered 20m for grid spacing in both x- and z-directions, which effectively includes more than six gridpoints in one wavelength. For forward simulation, the source is located at x=800m at a depth of 5m, and the receivers are placed at a depth of 10m at every 10m throughout the x-direction. The forward modeled data for the above model in the frequency domain (for some selected discreet frequency values) is given in figure 3.4. The wavefields corresponding to each frequencies are solved independently, so the problem of error propagation is is addressed to a good extent as compared the time domain approach.

The data in frequency domain, $D(x, z, \omega)$ can be converted to time domain $D(x, z, t)$ by taking the Fourier transform (Bailey and Swarztrauber, 1994) given by;

$$D(x, z, t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} D(x, z, \omega) e^{-i\omega t} d\omega \tag{3.2}$$

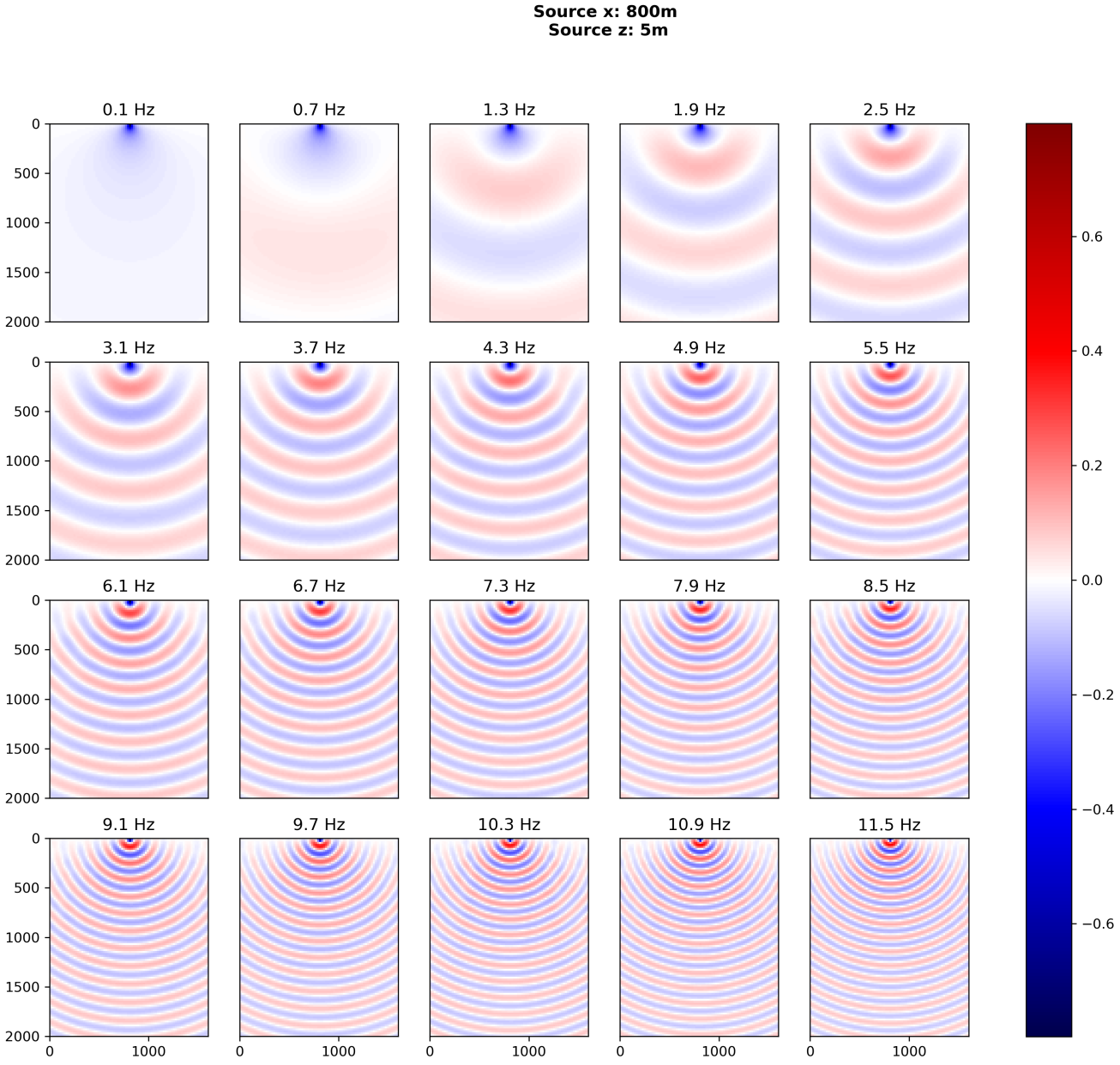

Figure 3.4: Real part of pressure wavefields for selected frequencies generated from model 1. x-axis of each panel represents the x-location and y-axis represents the depth.

The seismogram, after the Fourier transform of frequency data for model 1, is shown below.

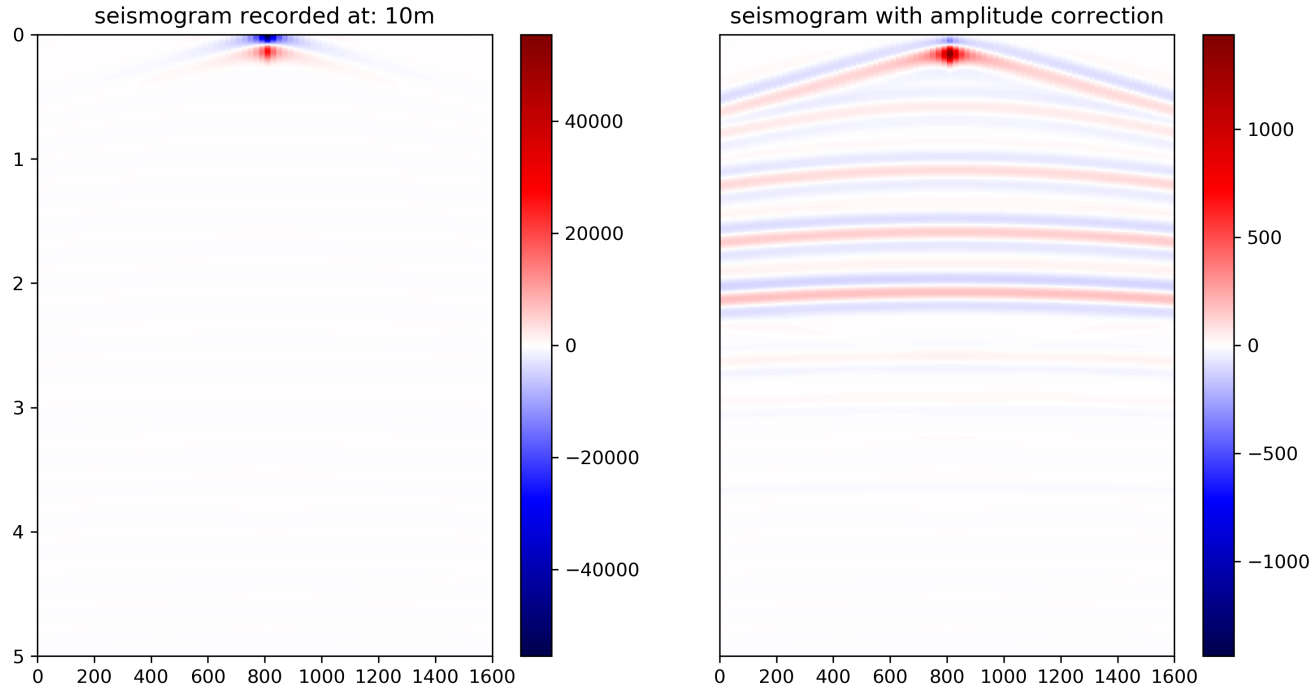

Figure 3.5: Seismogram after Fourier transform of frequency data obtained from five layer model. The left panel represents the raw seismogram, and the right panel represents the amplitude corrected seismogram.

In figure 3.5, the left panel shows the actual time series (raw seismogram) recorded at the receivers. The right panel corresponds to the seismogram after amplitude correction (after multiplying the time series with 1.5 power of time values, divergence correction). In the right panel, the direct arrival, reflections from each layer boundaries, and some multiples are clearly observed.

For the next test, a model as shown in figure 3.6 is considered with a background velocity of $1500ms^{-1}$, and a block buried inside with a velocity $1600ms^{-1}$ is considered. In this test, the source was placed at x=10m, and z=5m, and the generated pressure wavefields in the frequency domain are given in figure 3.7. The seismogram in the time domain generated from model 2 is shown in figure 3.8.

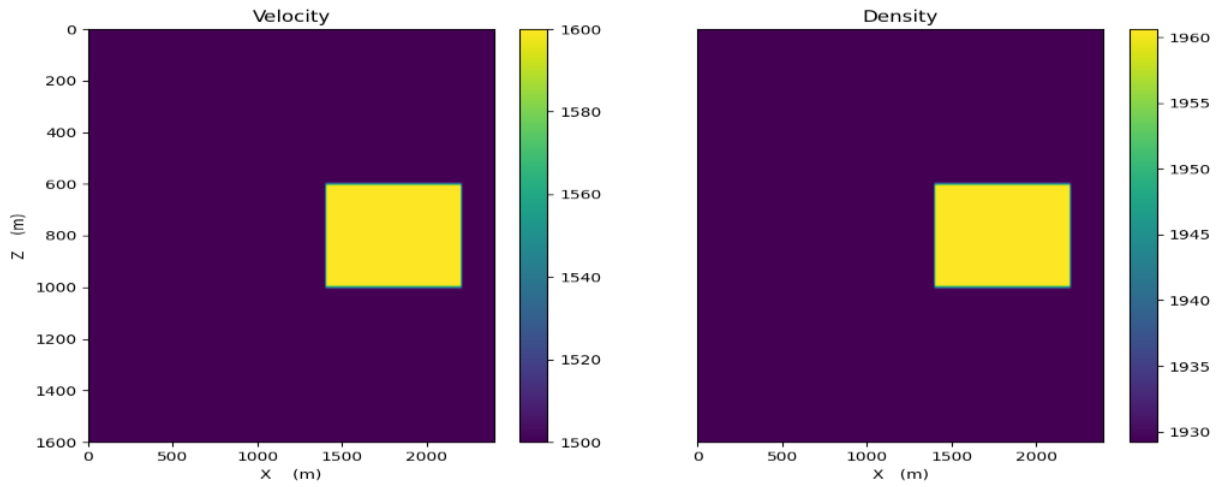

Figure 3.6: Model 2 used for checking diffraction events produced by the corner of the buried body. The background velocity is of $1500ms^{-1}$ and velocity of the buried body is $1600ms^{-1}$. The right panel represents the density.

