SEISMIC IMAGING BY NONLINEAR INVERSION

by

Werter Oliveira Silva

© Copyright by Werter Oliveira Silva, 2024

All Rights Reserved

A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Master of Science (Geophysics).

Golden, Colorado

Date _____

Signed: _____

Werter Oliveira Silva

Signed: _____

Dr. Paul Sava
Thesis Advisor

Golden, Colorado

Date _____

Signed: _____

Dr. Paul Sava
Professor and Head
Department of Geophysics

ABSTRACT

Imaging aims to create representations of internal object structures through indirect external physical measurements. In seismic exploration, for instance, seismic reflections on the Earth's surface are mapped into discontinuities in physical properties, revealing geological structures. Various seismic imaging techniques exist, differing in their approach to wave propagation (acoustic or elastic; isotropic or anisotropic), wave equation type (one-way or all-way), application domain (post-stack or pre-stack), numerical implementation (frequency or time domain; integral or differential forms), and other factors.

Migrations usually assume a linear relationship between data and image based on the Born approximation, and the image consists of a scalar parameter that describes the spatial distribution of subsurface reflectors. Since seismic data includes not only primary reflections but also multiples that do not satisfy the Born approximation, imaging is normally preceded by multiple attenuation to meet the linear assumption and avoid creating fake reflectors and crosstalk noise. However, multiples provide additional illumination and resolution because they can sample subsurface image points at angles different from those of the primary waves. Therefore, multiple attenuation ignores additional information that could be used to improve the image.

In this thesis, I introduce an acoustic nonlinear inversion imaging method, based on a parameterization of the wave equation that preserves the nonlinearity between data and image, defined as a vector instead of a scalar function. This parameterization separates propagation and dynamic effects. Wave propagation is ruled by a background velocity model, lacking sharp contrasts, while the dynamics of reflections is controlled by the image vector parameter I seek to invert. The vectorial nature of the image reflects the directional dependence of the reflectivity and its nonlinear dependence to the data enables multiple-scattering modeling to fit unprocessed data, containing multiples and ghots in addition to primaries.

TABLE OF CONTENTS

ABSTRACT . . . iii

LIST OF FIGURES . . . v

LIST OF SYMBOLS . . . vi

LIST OF ABBREVIATIONS . . . vii

ACKNOWLEDGMENTS . . . viii

DEDICATION . . . ix

CHAPTER 1 INTRODUCTION . . . 1

CHAPTER 2 SEISMIC IMAGING BY NONLINEAR INVERSION . . . 3

2.1 Abstract . . . 3

2.2 Introduction . . . 3

2.3 Theory . . . 5

2.3.1 Wave Equation Parameterization . . . 5

2.3.2 Inversion Setup and Gradient Derivation . . . 7

2.4 Numerical Examples . . . 10

2.4.1 Sensitivity Kernel . . . 10

2.4.2 Two Horizontal Reflectors . . . 14

2.4.3 Marmousi Model . . . 18

2.4.4 Misfit Convergence: Adjoint vs Time-reversal . . . 18

2.5 Discussions . . . 23

2.6 Conclusions . . . 24

2.7 Acknowledgments . . . 25

2.8 Appendix A: Adjoint Wave Equation Operator . . . 25

CHAPTER 3 CONCLUSIONS . . . 28

REFERENCES . . . 31

LIST OF FIGURES

Figure 2.1 (a) Sensitivity kernel for one source (red star) and one receiver (green triangle) for: (a) LSRTM, (b) component x and (c) component z of the nonlinear imaging. The blue and white arrows indicate phase changes in the migration and tomography kernels, respectively. 11

Figure 2.2 The top row depicts snapshots of forward u and adjoint wavefield λ at time t = 0.373 s. The middle row shows the gradient components of u in the horizontal direction u_x and in the vertical direction u_z. The bottom row shows image components formed at that particular time by multiplying u_x by λ and u_z by λ. The white arrows highlight the formation of the left component of the tomography kernel in Figure 2.1(b) and in Figure 2.1(c), having the same phase. The blue arrows point to the formation of the migration kernel in these figures. 12

Figure 2.3 The top row depicts snapshots of forward u and adjoint wavefield λ at time $t = 0.932\,s$. The middle row shows the gradient components of u in the x direction u_x and in the z direction u_z. The bottom row shows image components formed at that particular time by multiplying u_x by λ and u_z by λ. The white arrows highlight the formation of the right component of the tomography kernel in Figure 2.1(b) and in Figure 2.1(c), having opposite phase. The blue arrows point to the formation of the migration kernel in these figures. 13

Figure 2.4 (a) Velocity model, (b) density model, (c) horizontal component and (d) vertical component of reflectivity calculated as the normalized gradient of impedance: $\frac{\nabla Z}{Z}$. 15

Figure 2.5 (a) Simulated shot gather at shot position x = 600 m and z = 10 m, (b) first iteration of LSRTM, which corresponds to migration, (c) and (d) are the components x and z of the first iteration of nonlinear imaging, respectively. 16

Figure 2.6 (a) LSRTM image, (b) x and (c) z components of the nonlinear imaging, after 35 iterations. 17

Figure 2.7 (a) Marmousi velocity model, (b) density model and (c) smooth version of (a) used in migrations (first iteration of inversion) and inversions. 19

Figure 2.8 (a) Reflectivity component x and (b) reflectivity component z, obtained through $\frac{\nabla Z}{Z}$ using the velocity and density models in Figure 2.7(a) and Figure 2.7(b), respectively. . . 20

Figure 2.9 (a) First iteration of LSRTM, which is equivalent to an RTM (Reverse Time Migration), (b) x and (c) z components of the image vector obtained from the first iteration of the nonlinear imaging method. 21

Figure 2.10 (a) LSRTM result after 35 iterations, (b) x and (c) z components of the image vector after 35 iterations of the nonlinear imaging method. 22

Figure 2.11 Comparison of the data misfit convergence for the nonlinear imaging method when the correct adjoint or the time-reversal only is used to compute the adjoint wavefield λ as part of the misfit gradient. The misfit level of 5% is indicated by the dashed red line. . . . 23

LIST OF SYMBOLS

Acoustic impedance . . . . . Z

Acoustic seismic wavefield . . . . . u

Adjoint acoustic wavefield . . . . . λ

Adjoint symbol for linear operators . . . . . †

Divergence operator $\nabla\cdot$ . . . . . $\mathbf{D}$

Function for implicit wave equation definition . . . . . $\mathcal{F}$

Function that produce a diagonal matrix operator from the argument . . . . . $diag(\cdot)$

Gradient operator ∇ . . . . . $\mathbf{G}$

Inner product in the function space $\mathcal{U}$ . . . . . $\langle\cdot,\cdot\rangle_{\mathcal{U}}$

LSRTM operator . . . . . $\mathbf{B}$

Laplacian operator ∇^2 . . . . . $\mathbf{L}$

Least squares norm . . . . . $\|\cdot\|$

Physical space variable . . . . . $\mathbf{x}$

Sampling operator . . . . . $\mathbf{R}$

Wave equation operator . . . . . $\mathbf{A}$

LIST OF ABBREVIATIONS

Full Waveform Inversion . . . . . FWI

Least Squares Migration . . . . . LSM

Least Squares Reverse Time Migration . . . . . LSRTM

Limited-memory Broyden–Fletcher–Goldfarb–Shanno . . . . . L-BFGS

ACKNOWLEDGMENTS

I am truly grateful to Petrobras for granting me the opportunity to pursue a master's degree. The support provided by Petrobras is a significant contribution to my professional and personal growth.

I would like to thank my advisor, Dr. Paul Sava, for the invaluable discussions, guidance and encouragement. His constructive questions and comments were instrumental in shaping my technical skills and in refining my writing and presentation abilities. Moreover, his Wavefield Seismic Imaging course gave me the fundamental knowledge to dive into this fascinating topic.

I would like to thank Dr. Yaoguo Li and Dr. Ge Jin for agreeing to be members of my thesis committee. I consider myself lucky to have been their students and have learned so much about inversion theory and geophysical computing that aided in forming the background to conduct my research.

I appreciate the research group meetings led by Paul, along with Khalid Almuteri, Iga Pawelec, Nicholas Dorogy and Tiago Cabral. Not only did I learn from their presentations and discussions, but they also contributed to a pleasant atmosphere. My gratitude extends to the other students and professors in the Department of Geophysics with whom I interacted during classes, seminars, and leisure times. Special thanks to Noelle Vance and Lynn Lundebrek for their administrative support.

My gratitude to my friends and office mates, Ashish Kumar, Derrick Chambers, and Job Urang, with whom I have shared good conversations about research and also enjoyable moments of laughter. A special thanks goes to my Petrobras colleague Alan Souza for his incentive and valuable suggestions.

This thesis would not have been possible without the support of my family. My wife Iracelle, our children, Ian and Isis, along with my parents, have been my pillars of strength and inspiration. Their constant encouragement has kept me going when it seemed impossible. This achievement is as much theirs as it is mine, and I am profoundly grateful for their presence in my life.

To my beloved children, Ian and Isis.

CHAPTER 1

INTRODUCTION

Imaging creates representations of object interiors from indirect physical measurements at the exterior of the object and has applications across diverse fields. For example, medical imaging creates visual representations of the interior of the human body for the diagnosis of medical conditions (Hussain *et al.*, 2022). Non Destructive Testing uses imaging to analyze material integrity without causing damage, making it an indispensable tool in many industries, like advanced manufacturing, aerospace, and infrastructure (Dwivedi *et al.*, 2018). The focus of the thesis is on seismic imaging, a powerful technique for studying subsurface geological structures by mapping seismograms recorded on the Earth's surface to its interior (Scales, 1995).

The Earth subsurface can be represented by the spatial distribution of its physical properties, such as velocity and impedance, and also by their contrasts. Seismic data contain two important sources of information, traveltime and amplitude. As demonstrated by Jannane *et al.* (1989), traveltimes are predominantly influenced by the low wavenumber component of velocity governing wave propagation, while amplitudes are sensitive to the high wavenumber components of velocity that correspond to contrasts, giving rise to observed seismic reflections. Stolt & Weglein (2012) state that the physics of wave propagation and reflection partially decouple because they inhabit regions of different wavelengths, which is known as the information gap that can be recovered from seismic data (Claerbout, 1985; Tarantola, 1986). As a result, processing seismic reflections has two main interrelated goals, namely determining the velocity model and generating a seismic image. The concept of a seismic image is not precisely defined, but is related to a measure of reflection strength (Stolt & Weglein, 2012), which provides a high-frequency view of the subsurface delineating geological discontinuities. These discontinuities, or reflectors, mark geological facies and time changes (Sheriff & Geldart, 1995), making the seismic image a key element in interpreting subsurface geological structures.

Over the past decades, many seismic imaging algorithms have been developed, influenced by challenges posed by geological complexity and computational costs. For instance, the use of ray theory (Cervenỳ, 2001) to approximate waves accelerates numerical computations, but often fails to account for complex phenomena such as wavefront triplications (Biondi, 2006), resulting in inaccurate images. To overcome this limitation, wave-equation methods have been developed, exploiting the one-way approximation of the full wave equation. In this framework, the wave equation is split into up- and down-going operators (Claerbout, 1985), and the image is formed by crosscorrelating source and receiver wavefields according to the imaging principle, stating that reflectors exist where these wavefields coincide in time (Claerbout, 1971). However, these methods are limited by the maximum reflector dip they can image (Mulder & Plessix, 2004). Reverse time migration

(Baysal *et al.*, 1983) is a two-way alternative technique to angle limitation and became a current practice in industry with the advancement of computing power.

Imaging can be further classified into adjoint and inversion-based methods (Claerbout, 1992). The adjoint, also known as migration, has the purpose to undo the modeling representing the seismic experiment that generates the observed data (Berkhout, 2012). The accuracy of this approximation is degraded by spatial aliasing, limited aperture, and nonuniform illumination due to complex overburden (Wang *et al.*, 2017). To mitigate these problems and enhance image resolution, Least Squares Migration methods have been developed (Dai & Schuster, 2013; Nemeth *et al.*, 1999; Valenciano *et al.*, 2009). All adjoint and inverse techniques rely on the Born approximation (Bleistein *et al.*, 2001), which considers a linear relationship between the image and the data, assumed to contain only single-scattered (primary) events. In inversion-based approaches, the image is parameterized in terms of a perturbation representing rapid changes of a physical property, such as velocity (Dai & Schuster, 2013), and the goal is to determine a perturbation model that explains the observed data. All the acoustic wave-equation imaging methods produce an image that is a scalar as a function of physical space, and assume that the input data are free of multiples and ghosts. This assumption implies that the data have to undergo seismic preprocessing for multiple attenuation and ghost removal before they are ready for imaging (Yilmaz, 2001). However, multiples provide additional illumination and can contribute to imaging, instead of been considered as noise (Berkhout, 2014).

In Chapter 2, titled "Seismic imaging by nonlinear inversion", I propose a two-way acoustic imaging method that can handle both single and multiple scattered events in the data. Instead of using the conventional approach of linearizing the acoustic wave equation to restrict modeling to primaries only, to fit seismic preprocessed data free of multiples and ghosts, I parameterize the forward acoustic wave equation by a quantity that explains multiple scattering in order to fit unprocessed data resulting from single and multiply scattered waves. This quantity is an image vector that is nonlinearly related to the data and better represents the contrast of physical properties according to their directionality than scalar images.

CHAPTER 2

SEISMIC IMAGING BY NONLINEAR INVERSION

Paper submitted to *Geophysics*.