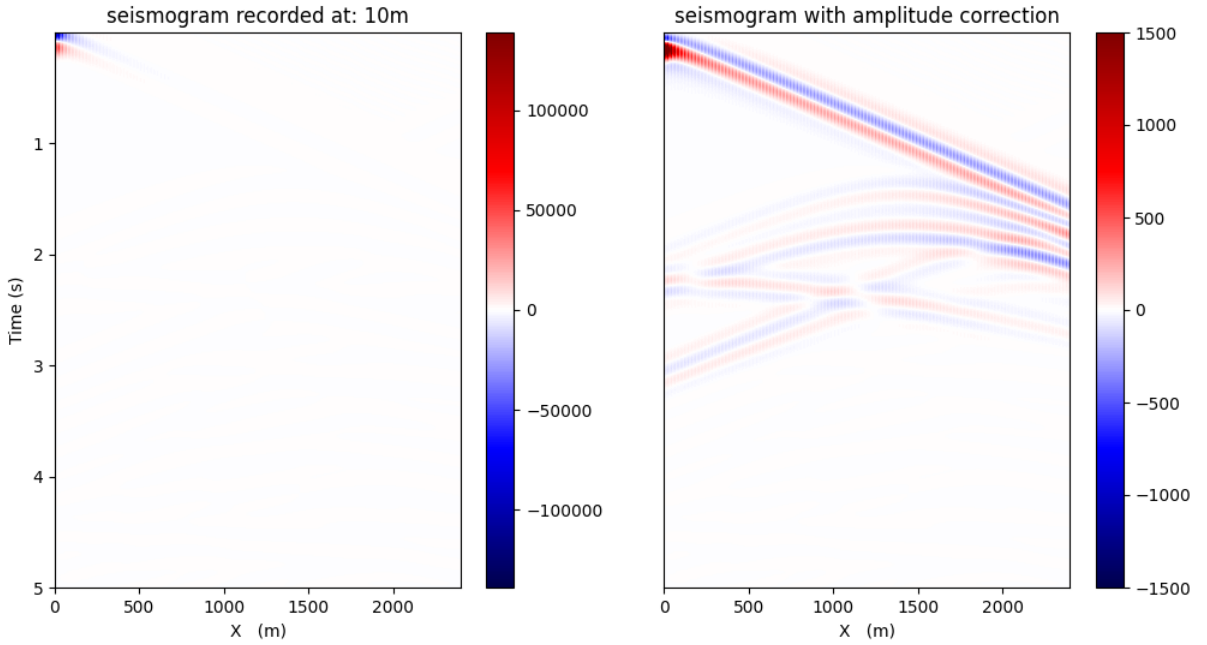

Figure 3.8: Seismogram generated from the buried box model. The left panel represents the raw seismogram and the right panel represents the amplitude corrected corrected seismogram.

Similar to the previous case, the left panel in figure 3.8 corresponds to the raw seismogram, and

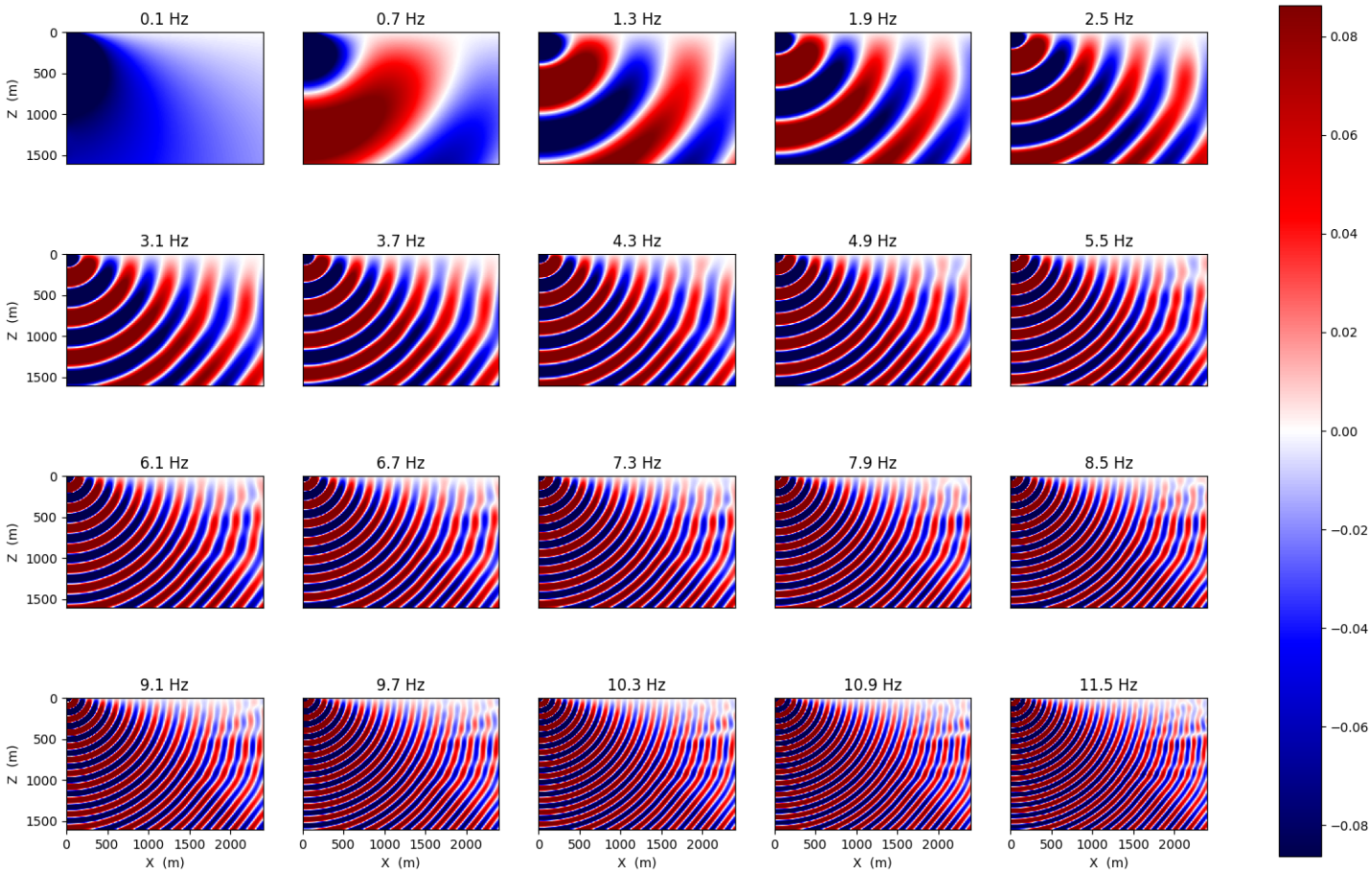

Figure 3.7: Real part of pressure wavefields for selected frequencies generated from model 2.

the right corresponds to amplitude corrected seismogram. In the right panel, the direct arrival is clearly visible. The reflected events from the bottom and top surfaces of the buried box are also observed. Apart from these, the diffraction events from the top and bottom corners of the box are also present in the seismogram.

From the above two tests, the forward algorithm was able to generate all the possible seismic events for each model, and it is important to note that there were no reflected events from bottom/right/left boundaries, which proves that the PML boundary condition is effectively implemented in the forward algorithm. If the boundary condition is not properly implemented, there will be remnant energies coming from the edges. Since it is not the case for previous results, we can infer that the boundary layers are absorbing all energies and provides an unbounded domain for modeling.

Apart from these above numerical examples, we also attempted to further establish the robustness of the forward scheme by comparing our result with the result available at "https://reproducibility.org/RSF/book/data/marmousi/paper.pdf". Here we compare the data generated on the Marmousi model using our forward algorithm and data generated using Madagascar. The comparison of the modeled data is given below. In the example, the shot was fired at x=5000m at a depth of 10m. The receivers are placed every 12.5m at a depth of 25m. In figure 3.9, the right panel corresponds to the data generated using our forward scheme, and the left panel is taken from the above document. If we compare them, it is observed that both panels have similar reflection and diffraction events. They closely resemble to each other. Towards the right bottom of both panels, the transmitted waves are distinctly visible. Consistent diffraction events on both panels are observed over arrival times between 0.6 and 1s. This close resemblance between the above panels indicates that the data generated using our forward scheme is reliable and robust.

All the above tests establish the robust performance of the forward algorithm. From the first experiment, we verified the arrival time of the reflection event for a simple model. Then the next two tests show that the forward engine is able to generate all the possible seismic events. Then the last comparison of data generated using the Marmousi model also points in favor. From all these tests, we are confident that the created forward scheme is able to handle the acoustic approximation well enough with appropriate boundary conditions and is capable of generating realistic data for waveform inversion.

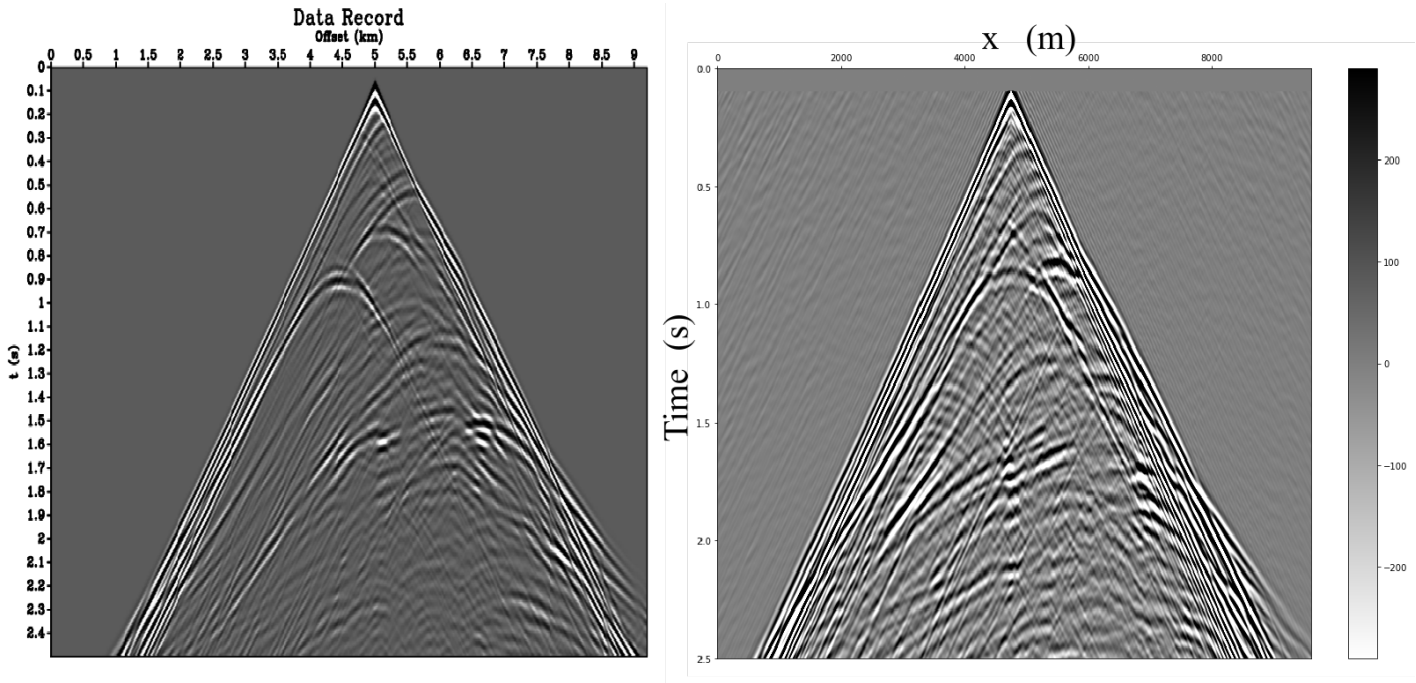

Figure 3.9: Comparison of seismogram from the Marmousi velocity model. The left panel represents the data generated using Madagascar, and the right panel represents the data generated using our forward algorithm.

3.2 Full Waveform Inversion

For the initial inversion tests, a synthetic model is considered, and data is generated in the frequency domain for four frequencies. These data are used for inversion using the GN algorithm. For the first test, a model with two boxes with velocity $1600ms^{-1}$, buried inside a background velocity of $1500ms^{-1}$, is considered. The model extends up to 2000m in the x-direction and 1200m in the z-direction. One of the two boxes is buried at a shallow depth at z=150m and the another at a deeper depth at z=900m. A half-space model with a velocity of $1500ms^{-1}$ (background velocity, no information about the boxes) was considered as the initial model. In the beginning, the inversion is done only using frequencies 0.5, 1, 1.5, and 2Hz. The result from the inversion comparing with the initial and true model is given below.