Werter Silva[1,2,3,4], Paul Sava[4]

2.1 Abstract

We introduce a nonlinear imaging method to address crosstalk artifacts caused by seismic multiples and to compensate for uneven illumination and ghosts. Our method provides enhanced resolution and represents subsurface discontinuities by a vector quantity highlighting the structural orientation and wavefield propagation direction. We emphasize the critical role of the adjoint wave equation in computing accurate gradients of the imaging objective function and its ability to accelerate the convergence of numeric optimization schemes. We demonstrate our method on synthetic examples and illustrate the key distinction with conventional methods generating scalar images from primaries only. We also demonstrate the faster convergence of the data misfit when using the correct adjoint wave equation relative to the time reversal of the forward modeling operator. Additionally, we show that the sensitivity kernels of our nonlinear method are effective in inversion without contaminating the image with low-wavenumber artifacts.

2.2 Introduction

Seismic imaging techniques have traditionally been based on the single-scatter assumption or first-order Born approximation (Bleistein *et al.*, 2001; Sava & Hill, 2009), which assumes that the data contain only primary events (Claerbout, 1985; Yilmaz, 2001). Consequently, seismic data must be pre-processed before they can be used with conventional seismic imaging algorithms. Preprocessing typically involves removing surface and internal multiple reflections (Dutta *et al.*, 2019; Verschuur *et al.*, 1992; Weglein *et al.*, 1997), as well as ghosts (Soubaras, 2010). However, multiples provide additional illumination and resolution since they sample the same subsurface image point from different angles (Lu *et al.*, 2015), and therefore could aid imaging, and thus should not be treated as noise (Berkhout, 2014).

The idea of incorporating multiple reflections into seismic imaging is not new. Berkhout & Verschuur (1994) proposed migrating multiples by separating them from primary sources and treating them as secondary sources at receiver positions. Similar approaches that depend on primary and multiple separation can be

[1]Primary researcher and author.
[2]Author for correspondence. Email: wertersilva@mines.edu.
[3]Petrobras, Rio de Janeiro, Brazil.
[4]Center for Wave Phenomena, Colorado School of Mines, Golden, Colorado, United States of America.

found in Whitmore *et al.* (2010) and Lu *et al.* (2015). Separation of primaries and multiples, however, relies on the ability to distinguish them, which in many cases is a difficult task. Berkhout & Verschuur (2011); Davydenko & Verschuur (2017); Lu *et al.* (2018a) propose inversion-based full-wavefield migration utilizing one-way wave equation propagators. These approaches can handle multiples and address uneven illumination that is observed in conventional migration and also increase resolution; however, they inherit the one-way limitations in imaging steep dips due to the limited angle range involved in the approximation of the full wavefield (Mulder & Plessix, 2004).

Two-way modeling and migration have become standard practice for addressing complex geological settings. Reverse time migration (RTM) is a two-way imaging method that is not limited by steep dips (Baysal *et al.*, 1983). Least squares reverse time migration (LSRTM) provides additional improvement by compensating for the uneven subsurface illumination caused by the acquisition footprint and geologic structure, while enhancing the image resolution (Dai & Schuster, 2013). Nevertheless, these techniques presume a linear relationship between the image and the data, disregarding multiply scattered waves.

To image primary and multiples within a single inversion framework, the forward modeling engine should reflect the nonlinear dependence between image and data, or use cascaded Born modeling to account for higher scattering orders at additional cost. Davydenko & Verschuur (2021) reformulate the acoustic wave equation in terms of a system of first-order equations that relate particle velocity and pressure and introduce a scattering term responsible for generating up- and down-going wavefields. Wu & Alkhalifah (2017) formulate the scattering term as a perturbation of the slowness squared, similar to LSRTM. While in conventional LSRTM it is assumed that a first-order perturbation in slowness squared implies a first-order perturbation in the wavefield, Wu & Alkhalifah (2017) incorporate higher-order perturbation terms, preserving nonlinearity and enabling multiple scattering. In both formulations (Davydenko & Verschuur, 2021; Wu & Alkhalifah, 2017), the perturbation that represents the seismic image is a scalar function.

Seismic imaging is associated with mapping physical property contrasts of the subsurface, which could be described by spatial derivatives that have directionality. In a horizontally layered medium, for example, the horizontal contrasts are zero, while the strongest contrasts are in the vertical direction. However, when a model exhibits steep flanks, the vertical contrast of these features becomes relatively small compared to the horizontal. As a result, a vector image provides more information about the contrasts in the subsurface compared to a scalar counterpart.

Using the variable density acoustic wave equation, two modeling engines based on vector reflectivity have recently been proposed. Whitmore *et al.* (2020) replace the density term in the acoustic wave equation by its definition in terms of impedance, which after working out the vector-calculus identities leads to a term containing the gradient of impedance, named vector reflectivity. McLeman *et al.* (2021) use the same calculus

identities, but applied directly to the density term, leading to a term containing the density gradient, which is augmented by the velocity gradient to represent an approximation to the normal vector reflectivity. We show that assuming a migration velocity without sharp contrasts, the first approach simplifies to the second. These two approaches have been used in the context of FWI imaging (McLeman *et al.*, 2021; Yang *et al.*, 2021), where the FWI framework is used to invert for velocity and image.

In this paper, we use the variable density acoustic wave equation inspired by these two vector formulations to design a nonlinear imaging inversion problem, where the density gradient controls the multiple scattering, while the migration velocity is assumed known and responsible for the wave propagation. We do not recover density contrasts, but use a wave equation able to explain primary and multiple events observed in the data, either coming from density or velocity contrasts, or their combination —a term we refer to as image vector. We set up an inverse problem to recover this vector term, derive the gradient expression of the objective function, and also the adjoint of forward modeling (Claerbout, 2014). We compare our approach to that of Wu & Lu (2023) and show that we obtain a different adjoint operator that correctly satisfies the dot product. We demonstrate our method on synthetic examples and compare the convergence of the data misfit when using the incorrect time reversal of the forward modeling operator instead of the correct adjoint modeling operator.

2.3 Theory

In this section, we formulate a nonlinear inverse problem aimed at determining the image vector parameter. We investigate two distinct parameterizations of the variable-density acoustic wave equation in relation to this parameter. One formulation includes a gradient term concerning velocity, while the other does not. In both instances, the wavefield exhibits a nonlinear relation with the image, allowing the forward simulation to predict not only primary waves but also multiple reflections of unlimited orders, and ghosts, even in the absence of sharp velocity changes, as long as there are variations in the vector image. This characteristic eliminates the need for preprocessing observed data for multiple and ghost attenuation; instead, we use them in the imaging process. We derive two different acoustic wave equation parameterizations and show that they are similar under the assumption of a smooth background velocity model. We also pose the nonlinear imaging inverse problem, derive the objective function gradient and associated adjoint wave equation.

2.3.1 Wave Equation Parameterization

Beginning with the variable-density acoustic wave equation (Kosloff & Baysal, 1982),

$$\frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} - \rho\nabla\cdot\left(\frac{1}{\rho}\nabla u\right) = s, \tag{2.1}$$

where $u(\mathbf{x},t)$ is the pressure depending on position $\mathbf{x}$ and time t, $v(\mathbf{x})$ and $\rho(\mathbf{x})$ are the spatially-variable velocity and density, and $s(\mathbf{x}_s,t)$ is the source term at source position $\mathbf{x}_s$, we can derive an equivalent form of the wave equation in terms of impedance $Z(\mathbf{x})$ from the definition $Z=\rho v$ (Whitmore *et al.*, 2020):

$$\frac{1}{v^2}\frac{\partial^2 u}{\partial t^2}-\frac{Z}{v}\nabla\cdot\left(\frac{v}{Z}\nabla u\right)=s. \tag{2.2}$$

Applying the vector calculus product rule for the second term on the left-hand side of equation 2.2, we obtain

$$\frac{1}{v^2}\frac{\partial^2 u}{\partial t^2}-\frac{\nabla v}{v}\cdot\nabla u+\mathbf{m}\cdot\nabla u-\nabla^2 u=s. \tag{2.3}$$

The third term in equation 2.3 involves the vectorial quantity $\mathbf{m}=\frac{\nabla Z}{Z}$, which represents a relative impedance contrast. It is equivalent to the vector reflectivity described in Whitmore *et al.* (2020), except for one-half multiplicative factor applied to the relative impedance gradient $\frac{\nabla Z}{Z}$, that is canceled by the product of the factor 2 by $\mathbf{m}$. The wavefield u is linked to the model $\mathbf{m}$ in a nonlinear way, which can be inferred from the fact that $\mathbf{m}$ multiplies ∇u, while u is dependent on $\mathbf{m}$. As a consequence, equation 2.3 can be used to simulate not only primary events, but also refractions, internal and surface-related multiples even when the velocity is smooth. This formulation contrasts with the Born approximation used in LSRTM, where the wavefield is linearly related to the model parameter that represents the image and interacts with it only once to generate primary events.

In the context of FWI, Yang *et al.* (2020) use equation 2.3 to generate the scattered wavefield necessary to create the tomographic term (Schuster, 2017) without having to resort to the first-order Born approximation (Yao *et al.*, 2014). This sensitivity kernel is responsible for providing low-wavenumber velocity updates along the reflection wavepath. However, because the migration sensitivity kernel is an order of magnitude higher than the tomography kernel, updating the brackground velocity often involves a kernel separation (Yao *et al.*, 2020). As we focus on inversion for imaging, we do not perform kernel separation because the high-wavenumber migration kernel dominates the inversion.

A simpler and more cost-effective formulation of the wave equation containing fewer terms can be derived from equation 2.1 by expanding the term $\nabla\cdot\left(\frac{1}{\rho}\nabla u\right)$ to obtain

$$\frac{1}{v^2}\frac{\partial^2 u}{\partial t^2}+\frac{\nabla\rho}{\rho}\cdot\nabla u-\nabla^2 u-s=0, \tag{2.4}$$

Density contrasts also generate reflections, but McLeman *et al.* (2021) evoke the approximation of the normal incidence reflection (Shuey, 1985)

$$R(\theta=0)\approx\frac{\Delta v}{v}+\frac{\Delta\rho}{\rho} \tag{2.5}$$

and propose augmenting the equation 2.4 such that $\mathbf{m}$ represents the intercept-reflectivity in the equation

$$\frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} + \mathbf{m}\cdot\nabla u - \nabla^2 u - s = 0. \tag{2.6}$$

Instead of considering augmentation, we recognize that Equation 2.3 reduces to equation 2.4 when the velocity does not have abrupt changes and the term $\frac{\nabla v}{v}$ can be ignored. This is a reasonable simplification, since imaging is mainly driven by the low-wavenumber components of the background velocity that affect the kinematics of wave propagation (Wang *et al.*, 2019). As a smooth velocity does not generate reflections, necessary to fit the data in the inversion problem, the parameter $\mathbf{m}$ assumes the responsibility of explaining the reflections. We refer to $\mathbf{m}$ as the image vector, without trying to relate to a specific medium property contrast, which in the acoustic approximation can be density contrasts, velocity contrasts, or its combination (impedance contrasts). The reasoning is similar to the LSRTM image definition, based on the separation of scales, where the image is considered a high-wavenumber perturbation of the velocity that has the role of scattering the wavefield and producing reflections, while the background velocity is responsible for propagation effects (Dai & Schuster, 2013). In the nonlinear imaging method, the background velocity is also assigned the role of propagating the wavefield, while the parameter $\mathbf{m}$ in equation 2.6 has the role of generating reflections. Because of the nonlinear relation between the wavefield u and the parameter $\mathbf{m}$, equation 2.6 can model multiples and wave ghosts in addition to primaries. As a consequence, multiple reflections and wave ghosts need not be removed in the nonlinear imaging formulation as is required by LSRTM and other algorithms to satisfy the assumed linear dependence between data and image.

2.3.2 Inversion Setup and Gradient Derivation

To formulate seismic imaging as an inverse problem, we need two components. First, we need to formulate an objective function using a mapping that links the model (i.e., the image) to the data. In this case, we adopt the least squares mismatch between the observed and predicted data as the optimization criteria. The connection between the model and the data is provided by the wave equation 2.4. Second, to minimize the objective function using a local optimization method (Nocedal & Wright, 1999), we need to formulate an efficient objective function gradient, using the adjoint state method (Plessix, 2006; Talagrand & Courtier, 1987). In the following, we formulate the inverse problem and derive the gradient using the adjoint state method.

Mathematically, the nonlinear seismic imaging inverse problem can be stated as

$$\begin{aligned} \text{argmin}_{\mathbf{m}} \quad & J(u(\mathbf{m})) = \int_0^T \frac{1}{2}\|\mathbf{R}u(\mathbf{m}) - d\|_{\mathbf{x}}^2 \, dt \\ \text{subject to} \quad & \mathcal{F}(u,\mathbf{m}) = \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} + \mathbf{m}\cdot\nabla u - \nabla^2 u - s = 0, \end{aligned} \tag{2.7}$$

where J is a scalar objective function measuring the square misfit between the predicted $\mathbf{R}u$ and observed data d, T is the maximum recording time, $\mathbf{R}$ is a restriction operator that samples the wavefield u at the receiver positions, $\|\cdot\|_{\mathbf{x}}$ is the ℓ^2 norm over the physical space described by variable $\mathbf{x}$, with time integration made explicit for convenience. We assume an implicit summation over shots and receivers. The wavefield u is defined by wave equation 2.4, given in an implicit form described by function $\mathcal{F}$.