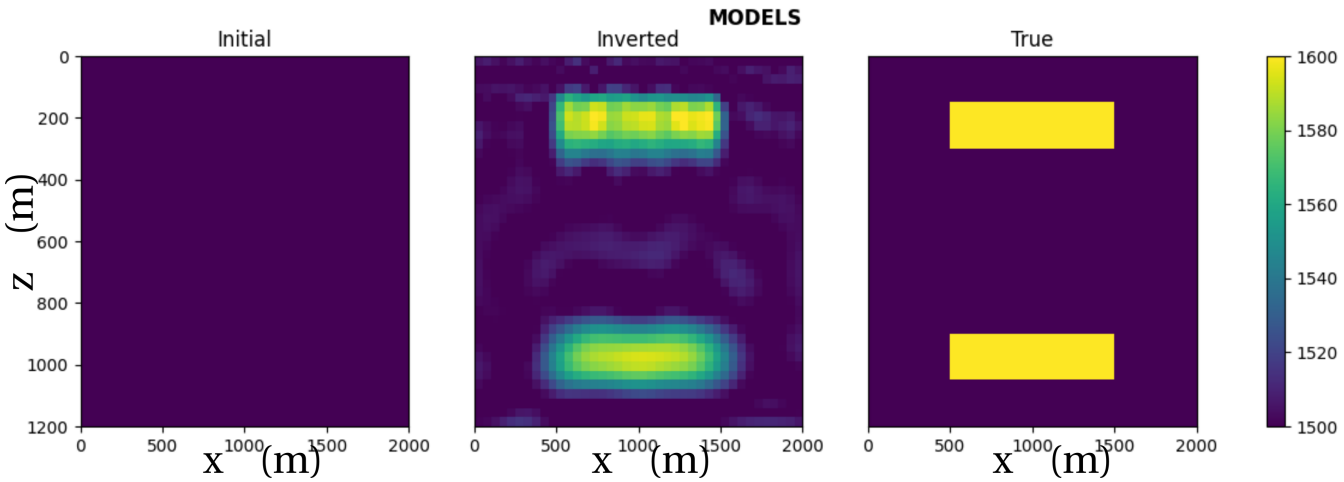

Figure 3.10: Inversion test with two buried boxes using discreet frequencies 0.5, 1, 1.5, 2Hz.

In figure 3.10, the left panel indicates the initial model, the middle panel shows the inverted model, and the right panel displays the true model. If we compare the inverted model and true model, it is very evident that without any prior information about the buried blocks, the inversion scheme was able to detect them and shows a close match with the true model. In figure 3.11, the left panel illustrates the horizontal profile of the velocity at a constant depth of 210m, and the right panel shows the vertical profile of the velocity at a constant x-location; x=1500m. In both panels, the green line represents the inverted profile, blue represents the true profile, and orange represents the initial profile. It is observed that the green and blue curves show a close match up to a good extent.

For the second test, the effect of the number of discrete frequency values are investigated. So in

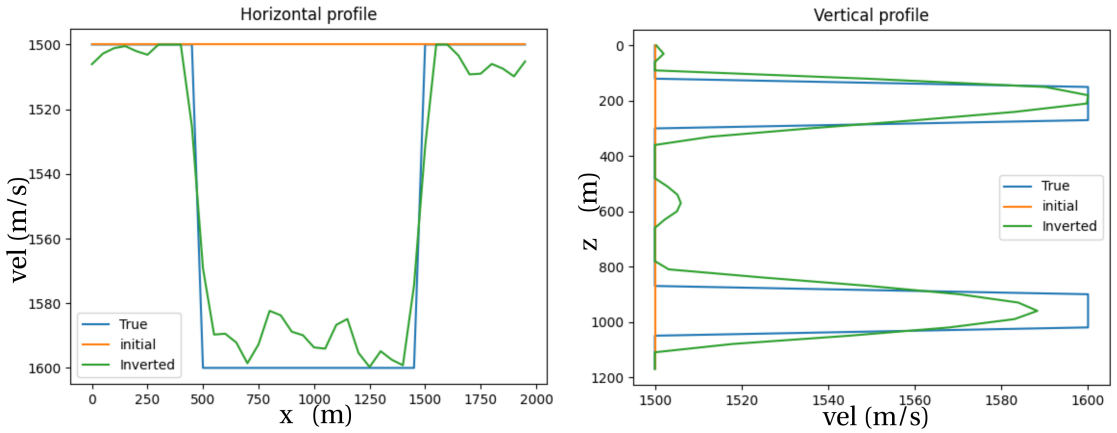

Figure 3.11: Velocity profile comparison of models. The left panel represents the horizontal profile at z=210m, and the right panel represents the vertical profile at x=1500m (test 1).

this test, apart from the above frequency values, 4Hz and 6Hz are also used. The inversion result from the test is given below.

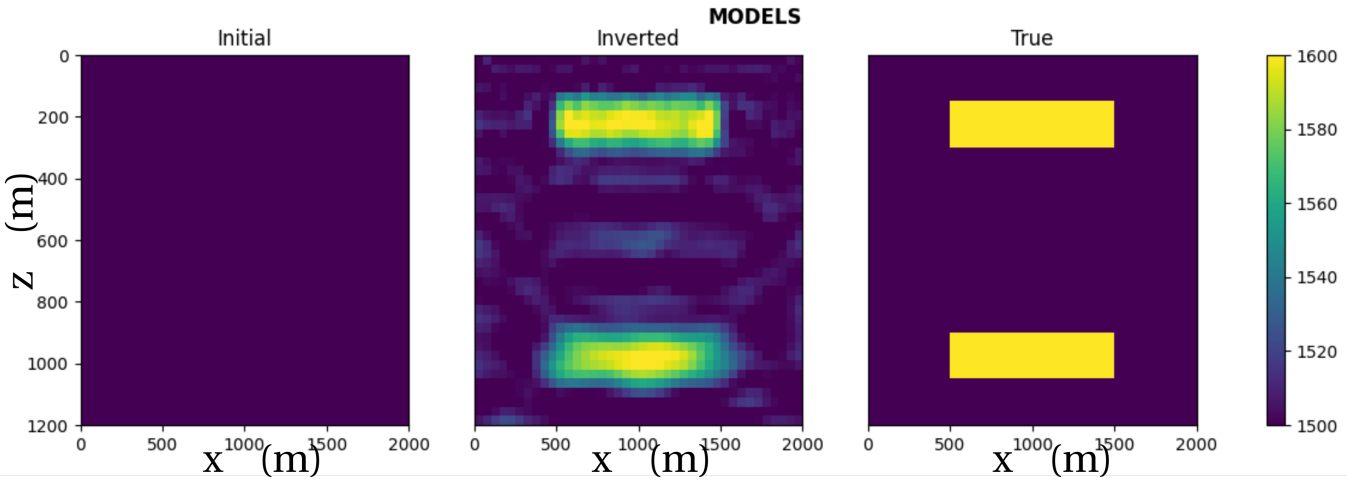

Figure 3.12: Inversion test with two buried boxes using discreet frequencies 0.5, 1, 1.5, 2, 4, 6Hz

If we compare figure 3.10 and figure 3.12, in figure 3.12, the buried box at the shallow depth is defined sharply; however it contains artifacts in the background velocity region between the boxes. In order to better visualize the result, we could look at a particular vertical profile going through both boxes. The right panel in figure 3.11 and figure 3.13 shows the vertical profile. In both images, both the true and inverted profiles match inside the boxes, but due to the presence of higher fre-

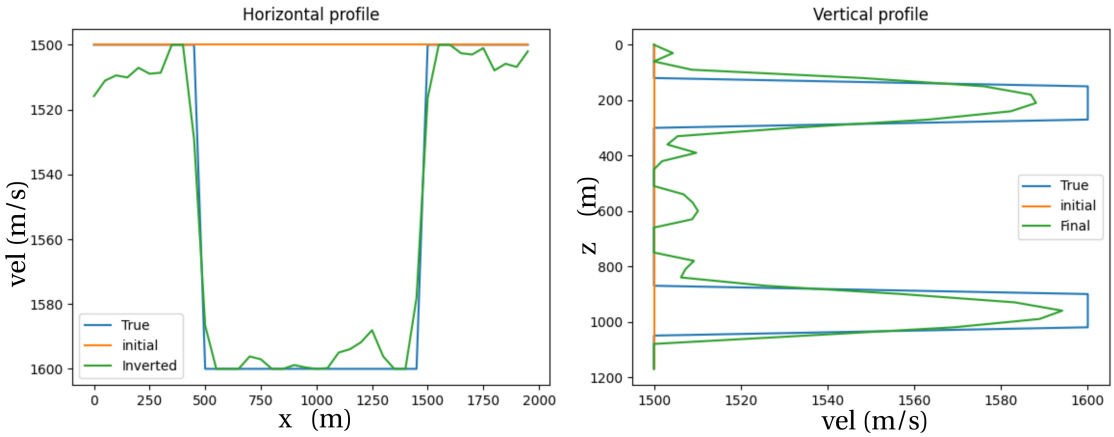

Figure 3.13: Velocity profile comparison of models. The left panel represents the horizontal profile at z=210m, and the right panel represents the vertical profile at x=1500m (test 2).

quencies, the inverted profile shows a wiggly nature between the boxes in figure 3.13. Conversely, the horizontal profile points that the presence of higher frequency values is better, and the inverted profiles match more correctly with the true. This test shows that for estimating the velocities with higher accuracy in the horizontal direction, higher frequency values are necessary, however, it may affect the vertical information.

For the next test, we considered a synthetic model with several dipping reflectors and one body with high-velocity ($1650 ms^{-1}$) at a shallow depth and another body with low-velocity($1400 ms^{-1}$) at a deeper depth. The domain spreads 4400m in the x-direction and 600m in the z-direction. For the inversion, the initial model is considered without the two bodies.

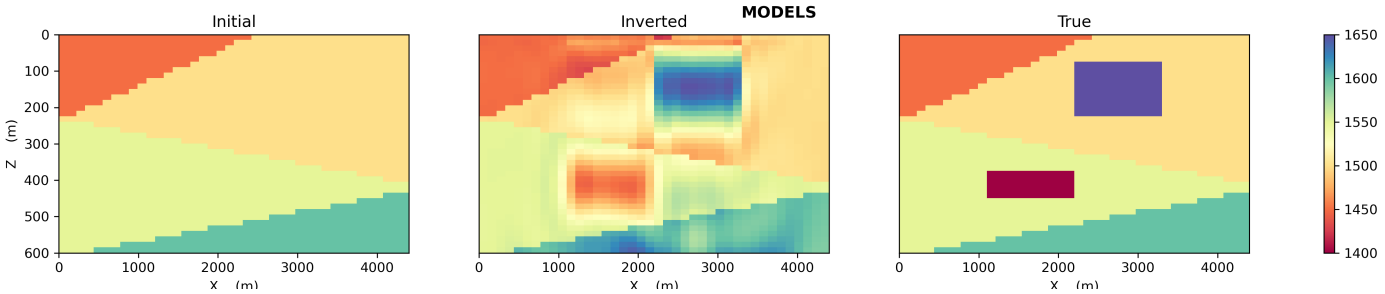

Figure 3.14: Test on detecting shallow high-velocity and deep low-velocity bodies using GN. The panels from left to right represent the initial, the inverted, and the true model, respectively. The colorbar gives the velocity information in ms^{-1}.

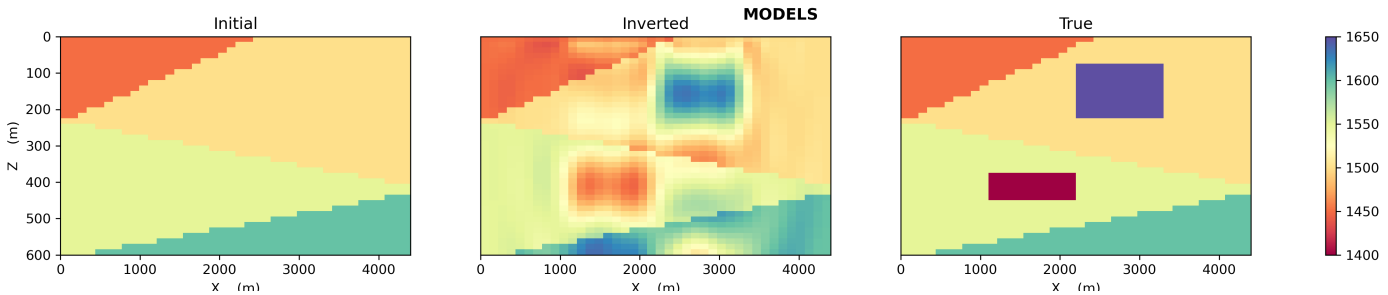

Figure 3.15: Test on detecting shallow high-velocity and deep low-velocity bodies using NLCG with the same source-receiver configuration and optimization parameters as GN.

In figure 3.14 and 3.15, the right panel represents the true velocity model with dipping reflectors and the two anomalous bodies; the left panel represents the initial model without the bodies. The middle panel represents the inverted model. To compare the results with GN and NLCG, both inversions were performed with the same configurations of source and receivers as well as the same optimization parameters. Figure 3.14 represents the inversion result using the GN method, and figure 3.15 represents the results using NLCG. In figure 3.14, both the shallow and the deeper bodies are retrieved with good accuracy. In figure 3.15, the NLCG scheme has successfully reconstructed the anomalous bodies. However, the GN result is giving sharp boundaries of the bodies, and the NLCG result has artifacts that looks like the velocity of the bodies are smeared with the background velocities. It is also observed that the bottom layer in the inverted model has some wrong update in the NLCG result, even though the modeled data closely matches the actual data. This clearly shows the non-uniqueness of the model parameters. The shallow body is reconstructed with very higher accuracy in the GN result than NLCG.

The horizontal and vertical profiles of these models at selected depth and x-location for GN and NLCG, respectively, are shown in figure 3.16 and 3.17. The left panels represent the horizontal profile at a constant depth of 150m. The profile passes through the shallow high-velocity body. The body can be seen as a high-velocity zone between x=2200 and x=3200. In figure 3.16, the inverted profile (green) almost perfectly matches the true profile (blue) inside the body. It is to be noted that the inverted profile perfectly matches the true profile so that it gives a sharp definition of the boundaries of the body. However, in figure 3.17, the matching is not as perfect as GN. The sharp edge is not reconstructed using NLCG. In the right panel, the vertical profile through a constant x-location, x=1650m, passing through the deep low-velocity body, both methods give a good match of the inverted profile with the true profile, but the amount of update is insufficient in both cases. This test demonstrates the ability of FWI to detect bodies with anomalous velocities at variable

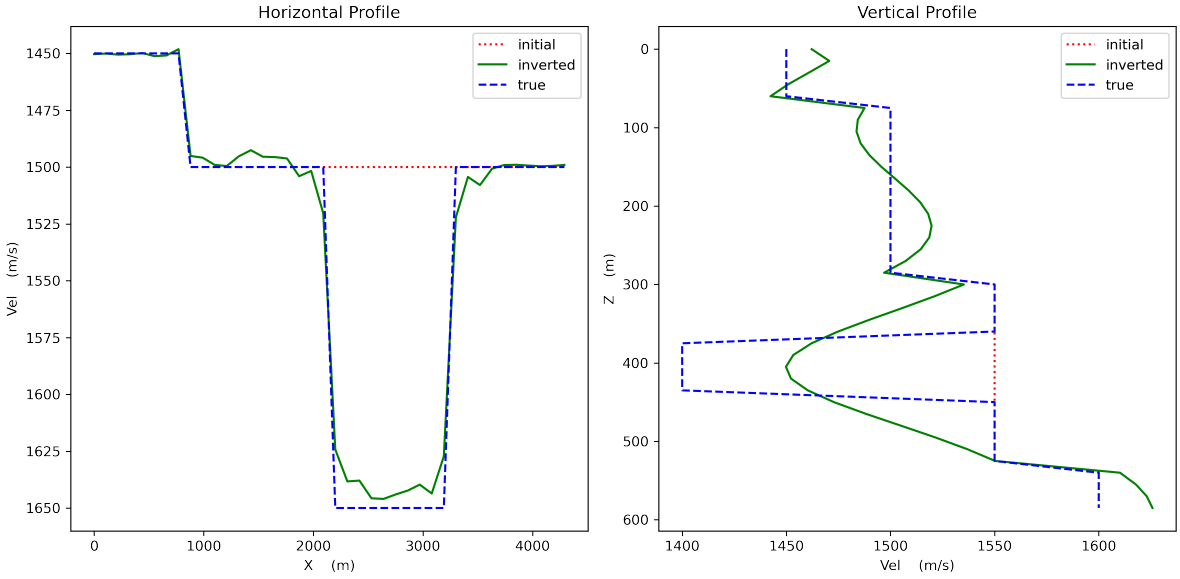

Figure 3.16: Comparison of the velocity profile of models from GN method. The left panel represents the horizontal profile at z=150m, and the right panel represents the vertical profile at x=1650m.

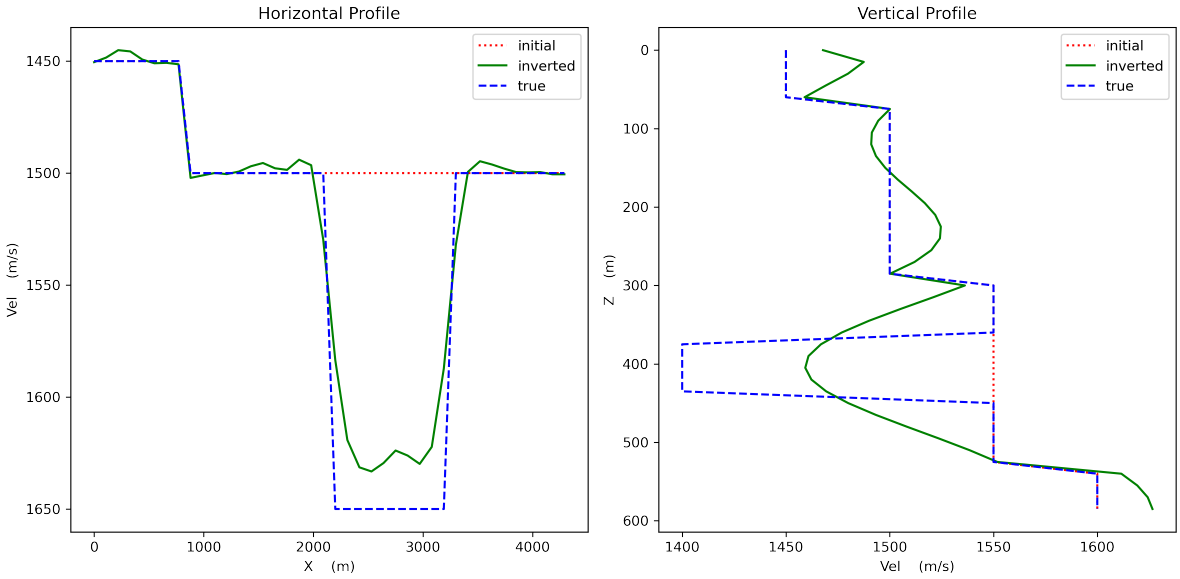

Figure 3.17: Comparison of the velocity profile of models from NLCG method. The left panel represents the horizontal profile at z=150m, and the right panel represents the vertical profile at x=1650m.

depths inside the subsurface.

The evolution of error between modeled and observed data from the beginning to the end of inversion for both methods are shown below,

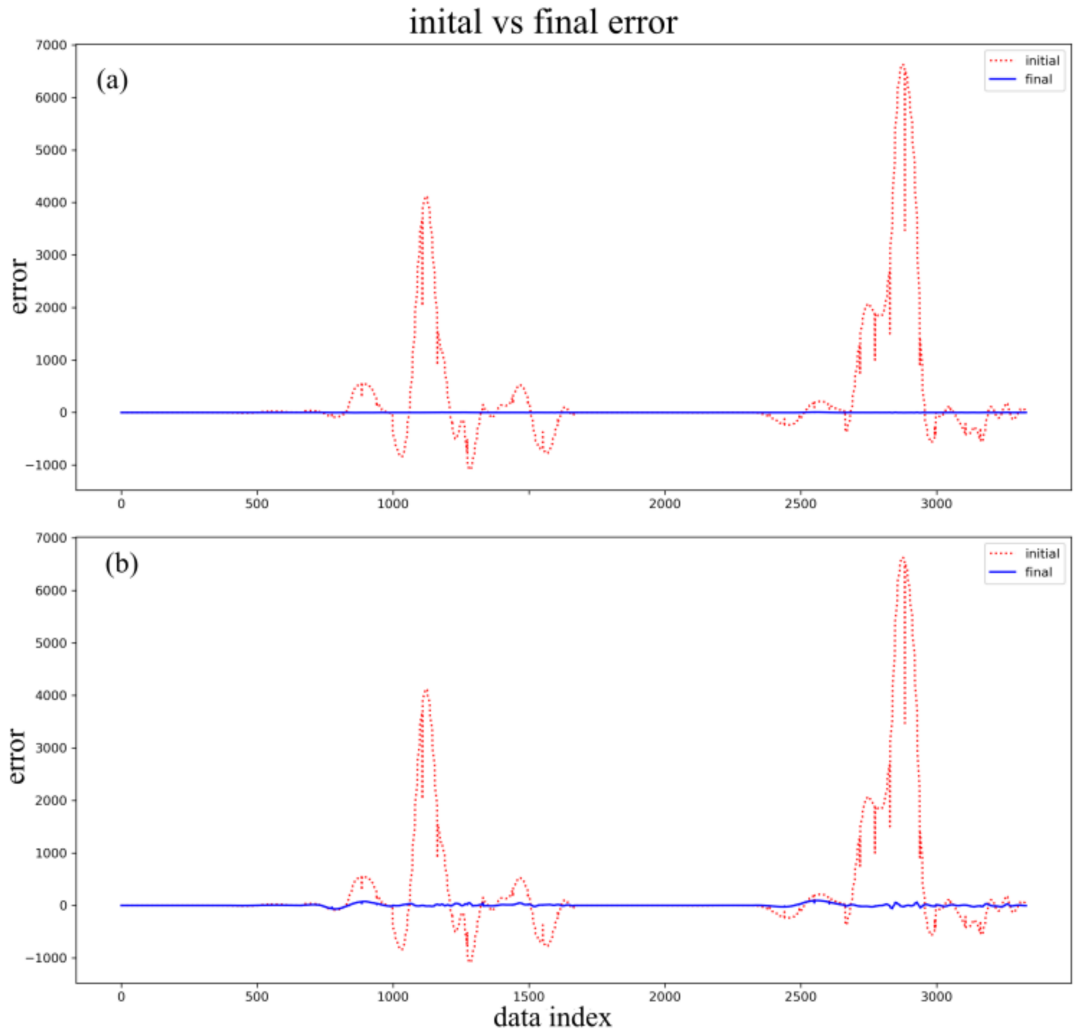

Figure 3.18: Difference between data and synthetic data from the initial model (dotted red) vs difference between data and synthetic data from the inverted model (solid blue) from the above synthetic model usingg (a) GN and (b) NLCG.

To check the quality of the update retrieved by the FWI scheme, it is also possible to look at the difference between the observed data and the data modeled using the initial model (initial error) and the difference between the observed data and the data modeled from the inverted model (final

error). The difference is calculated sample by sample, so the x-axis represents the sample index and y-axis is the error. In figure 3.18, the dotted red curve represents the difference between observed data and modeled data using the initial model, and the solid blue curve represents the difference between observed data and modeled data using the inverted model. The misfit plots demonstrate a good agreement between observed and inverted data for GN. In NLCG, there is a slight variation from zero for some data samples; however, the misfit is within the tolerance limits. This NLCG result was obtained after 30 iterations, and the GN result was obtained after 8 iterations.