Using the chain rule, the derivative of J with respect to the model parameter $\mathbf{m}$, often referred to as gradient, is

$$\frac{dJ}{d\mathbf{m}} = \int_0^T \left(\frac{\partial J}{\partial u} \frac{\partial u}{\partial \mathbf{m}} \right) dt. \tag{2.8}$$

In equation 2.8, the term $\frac{\partial J}{\partial u}$ corresponds to the derivative of the squared waveform misfit with respect to u, which denotes the data difference. The Fréchet derivative $\frac{\partial u}{\partial \mathbf{m}}$ involves as many wave equation simulations as there are grid points in the discretized model (Liberzon, 2011) which is computationally expensive. However, the gradient of the objective function $\frac{dJ}{d\mathbf{m}}$ can be estimated efficiently using the adjoint state method, which only requires two wave simulations: one running forward in time and the other running backward in time (Plessix, 2006).

To avoid directly computing the costly term $\frac{\partial u}{\partial \mathbf{m}}$, we can use the wave equation 2.7 defined implicitly by $\mathcal{F} = 0$. Since this equation is zero everywhere by definition, its total derivative $\frac{d\mathcal{F}}{d\mathbf{m}}$ is also zero (Bradley, 2019). Therefore, we have

$$\frac{d\mathcal{F}}{d\mathbf{m}} = \frac{\partial \mathcal{F}}{\partial u} \frac{\partial u}{\partial \mathbf{m}} + \frac{\partial \mathcal{F}}{\partial \mathbf{m}} = 0, \tag{2.9}$$

from which we solve for $\frac{\partial u}{\partial \mathbf{m}}$ to obtain

$$\frac{\partial u}{\partial \mathbf{m}} = - \left(\frac{\partial \mathcal{F}}{\partial u} \right)^{-1} \frac{\partial \mathcal{F}}{\partial \mathbf{m}} \,. \tag{2.10}$$

The derivative $\frac{\partial \mathcal{F}}{\partial u}$ gives the wave equation operator

$$\mathbf{A} = \frac{1}{v^2} \frac{\partial^2}{\partial t^2} + \mathbf{m} \cdot \nabla - \nabla^2 \tag{2.11}$$

and the derivative of the wave equation with respect to the model parameter $\frac{\partial \mathcal{F}}{\partial \mathbf{m}}$ is

$$\frac{\partial \mathcal{F}}{\partial \mathbf{m}} = \nabla u. \tag{2.12}$$

Replacing equation 2.10 in equation 2.8, we get

$$\frac{dJ}{d\mathbf{m}} = \int_0^T \frac{\partial J}{\partial u} \left[- \left(\frac{\partial \mathcal{F}}{\partial u} \right)^{-1} \frac{\partial \mathcal{F}}{\partial \mathbf{m}} \right] dt \,. \tag{2.13}$$

Equation 2.13 does not make an explicit reference to the Fréchet derivative $\frac{\partial u}{\partial \mathbf{m}}$, yet it appears to have resulted in a more complex form that involves the inverse of the wave equation operator $\frac{\partial \mathcal{F}}{\partial u}$. However,

instead of first solving the term in square brackets in equation 2.13, we shift the brackets to a term that is easier to solve and label it $\lambda^\dagger$, where

$$\frac{dJ}{d\mathbf{m}} = \int_0^T \underbrace{\left[-\frac{\partial J}{\partial u} \left(\frac{\partial \mathcal{F}}{\partial u} \right)^{-1} \right]}_{\lambda^\dagger} \frac{\partial \mathcal{F}}{\partial \mathbf{m}} dt \,. \tag{2.14}$$

The term in brackets can be restated in the form of an adjoint wave equation

$$\left(\frac{\partial \mathcal{F}}{\partial u} \right)^\dagger \lambda = - \left(\frac{\partial J}{\partial u} \right)^\dagger , \tag{2.15}$$

where the superscript † indicates the adjoint operation, $\left(\frac{\partial \mathcal{F}}{\partial u}\right)^\dagger$ is the adjoint of the wave equation operator $\mathbf{A}$ (Appendix A), $-\left(\frac{\partial J}{\partial u}\right)^\dagger$ is the adjoint source labeled s_a in the following, and $\lambda(\mathbf{x}, t)$ is the adjoint wavefield. The adjoint wave equation is

$$\frac{1}{v^2} \frac{\partial^2 \lambda}{\partial t^2} - \nabla \cdot (\mathbf{m} \lambda) - \nabla^2 \lambda = s_a, \tag{2.16}$$

with

$$s_a = -\mathbf{R}^\dagger (\mathbf{R} u(\mathbf{m}) - d), \tag{2.17}$$

where $\mathbf{R}^\dagger$ is the adjoint of the sampling operator $\mathbf{R}$. Its role in equation 2.16 is to inject the data difference at the receiver positions.

Replacing the adjoint wavefield λ in equation 2.14, along with $\frac{\partial \mathcal{F}}{\partial \mathbf{m}} = \nabla u$, we obtain the gradient expression

$$\frac{dJ}{d\mathbf{m}} = \int_0^T \lambda^\dagger \nabla u \, dt \tag{2.18}$$

or

$$\frac{dJ}{d\mathbf{m}} = \int \delta(\tau) \lambda \star \nabla u \, d\tau, \tag{2.19}$$

where $\star$ represents a time crosscorrelation, sifted at zero-lag by the Dirac delta function $\delta(\tau)$ (Sava, 2014).

Equation 2.19 resembles the imaging condition defined in Claerbout (1971), also noted by Tarantola (1984). An important distinction in this case is that instead of correlating two scalar wavefields, we correlate the scalar adjoint wavefield λ with each component of the gradient of the forward wavefield ∇u. This is consistent with the fact that the reflectivity $\mathbf{m}$ is a vector, so its update must be driven by an object of the same dimension.

In summary, to obtain the gradient of the objective function with respect to the parameter $\mathbf{m}$, we solve two wave equations, forward and adjoint, multiply the adjoint wavefield by the gradient of the forward wavefield and integrate over time. The gradient expression can be used in any gradient-based optimization routine, such as conjugate gradient descent or L-BFGS to iteratively update the model given an initial guess

(Nocedal & Wright, 1999).

To validate our method, we perform tests using numerical examples, as discussed in the following section.

2.4 Numerical Examples

We show three examples. First, we analyze the sensitivity kernel for a source-receiver pair. Second, we demonstrate the method on a simple model containing two horizontal reflectors. Finally, we test the method on the Marmousi model.

2.4.1 Sensitivity Kernel

In this section, we compute the sensitivity kernel (Yao *et al.*, 2020) of the nonlinear imaging method and compare it with that of LSRTM. We consider a 2D constant-velocity model $v = 2.5$ km/s, a reflector at depth $z = 1.6$ km, and a pair source-receiver at positions $x = 1.0$ km and $x = 3.0$ km, respectively, both at depth $z = 0.5$ km.

The sensitivity kernel in Figure 2.1(a) only contains the migration component. It happens because the adjoint operator $\mathbf{B}^\dagger$ involved in the gradient computation of the least squares objective function does not depend on the parameter representing the image $\mathbf{m}$, considered a perturbation of the velocity model (Dai & Schuster, 2013). The gradient is given by

$$\frac{dJ}{d\mathbf{m}} = \mathbf{B}^\dagger \Delta d, \tag{2.20}$$

where Δd is the data difference between observed and predicted data. This data difference is backprojected into the model space by $\mathbf{B}^\dagger$ using a smooth velocity model, without creating backscattering events responsible for generating the tomography term.

On the other hand, the operator in equation 2.11 and its adjoint have a nonlinear dependence on the image vector parameter $\mathbf{m}$, and generate the tomography and migration kernel in Figure 2.1, for a perturbation in the form of a reflector at depth 1.6 km. Figure 2.1(b) denotes the horizontal component of the kernel, where a phase change in the migration isochrone indicated by the blue arrow can be observed. This phase change can be understood from the snapshots in Figure 2.2 at the particular time when the image point indicated by the blue arrow forms. Note that the snapshot of u_x changes polarity in the horizontal direction and that u_z changes polarity in the vertical direction, both relative to the source position. The phase change in Figure 2.1(b) occurs at the same position in the snapshot of u_x (Figure 2.2), directly below the source. As the snapshot of the adjoint wavefield λ does not show any phase change, the image formed by the product between the snapshots of u_x and λ reflects the phase change observed in u_x. Similarly, Figure 2.3 explains the opposite phase highlighted by the white arrows in the tomography kernel shown in Figure 2.1(c).

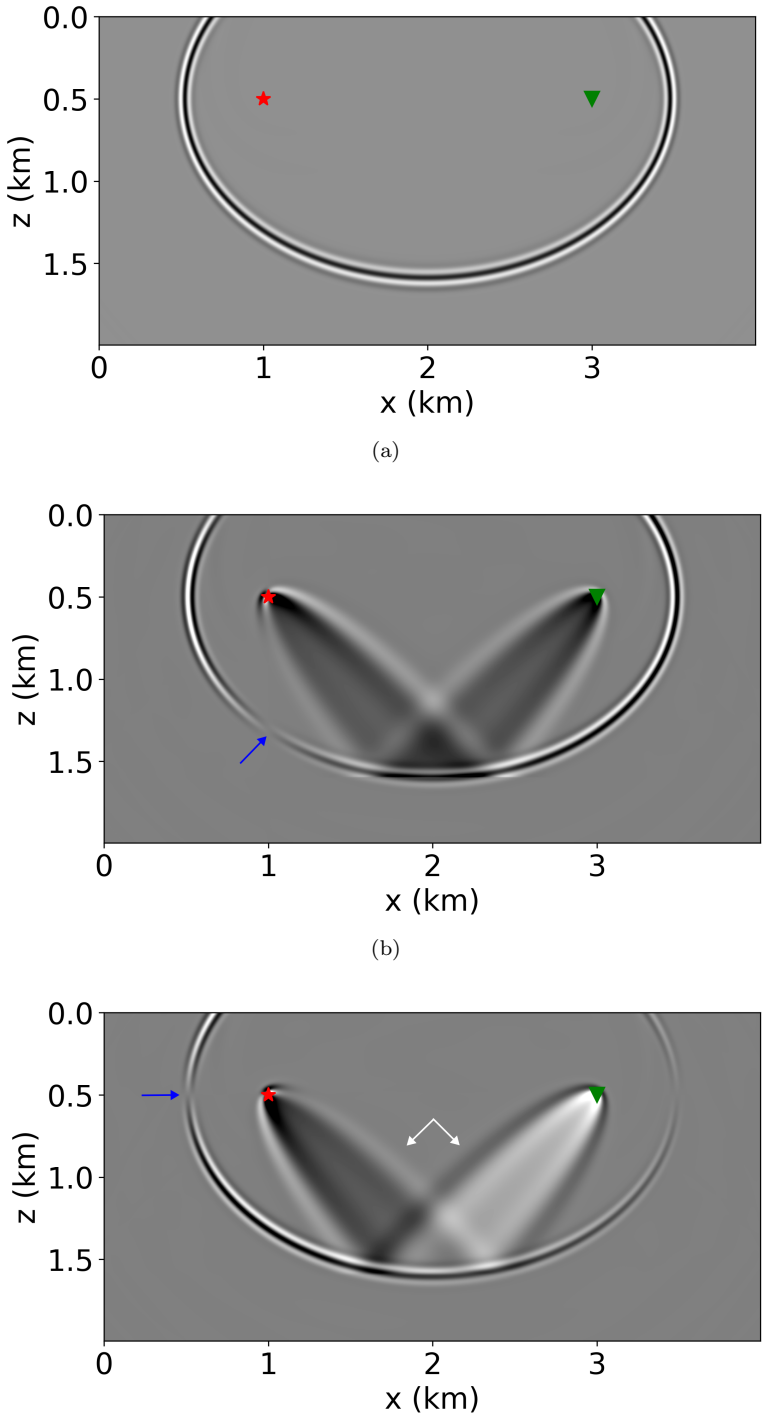


Figure 2.1 (a) Sensitivity kernel for one source (red star) and one receiver (green triangle) for: (a) LSRTM, (b) component x and (c) component z of the nonlinear imaging. The blue and white arrows indicate phase changes in the migration and tomography kernels, respectively.