For the next test, a domain of 750m in the z-direction and 5600m in the x-direction with a discretization of 20m in z-direction and 70m in x-direction is considered. For this test, the main objective is to rigorously check the robustness of the inversion algorithm, so the domain is populated with many blocks having checkers pattern. The size of each block is 195m in z-direction and 910m in x-direction. The checkers model contains blocks of velocity $1450ms^{-1}$ and $1550ms^{-1}$ alternately placed adjacent to each other. The inversion is done using the discreet frequency values 0.1, 0.2, 0.4, 0.9, 2, 4.5Hz. In order to rigorously and firmly test the algorithm, no prior information on the velocity of the checkers or the background velocity is incorporated in the initial model; i.e., the initial model is considered as a uniform velocity model of $1500ms^{-1}$.

In figure 3.19, the left column represent the initial model of uniform velocity ($1500ms^{-1}$) without any prior information about the true model. The middle column represent the inverted model, and the right column represent the true checkerboard model. The top row corresponds to GN inversion and bottom row corresponds to NLCG. It is very evident that the inverted models have tried to correctly retrieve the information about the checkerboard model over the whole inversion domain. If we compare the GN and NLCG results at shallower depths, GN result is superior to the NLCG result. As we go into deeper depths, both results show a degrading trend in the quality of updates. Similar to the previous synthetic model tests, GN is able to better delineate the edges/ boundaries of the checkers than NLCG. The NLCG presents a smearing of velocity like artifacts significantly as compared to GN, even at shallow depths.

The horizontal and vertical profiles of these models at selected depth and x-location, respectively, for GN and NLCG are shown in figure 3.20 and 3.21. The left panels represent the horizontal profile at a constant depth of 225m. The true profile alternates between $1450ms^{-1}$ and $1550ms^{-1}$ and the initial profile is a constant velocity of $1500ms^{-1}$. The inverted profile (green) matches the true profile (blue) to a good extent in both cases. However, the inverted profiles exhibit an oscillatory nature, and the amount of update seems insufficient. The oscillatory nature is more prominent in

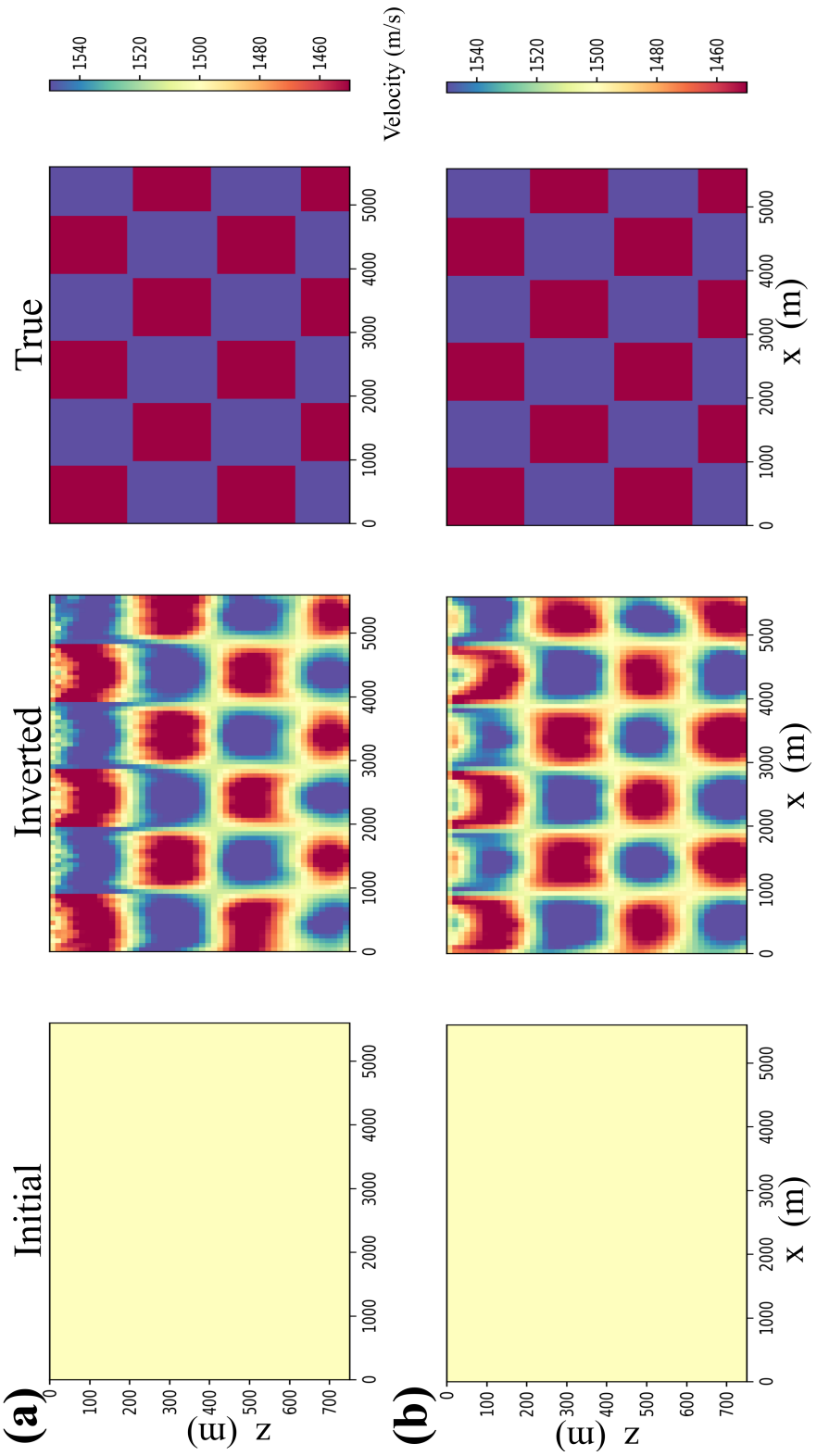

Figure 3.19: Inversion results from checkerboard test using (a) GN and (b) NLCG methods. The checkers have a dimension of 910m×195m (x × z). The left panels represents the starting model of a uniform velocity of $1500 ms^{-1}$, the middle panels represents the inverted model from each method, and right panels represents the actual model.

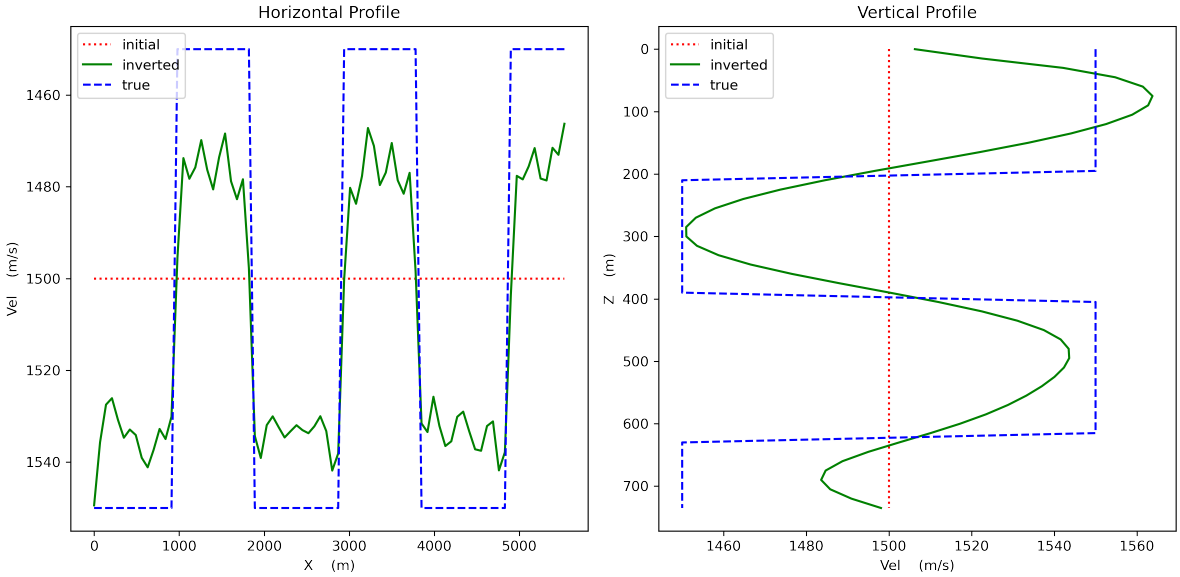

Figure 3.20: Velocity profile of models from GN. The left panel represents the horizontal profile at z=225m, and the right panel represents the vertical profile at x=1050m.

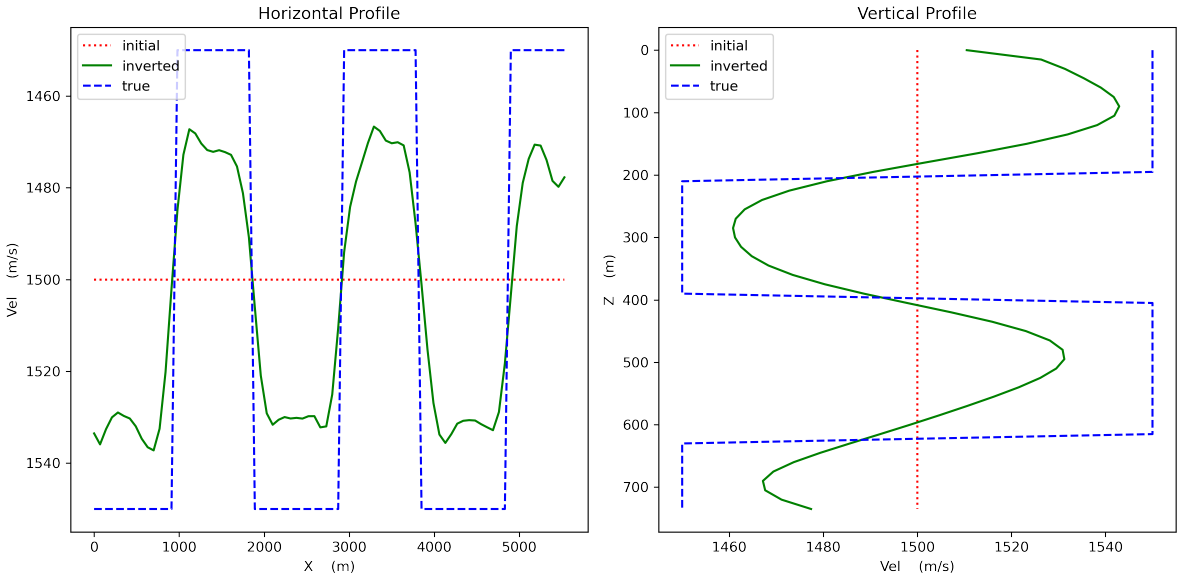

Figure 3.21: Velocity profile of models from NLCG. The left panel represents the horizontal profile at z=225m, and the right panel represents the vertical profile at x=1050m.

the GN result. In the right panels, the vertical velocity profile at x=1050m is shown for both results; the inverted profile closely matches the true profile with more correctness at the shallow depths, and at deeper depths, the amount of update is not enough for a good match, but the direction is appropriate. In the vertical profile, the amount of update is significantly insufficient in the NLCG result. In GN, at least at the shallow region, the update is sufficiently close to the true velocity value.