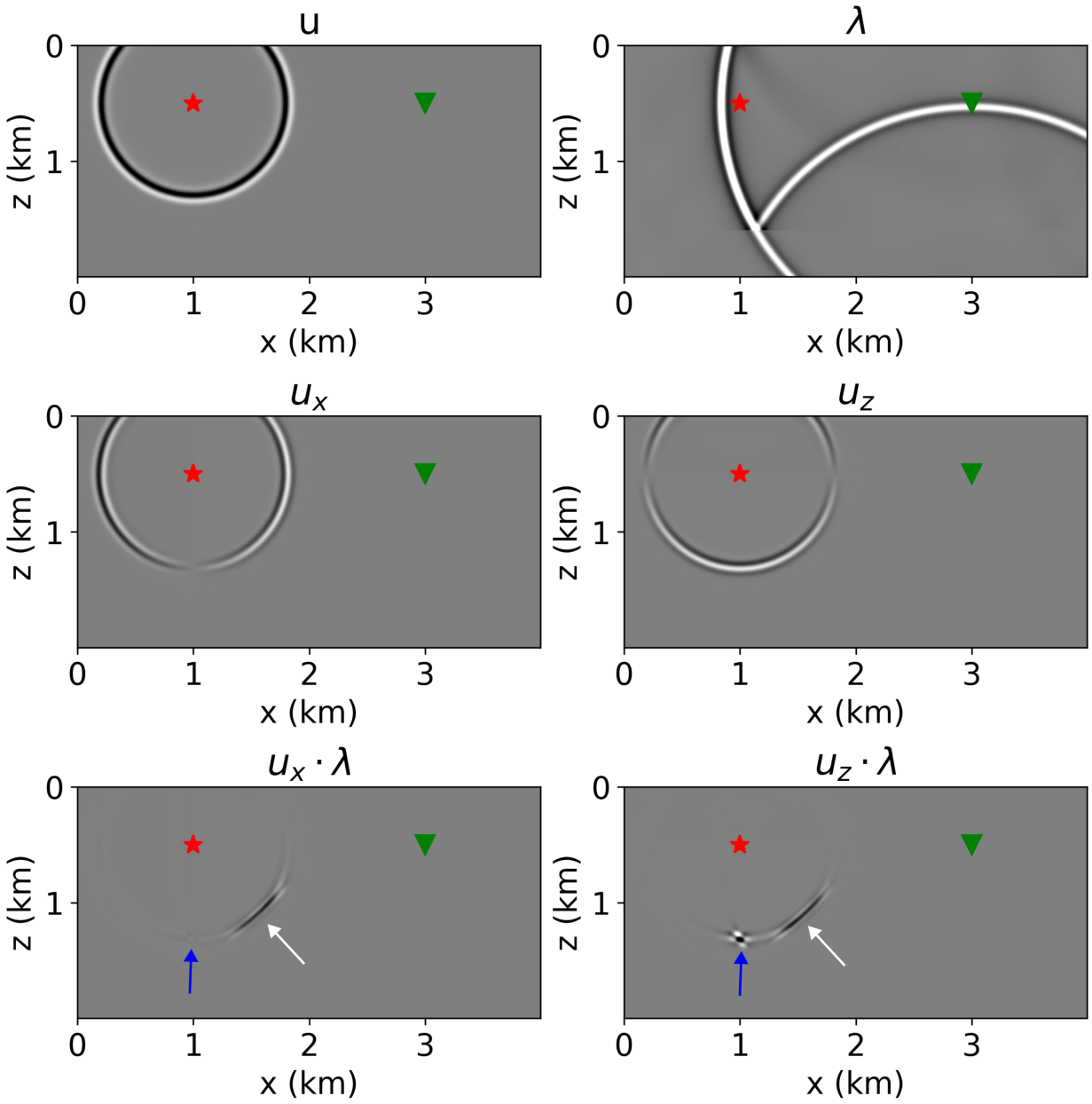


Figure 2.2 The top row depicts snapshots of forward u and adjoint wavefield λ at time t = 0.373 s. The middle row shows the gradient components of u in the horizontal direction u_x and in the vertical direction u_z. The bottom row shows image components formed at that particular time by multiplying u_x by λ and u_z by λ. The white arrows highlight the formation of the left component of the tomography kernel in Figure 2.1(b) and in Figure 2.1(c), having the same phase. The blue arrows point to the formation of the migration kernel in these figures.

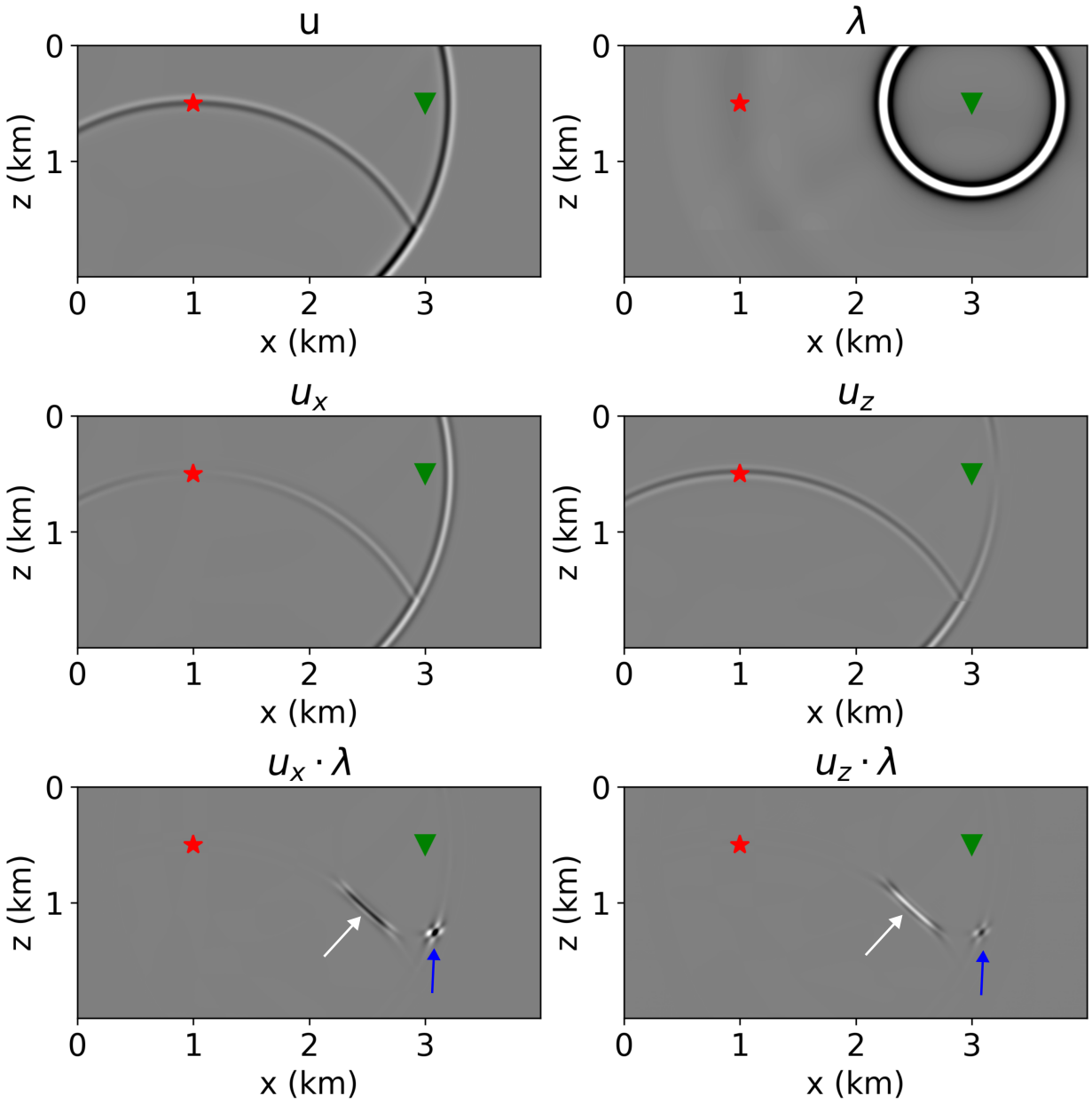


Figure 2.3 The top row depicts snapshots of forward u and adjoint wavefield λ at time $t = 0.932\,s$. The middle row shows the gradient components of u in the x direction u_x and in the z direction u_z. The bottom row shows image components formed at that particular time by multiplying u_x by λ and u_z by λ. The white arrows highlight the formation of the right component of the tomography kernel in Figure 2.1(b) and in Figure 2.1(c), having opposite phase. The blue arrows point to the formation of the migration kernel in these figures.