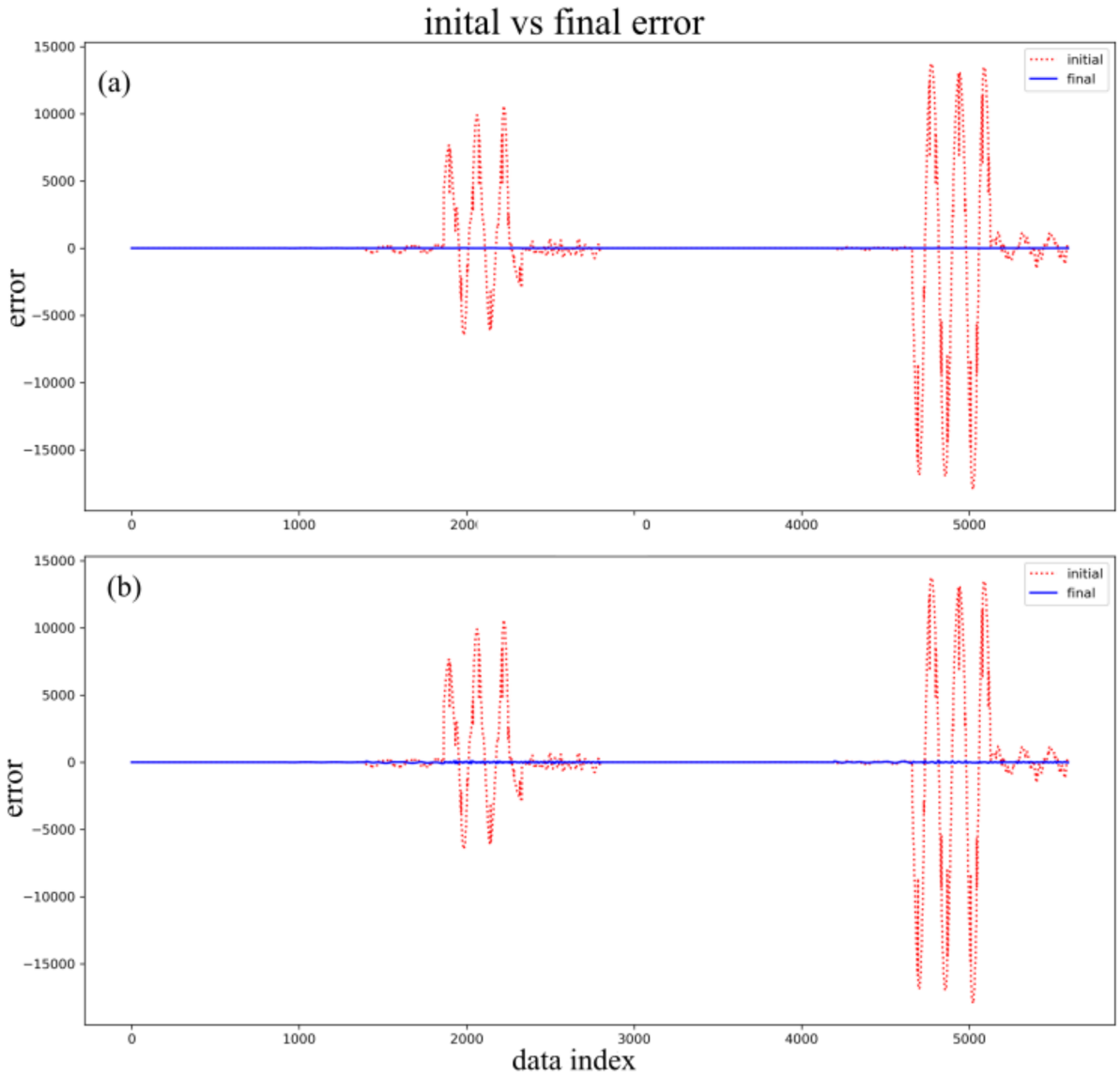

Figure 3.22: Difference between data and synthetic data from the initial model (dotted red) vs difference between data and synthetic data from the inverted model (solid blue) from checkerboard test using (a) GN method and (b) NLCG.

Figure 3.22 represents a comparison between the initial and the final error corresponding to GN and NLCG methods, respectively. It is clearly visible that the initial error lies between ±15000, however, the final error is almost zero for all data points in both cases. This clearly points out that the misfit function has been reduced to a minimum and the modeled data closely matches the actual data. If the final error curves of both methods are compared, GN shows a consistently better fit for all samples than NLCG. This points to fact that the GN method is able to better reconstruct the model parameters than NLCG.

The NLCG inversion result was obtained after 40 iterations, whereas the GN inversion result was obtained after 8 iterations. The normalized residual value over iteration for both methods is shown below, and it evidently demonstrates the superior convergence rate of GN toward the minima.

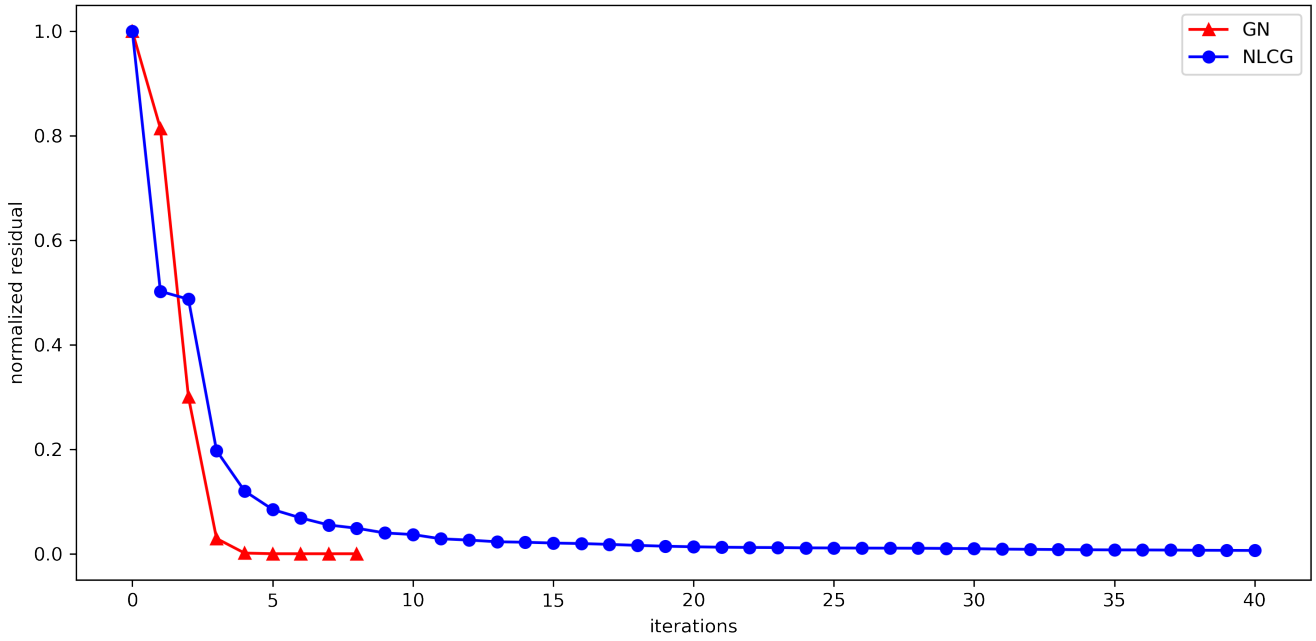

Figure 3.23: Normalized residual over iterations for GN and NLCG in checkerboard test. The red triangles correspond to the GN method and blue dots correspond to NLCG.

The following table3.1 and 3.2 summarizes the details of GN and NLCG inversion schemes respectively. The nRMS error is defined to assess the quality of update and it is defined as,

$$nRMS = \frac{1}{\sqrt{N_d}} \sqrt{\left[\frac{d^{obs} - d^{pred}}{abs(d^{obs})}\right]^T \frac{d^{obs} - d^{pred}}{abs(d^{obs})}} \tag{3.3}$$

For ideal case it will be near to zero. In our case, the values are near to zero, this could be because the data did not contain any noise.

Gauss-Newton			
Model	nITER	$\frac{err_{ini}}{err_{fin}}$	nRMS error
model 1 (synthetic model with anomalous bodies)	8	9.48e-7	0.011374
model 2 (checkerboard test)	8	5.47e-7	0.115504

Table 3.1: Summary of GN inversion for both models. nITER represents the number of total iterations performed and nRMS represents the normalized final rms error (i.e. ratio of initial to final error).

Non-linear conjugate gradient			
Model	nITER	$\frac{err_{ini}}{err_{fin}}$	nRMS error
model 1 (synthetic model with anomalous bodies)	30	1.89e-2	0.195804
model 2 (checkerboard test)	40	3.45e-3	0.137496

Table 3.2: Summary of NLCG inversion for both models.

In the previous two inversion tests, a rigorous comparison of inversion using the Gauss-Newton method and the Non-Linear Conjugate Gradient method is performed. In both cases, it is consistently observed that GN provides a sharper boundary of anomalous bodies and converges to minima at a faster rate than NLCG. It was observed that The NLCG results are susceptible to smearing of velocity resulting in blurred boundaries of the bodies. The inferences are completely based on the previous two tests, and it is not fair to establish a final conclusion regarding the superiority of the methods based on these two tests only. Both the algorithms require rigorous tests on benchmark models such as Marmousi or SEG/EAGE salt models and then apply both algorithms to real seismic data. More studies must be done on establishing stable optimization parameters by performing numerous tests with both methods.

In figure 3.2, the calculated arrival time matches with the reflection event in the seismogram. In figure 3.8, it is very evident that the forward engine is able to generate all the possible events from model 2 (figure 3.6) with no reflection from the boundaries. In figure 3.9, a comparison between generated data with data from another source is presented, and it is clear that both of the seismograms resemble closely, indicating our forward modeling scheme is able to accurately generate the data. All these examples of forward modeling results prove the reliable and robust performance of the scheme. If the data is generated without implementing proper boundary conditions, then strong unwanted reflection events from boundaries will be generated, and these kinds of event resulting from improper boundary conditions will tremendously tamper the quality of the inversion result. Since the authenticity of the forward engine is accessed and evaluated thoroughly, we moved to the inversion scheme. The inversion algorithm is successfully applied to several synthetic models. The inversion scheme was able to retrieve the information about the buried box (figure 3.10 and 3.12) starting from the initial model without having any prior information on the boxes. Even though the initial goal of the project was to create the inversion algorithm using the Gauss-Newton method, we also got the opportunity to explore the conjugate gradient methods. So we tried to create the inversion algorithm using the Non-Linear Conjugate Gradient method also. Figure 3.14 and 3.15 present the inversion results of synthetic test on detecting anomalous bodies using GN and NLCG methods, respectively. And figure 3.19 and ?? present the inversion results on checkerboard tests. The checkerboard test proves the robustness of both methods by successfully reconstructing the velocity information starting from no prior information. The checkerboard test is often widely used to assess the performance and competency of the inversion algorithms rigorously. Both methods gave comparable results, however, the convergence rate of GN is far superior. NLCG requires more iterations to converge to minima to reach comparable results of GN,however, the computation cost per iteration is less in case on NLCG.

4 ‖ Conclusion

Through this project, we made an attempt to develop the algorithm of 2D FWI of seismic data in the frequency domain, initially using the GN method and later using the NLCG method. The acoustic forward modeling was implemented using the fourth-order finite difference scheme on staggered-grids, with PML boundary conditions to create an unbounded domain with very little to no numerical error. The choice of frequency domain was made by accounting for its low computational cost for 2D cases due to the fact that only discreet frequencies are used for the inversion. If we use all the frequencies, it would be equivalent to implementing the inversion in the time domain. In the GN method, the second derivative of the misfit function, the Hessian matrix, is approximated in terms of the Fréchet derivative of the forward operator, the Jacobian matrix. Estimating the accurate Hessian is a time and resource-consuming task and sometimes nearly impossible to estimate. GN being a Newton method presents a faster rate of convergence to minima as compared to gradient-methods such as steepest descent. The gradient-based approaches require much lesser computational resources per iteration. The NLCG is a gradient-based method and demands less computational resources than the GN method, however, it requires more iterations to reach comparable results. In NLCG, the model update is performed using the gradient and a step length estimated using line search methods.

Through the forward simulation tests, we have established the sound quality of data generated using the algorithm. The comparison of data generated on the Marmousi model using Madagascar and our algorithm gives confidence in our scheme. The boundary condition was implemented with proper parameters so that we observe no remnant energies reflecting from the edges of the domain. The forward scheme is incorporated in the inversion scheme to generate the synthetic data for matching with the recorded data. The inversion algorithm is developed using GN and NLCG methods. We also have presented the results from both methods and done a comprehensive analysis of the inverted result from both methods. From the results obtained on these tests, we inferred that the GN method outperforms the NLCG in terms of the quality of the update as well as the number of iterations. It was consistently observed that the NLCG results appear to have a smearing of velocity artifact as compared to GN results. Also, the quality of both results at deeper depths diminished due to poor illumination. In the horizontal profile of GN results, we observe oscillatory behavior that require further investigation for optimal inversion and regularization parameters. The statement made on the outperformance of GN is completely made based on the two tests, and it requires further investigations on different benchmark models and real seismic data.

Since the modeling was performed using the fourth-order finite difference approximation on staggered-grid resulting in the 13-point stencil, it gives more fidelity to using less number of gridpoints (four gridpoints) per shortest wavelength. This minimum number of gridpoints can be further reduced using high-order approximations resulting in more stencil points with additional computational cost. It is also noted that the forward modeling contains only P-waves because of the acoustic approximation of the wave equation. This approximation could be partially valid for marine environments for controlled-source experiments but may fail on land data due to the presence of surface waves generated by the interference between both P- and S-waves. For more accurate imaging, elastic wave equation could be helpful. However, it requires massive computational resources due to the inversion of multi-parameters, and is highly non-linear, and each parameter have different sensitivity. In this study, we are minimizing the data residuals in a least-square sense. It is also possible to modify the misfit function to other norms such as the Wasserstein metric (optimal transport), a combination of different norms with varying weights, and to implement global optimization strategies such as Markov chain Monte Carlo, simulated annealing, and genetic algorithm.

We believe there are more innovations made in the field of seismic imaging. As a future aspect of this study, we intend to further investigate on regularization and stabilization of both methods for applying to benchmark models such as Marmoiusi, SEG/EAGE salt model, 2004 BP benchmark model, and then on real seismic data. We also intend to further investigate the performance of both GN and NLCG methods and investigate the possibility of creating a hybrid optimization strategy where we start the inversion with GN for a few iterations and change to NLCG for further iterations. We also intend to attempt on implementing elastic or visco-elastic approximation of the wave equation for improved imaging.

References

Abma, R. L., Ford, A., Rose-Innes, N., Mannaerts-Drew, H. and Kommedat, J. (2013). Continued development of simultaneous source acquisition for ocean bottom surveys. 75th Annual International Conference and Exhibition, EAGE, Extended Abstracts .

Anagaw, A. Y. (2014). Full waveform inversion using simultaneous encoded sources based on firstand second-order optimization methods. Ph.D. thesis.

Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P. and Virieux, J. (2013). Regularized seismic full waveform inversion with prior model information. Geophysics 78, doi:10.1190/geo2012-0104.1.

Bae, H. S., Shin, C., Min, D. and Calandra, H. (2010). 2-D waveform inversion in the laplace domain for acoustic-elastic coupled media. SEG Technical Program Expanded Abstracts 2010 doi:10.1190/1.3513937.

Bailey, D. H. and Swarztrauber, P. N. (1994). A fast method for the numerical evaluation of continuous fourier and laplace transforms. SIAM Journal on Scientific Computing 15: 1105–1110, doi:10.1137/0915067.

Bancroft, J. C. (2007). A practical understanding of pre- and poststack migrations : 4.1–4.118doi:10.1190/1.9781560801641.ch4.

Bell, J. B., Tikhonov, A. N. and Arsenin, V. Y. (1978). Solutions of ill-posed problems. Mathematics of Computation 32: 1320, doi:10.2307/2006360.

Ben-Hadj-Ali, H., Operto, S. and Virieux, J. (2008). Velocity model building by 3D frequency-domain, full-waveform inversion of wide-aperture seismic data. Geophysics 73: VE101–VE117, doi:10.1190/1.2957948.

Berenger, J.-P. (1994). A perfectly matched layer for the absorption of electromagnetic waves. Journal of Computational Physics 114: 185–200, doi:10.1006/jcph.1994.1159.

Borisov, D. and Singh, S. C. (2015). Three-dimensional elastic full waveform inversion in a marine environment using multicomponent ocean-bottom cables: a synthetic study. Geophysical Journal International 201: 1215–1234, doi:10.1093/gji/ggv048.

Bunks, C., Saleck, F. M., Zaleski, S. and Chavent, G. (1995). Multiscale seismic waveform inversion. Geophysics 60: 1457–1473, doi:10.1190/1.1443880.

Cantillo, J., Boelle, J.-L., Lafram, P. A. and Lecerf, D. (2010). Ocean bottom nodes (OBN) repeatability and 4d. In SEG Technical Program Expanded Abstracts 2010. Society of Exploration Geophysicists, 61–65.

Cao, Z. (2006). Analysis and application of the radon transform. doi:http://dx.doi.org/10.11575/PRISM/886.

Cerjan, C., Kosloff, D., Kosloff, R. and Reshef, M. (1985). A nonreflecting boundary condition for discrete acoustic and elastic wave equations. Geophysics 50: 705–708, doi:10.1190/1.1441945.

Dondurur, D. (2018). Acquisition and processing of marine seismic data. Elsevier doi:10.1016/c2016-0-01591-7.

Fletcher, R. (1986). Practical methods of optimization. Wiley, 2nd ed.

Fletcher, R. and Reeves, C. (1964). Function minimization by conjugate gradients. The Computer Journal 7: 149–154, doi:10.1093/comjnl/7.2.149.

Fomel, S. and Claerbout, J. F. (2003). Multidimensional recursive filter preconditioning in geophysical estimation problems. Geophysics 68: 577–588, doi:10.1190/1.1567228.

Freudenreich, Y. and Singh, S. (2000). Full waveform inversion for seismic data - frequency versus time domain. 62nd EAGE Conference amp; Exhibition doi:10.3997/2214-4609-pdb.28.c54.

George, A. and Liu, J. W.-H. (1981). Computer solution of large sparse positive definite systems. Prentice-Hall.

Herrmann, F. J., Li, X., Aravkin, A. Y. and Leeuwen, T. van (2011). A modified, sparsity-promoting, gauss-newton algorithm for seismic waveform inversion. SPIE Proceedings doi:10.1117/12.893861.

Hestenes, M. and Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. Journal of Research of the National Bureau of Standards 49: 409, doi:10.6028/jres.049.044.

Hustedt, B., Operto, S. and Virieux, J. (2004). Mixed-grid and staggered-grid finite-difference methods for frequency-domain acoustic wave modelling. Geophysical Journal International 157: 1269–1296, doi:10.1111/j.1365-246x.2004.02289.x.

Ikelle, L. T., Diet, J. P. and Tarantola, A. (1988). Linearized inversion of multioffset seismic reflection data in the ω-k domain: Depth-dependent reference medium. Geophysics 53: 50–64, doi:10.1190/1.1442399.

Kearey, P., Brooks, M. and Hill, I. (2002). An Introduction to Geophysical Exploration. Blackwell Science Ltd, 3rd ed.

Lailly, P. (1983). The seismic inverse problem as a sequence of before-stack migrations. Society for Industrial and Applied Mathematics, Conference on Inverse Scattering: Theory and Application : 206–220.

Mallick, S. and Frazer, L. N. (1987). Practical aspects of reflectivity modeling. Geophysics 52: 1355–1364, doi:10.1190/1.1442248.

Marfurt, K. J. (1984). Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations. Geophysics 49: 533–549, doi:10.1190/1.1441689.

Mora, P. (1987). Nonlinear two-dimensional elastic inversion of multioffset seismic data. Geophysics 52: 1211–1228, doi:10.1190/1.1442384.

Nemeth, T., Wu, C. and Schuster, G. T. (1999). Least-squares migration of incomplete reflection data. Geophysics 64: 208–221, doi:10.1190/1.1444517.

Nocedal, J. (1980). Updating quasi-newton matrices with limited storage. Mathematics of Computation 35: 773–782, doi:https://doi.org/10.2307/2006193.

Nocedal, J. and Wright, S. J. (2006). Numerical optimization. Springer, 2nd ed.

Oldham, R. (1906). The constitution of the earth. Quarterly Journal of the Geological Society of London : 456–457.

Petrov, P. V. and Newman, G. A. (2014). Three-dimensional inverse modelling of damped elastic wave propagation in the fourier domain. Geophysical Journal International 198: 1599–1617, doi:10.1093/gji/ggu222.

Plessix, R.-E. (2007). A helmholtz iterative solver for 3D seismic-imaging problems. Geophysics 72: SM185–SM194, doi:10.1190/1.2738849.

Plessix, R.- (2009). Three-dimensional frequency-domain full-waveform inversion with an iterative solver. Geophysics 74: WCC149–WCC157, doi:10.1190/1.3211198.

Polak, E. and Ribiere, G. (1969). Note sur la convergence de méthodes de directions conjuguées. Revue française d'informatique et de recherche opérationnelle. Série rouge 3: 35–43, doi:10.1051/m2an/196903r100351.

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, doi:10.1190/1.1444597.

Pratt, R. G. and Worthington, M. H. (1988). The application of diffraction tomography to cross-hole seismic data. Geophysics 53: 1284–1294, doi:10.1190/1.1442406.

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, doi:10.1111/j.1365-2478.1990.tb01846.x.

Ricker, N. (1953). The form and laws of propagation of seismic wavelets. Geophysics 18: 10–40, doi:10.1190/1.1437843.

Robein, E. (2010). Velocities, time-imaging and depth-imaging: Principles and methods. Earthdoc.

Sambridge, M. and Drijkoningen, G. (1992). Genetic algorithms in seismic waveform inversion. Geophysical Journal International 109: 323–342, doi:10.1111/j.1365-246x.1992.tb00100.x.

Sheriff, R. E. and Geldart, L. P. (1987). Exploration seismology. Cambridge University Press.

Shewchuk, J. R. (1994). An introduction to the conjugate gradient method without the agonizing pain. Carnegie-Mellon University. Department of Computer Science.

Shin, C., Jang, S. and Min, D.-J. (2001). Improved amplitude preservation for prestack depth migration by inverse scattering theory. Geophysical Prospecting 49: 592–606, doi:10.1046/j.1365-2478.2001.00279.x.

Shipp, R. M. and Singh, S. C. (2002). Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data. Geophysical Journal International 151: 325–344, doi:10.1046/j.1365-246x.2002.01645.x.

Sirgue, L. and Pratt, R. G. (2004). Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. Geophysics 69: 231–248, doi:10.1190/1.1649391.

Sirgue, L., T. Etgen, J. and Albertin, U. (2008). 3D frequency domain waveform inversion using time domain finite difference methods. 70th EAGE Conference and Exhibition incorporating SPE EUROPEC 2008 doi:10.3997/2214-4609.20147683.

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. Geophysics 49: 1259–1266, doi:10.1190/1.1441754.

Tarantola, A. (1987). Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation. Elsevier Science.

Thurber, C. and Ritsema, J. (2007). Theory and observations – seismic tomography and inverse methods. Treatise on Geophysics : 323–360doi:10.1016/b978-044452748-6.00009-2.

Vigh, D. and Starr, E. W. (2008). Comparisons for waveform inversion, time domain or frequency domain? SEG Technical Program Expanded Abstracts 2008 doi:10.1190/1.3059269.

Virieux, J. (1984). SH-wave propagation in heterogeneous media: Velocity-stress finite-difference method. Geophysics 49: 1933–1942, doi:10.1190/1.1441605.

Virieux, J., Asnaashari, A., Brossier, R., Métivier, L., Ribodetti, A. and Zhou, W. (2017). An introduction to full waveform inversion. Encyclopedia of Exploration Geophysics doi:10.1190/1.9781560803027.entry6.

Virieux, J. and Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. Geophysics 74: WCC1–WCC26, doi:10.1190/1.3238367.

Wang, H., Singh, S. C. and Calandra, H. (2014). Integrated inversion using combined wave-equation tomography and full waveform inversion. Geophysical Journal International 198: 430–446, doi:10.1093/gji/ggu138.

Wang, J. and Dong, L. (2021). Memory-efficient frequency-domain gauss–newton method for wave-equation first-arrival traveltime inversion. Near Surface Geophysics 19: 415–428, doi:10.1002/nsg.12161.

Warner, M., Ratcliffe, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N., Vinje, V., Štekl, I., Guasch, L. and Win, C. e. a. (2013). Anisotropic 3D full-waveform inversion. Geophysics 78: R59–R80, doi:10.1190/geo2012-0338.1.

Wu, R. and Toksöz, M. N. (1987). Diffraction tomography and multisource holography applied to seismic imaging. Geophysics 52: 11–25, doi:10.1190/1.1442237.

Wu, R.-S. and Aki, K. (1988). Multiple scattering and energy transfer of seismic waves—separation of scattering effect from intrinsic attenuation ii. application of the theory to hindu kush region. Scattering and Attenuations of Seismic Waves, Part I : 49–80doi:10.1007/978-3-0348-7722-0_4.

Yang, P., Gao, J. and Wang, B. (2015). A graphics processing unit implementation of time-domain full-waveform inversion. Geophysics 80, doi:10.1190/geo2014-0283.1.

Yilmaz, Öz. (2001). Seismic data analysis: Processing, inversion, and interpretation of seismic data. Society of exploration geophysicists.

A ‖ 13 Point stencils from fourth-order discretization of the wave equation

From the Eq. 2.5, the stencil points are as given below;

$$
\begin{aligned}
C_1 &= \frac{\omega^2}{\kappa_{i,j}} - \frac{1}{\Delta^2\xi_i}\left[\frac{9}{8}\frac{9}{8}\left(\frac{b_{i+\frac{1}{2},j}}{\xi_{i+\frac{1}{2}}} + \frac{b_{i-\frac{1}{2},j}}{\xi_{i-\frac{1}{2}}}\right) + \frac{1}{24}\frac{1}{24}\left(\frac{b_{i+\frac{3}{2},j}}{\xi_{i+\frac{3}{2}}} + \frac{b_{i-\frac{3}{2},j}}{\xi_{i-\frac{3}{2}}}\right)\right] \\
&\quad - \frac{1}{\Delta^2\xi_j}\left[\frac{9}{8}\frac{9}{8}\left(\frac{b_{i,j+\frac{1}{2}}}{\xi_{j+\frac{1}{2}}} + \frac{b_{i,j-\frac{1}{2}}}{\xi_{j-\frac{1}{2}}}\right) + \frac{1}{24}\frac{1}{24}\left(\frac{b_{i,j+\frac{3}{2}}}{\xi_{j+\frac{3}{2}}} + \frac{b_{i,j-\frac{3}{2}}}{\xi_{j-\frac{3}{2}}}\right)\right] \\
C_2 &= \frac{1}{\Delta^2\xi_i}\left(\frac{9}{8}\frac{9}{8}\frac{b_{i-\frac{1}{2},j}}{\xi_{i-\frac{1}{2}}} + \frac{9}{8}\frac{1}{24}\frac{b_{i+\frac{1}{2},j}}{\xi_{i+\frac{1}{2}}} + \frac{1}{24}\frac{9}{8}\frac{b_{i-\frac{3}{2},j}}{\xi_{i-\frac{3}{2}}}\right) \\
C_3 &= \frac{1}{\Delta^2\xi_i}\left(\frac{9}{8}\frac{9}{8}\frac{b_{i+\frac{1}{2},j}}{\xi_{i+\frac{1}{2}}} + \frac{9}{8}\frac{1}{24}\frac{b_{i-\frac{1}{2},j}}{\xi_{i-\frac{1}{2}}} + \frac{1}{24}\frac{9}{8}\frac{b_{i+\frac{3}{2},j}}{\xi_{i+\frac{3}{2}}}\right) \\
C_4 &= \frac{1}{\Delta^2\xi_j}\left(\frac{9}{8}\frac{9}{8}\frac{b_{i,j-\frac{1}{2}}}{\xi_{j-\frac{1}{2}}} + \frac{9}{8}\frac{1}{24}\frac{b_{i,j+\frac{1}{2}}}{\xi_{j+\frac{1}{2}}} + \frac{1}{24}\frac{9}{8}\frac{b_{i,j-\frac{3}{2}}}{\xi_{j-\frac{3}{2}}}\right) \\
C_5 &= \frac{1}{\Delta^2\xi_j}\left(\frac{9}{8}\frac{9}{8}\frac{b_{i,j+\frac{1}{2},j}}{\xi_{j+\frac{1}{2}}} + \frac{9}{8}\frac{1}{24}\frac{b_{i,j-\frac{1}{2}}}{\xi_{j-\frac{1}{2}}} + \frac{1}{24}\frac{9}{8}\frac{b_{i,j+\frac{3}{2}}}{\xi_{j+\frac{3}{2}}}\right) \\
C_6 &= -\frac{1}{\Delta^2\xi_i}\left(\frac{9}{8}\frac{1}{24}\frac{b_{i-\frac{1}{2},j}}{\xi_{i-\frac{1}{2}}} + \frac{1}{24}\frac{9}{8}\frac{b_{i-\frac{3}{2},j}}{\xi_{i-\frac{3}{2}}}\right) \\
C_7 &= -\frac{1}{\Delta^2\xi_i}\left(\frac{9}{8}\frac{1}{24}\frac{b_{i+\frac{1}{2},j}}{\xi_{i+\frac{1}{2}}} + \frac{1}{24}\frac{9}{8}\frac{b_{i+\frac{3}{2},j}}{\xi_{i+\frac{3}{2}}}\right) \\
C_8 &= -\frac{1}{\Delta^2\xi_i}\left(\frac{9}{8}\frac{1}{24}\frac{b_{i,j-\frac{1}{2}}}{\xi_{j-\frac{1}{2}}} + \frac{1}{24}\frac{9}{8}\frac{b_{i,j-\frac{3}{2}}}{\xi_{j-\frac{3}{2}}}\right) \\
C_9 &= -\frac{1}{\Delta^2\xi_j}\left(\frac{9}{8}\frac{1}{24}\frac{b_{i,j+\frac{1}{2}}}{\xi_{j+\frac{1}{2}}} + \frac{1}{24}\frac{9}{8}\frac{b_{i,j+\frac{3}{2}}}{\xi_{j+\frac{3}{2}}}\right) \\
C_{10} &= \frac{1}{\Delta^2\xi_i}\left(\frac{1}{24}\frac{1}{24}\frac{b_{i-\frac{3}{2},j}}{\xi_{i-\frac{3}{2}}}\right) \\
C_{11} &= \frac{1}{\Delta^2\xi_i}\left(\frac{1}{24}\frac{1}{24}\frac{b_{i+\frac{3}{2},j}}{\xi_{i+\frac{3}{2}}}\right) \\
C_{12} &= \frac{1}{\Delta^2\xi_j}\left(\frac{1}{24}\frac{1}{24}\frac{b_{i,j-\frac{3}{2}}}{\xi_{j-\frac{3}{2}}}\right)
\end{aligned}
\tag{A.1}
$$

$$C_{13} = \frac{1}{\Delta^2 \xi_j} \left(\frac{1}{24} \frac{1}{24} \frac{b_{i,j+\frac{3}{2}}}{\xi_{j+\frac{3}{2}}} \right)$$

B ‖ Inversion algorithm

The algorithm for generating the synthetic data and the Jacobian matrix is given below. F and s are the discreet frequency values and corresponding source amplitude, respectively, T is the transformation matrix for interpolating the data from gridpoints to receiver locations.

Algorithm 1 Generate synthetic data and Jacobian matrix

Require: $F,\ s,\ T,\ C_1,\ \ldots,\ C_{13}$
  for $i \leftarrow 1\ \ to\ \ length(F)$ do
    $M \leftarrow diag([C_{12}, C_{10}, C_8, C_6, C_4, C_2, C_1, C_3, C_5, C_7, C_9, C_{11}, C_{13}],$
      $[-3nx, -2nx, -nx, -3, -2, -1, 0, 1, 2, 3, nx, 2nx, 3nx])$
    $p \leftarrow M^{-1}s$
    $P^{mod} \leftarrow append(Tp)$
    $G1 \leftarrow -\frac{\partial M}{\partial m}p$
    $G \leftarrow append(G1)$
    $R1 \leftarrow (M^{-1})^T T^T$
    $R \leftarrow append(R1)$
  end for
  $J \leftarrow G^T R$
  return $P^{mod}\ and\ J$

The algorithm for inversion using the Gauss-Newton method is given below. [1]

Algorithm 2 Inversion using GN

Require: $P^{obs},\ m^{initial},\ m^{ref}$
  $m \leftarrow m^{initial}$
  $E^{rms} \leftarrow 100$
  $W_d \leftarrow diag(\frac{100}{|P^{obs}|})$
  $W_m \leftarrow I_d$
  while $iter \leq max\_iter\quad or\quad E^{rms} \geq tolerence$ do
    $P^{mod},\ J \leftarrow$ call Algorithm 1
    $\Delta r \leftarrow P^{obs} - P^{mod}$
    $\lambda \leftarrow \frac{||J^T W_d^T W_d J||_\infty}{n^{iter}}$
    $m \leftarrow \left(J^T W_d^T W_d J + \lambda W_m^T W_m\right)^{-1} \left(J^T W_d^T W_d(\Delta r + Jm) + \lambda W_m^T W_m m^{ref}\right)$
    $E^{rms} \leftarrow \sqrt{\frac{\Delta r^T \Delta r}{length(dr)}}$
    $iter \leftarrow iter + 1$
  end while
  $m^{inv} \leftarrow m$
  return m^{inv}

[1] value of n for λ can be decided by trial and error.

The algorithm to estimate the scaled gradient of the objective function is given below.

Algorithm 3 Generate gradient of objective function

Require: $J,\ \Delta r,\ W_m,\ W_d,\ m,\ m^{ref},\ \lambda$
  $grad \leftarrow \frac{-2}{N_d} J^T W_d^T W_d \Delta r + \frac{2\lambda}{N_p}(m - m^{ref})(W_m^T W_m)^T$
  return $grad$

The algorithm for inversion using the Non-Linear Conjugate Gradient method is given below. [2]

Algorithm 4 Inversion using NLCG

Require: $P^{obs},\ m^{initial},\ m^{ref},\ \lambda$
  $m \leftarrow m^{initial}$
  $E^{rms} \leftarrow 100$
  $W_d \leftarrow diag(\frac{100}{|P^{obs}|})$
  $W_m \leftarrow I_d$
  while $iter \leq max\_iter \quad or \quad E^{rms} \geq tolerence$ do
    $P^{mod},\ J \leftarrow$ call Algorithm 1
    $\Delta r \leftarrow P^{obs} - P^{mod}$
    $grad \leftarrow$ call Algorithm 3
    $g \leftarrow -grad$
    if $iter == 1$ then
      $\beta \leftarrow 0$
    else
      $\beta \leftarrow \frac{g^T(g - g_0)}{g_0^T g_0}$
    end if
    $h \leftarrow g + \beta h_0$
    $\alpha \leftarrow \frac{(Jh)^T\ \Delta r}{(Jh)^T (Jh)}$
    $m \leftarrow m + \alpha h$
    $g_0, h_0 = g, h$
    $E^{rms} \leftarrow \sqrt{\frac{\Delta r^T \Delta r}{length(dr)}}$
    $iter \leftarrow iter + 1$
  end while
  $m^{inv} \leftarrow m$
  return m^{inv}

[2] All codes used in this study are available in python at https://github.com/mssujith/