2.4.2 Two Horizontal Reflectors

In this section, we compare the performance of nonlinear imaging with RTM and LSRTM using a 2D model with two reflectors. The focus is on how these methods handle seismic data that contain both primary and multiple reflections, in addition to ghosts.

We consider a 2D model with a grid spacing of 2 m in the x and z directions. The velocity in Figure 2.4(a) is constant and the density model has a high impedance layer embedded in a constant background (Figure 2.4(b)). The product of these two forms the impedance, whose logarithm gradient $\nabla(lnZ)$ produces the horizontal and vertical reflectivity in Figure 2.4(c) and Figure 2.4(d), respectively. With this vector reflectivity and a Ricker wavelet having its frequency peak of 30 Hz, we model 30 shot gathers with a free surface boundary condition. The shots are evenly distributed on the surface at z = 10 m. The receivers are placed at the same depth and spread across all grid points on the surface. Figure 2.5(a) shows an example of a shot gather at position x = 0.6 km.

Using the modeled shot gathers as input, we perform the migration (first iteration) of linear (LSRTM) and nonlinear imaging. The results are shown in Figure 2.5. For linear imaging, the direct arrival is muted in the data prior to migration. We can see in Figure 2.5(b) that more than two reflectors appear in the image, which is due to multiples getting back projected from the data to the model space, thus creating fake reflectors. In addition to that, reflectors are not zero-phase, but present a phase distortion caused by the ghost "derivative" effect. Figure 2.5(c) and Figure 2.5(d) show the horizontal and vertical components of nonlinear migration. The image of the horizontal component contains weak energy compared to the vertical component, which is expected because the true model has nonzero values only in the vertical component. The images also contain fake reflectors due to the multiples present in the data. The reflectors are now zero-phase, but the significant side lobes are indicative of the ghost presence.

Figure 2.6 presents a comparison between linear (LSRTM) and nonlinear imaging results after 35 iterations. In the case of LSRTM, the images show sharper reflectors and more uniformly distributed amplitudes compared to migration due to wavelet deconvolution and balanced illumination. However, LSRTM still exhibits the ghost effect and highlights multiple-induced reflectors. On the other hand, Figure 2.6(c) displays the vertical component of nonlinear imaging, which successfully recovers the two reflectors while attenuating spurious events and ghosts. The nonlinear modeling engine fits not only primary events, but also multiples. An unintended consequence of striving to recover sharp contrasts from band-limited data is the emergence of "ringing" (Lu *et al.*, 2018b), masquerading as high-frequency "thin reflectors" surrounding the main ones.

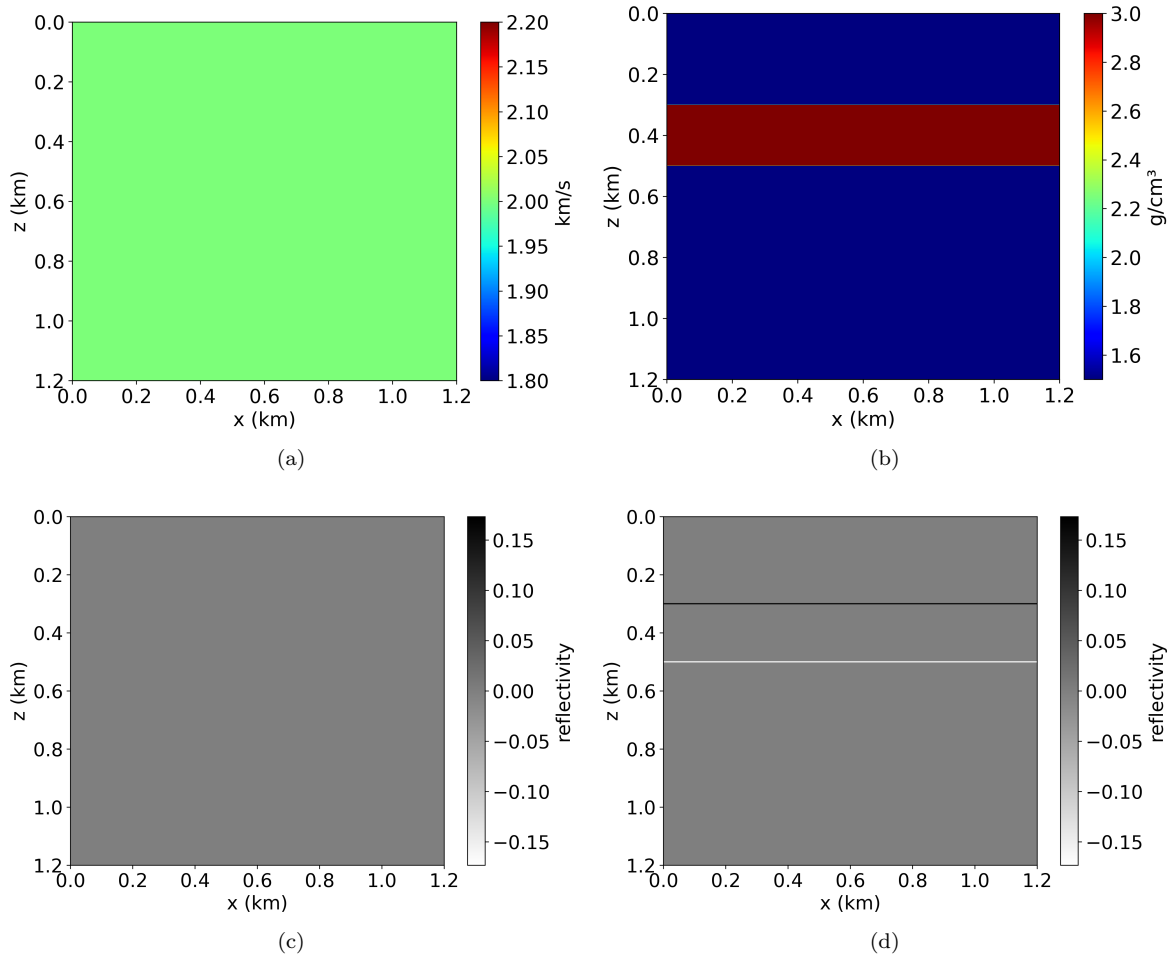


Figure 2.4 (a) Velocity model, (b) density model, (c) horizontal component and (d) vertical component of reflectivity calculated as the normalized gradient of impedance: $\frac{\nabla Z}{Z}$.

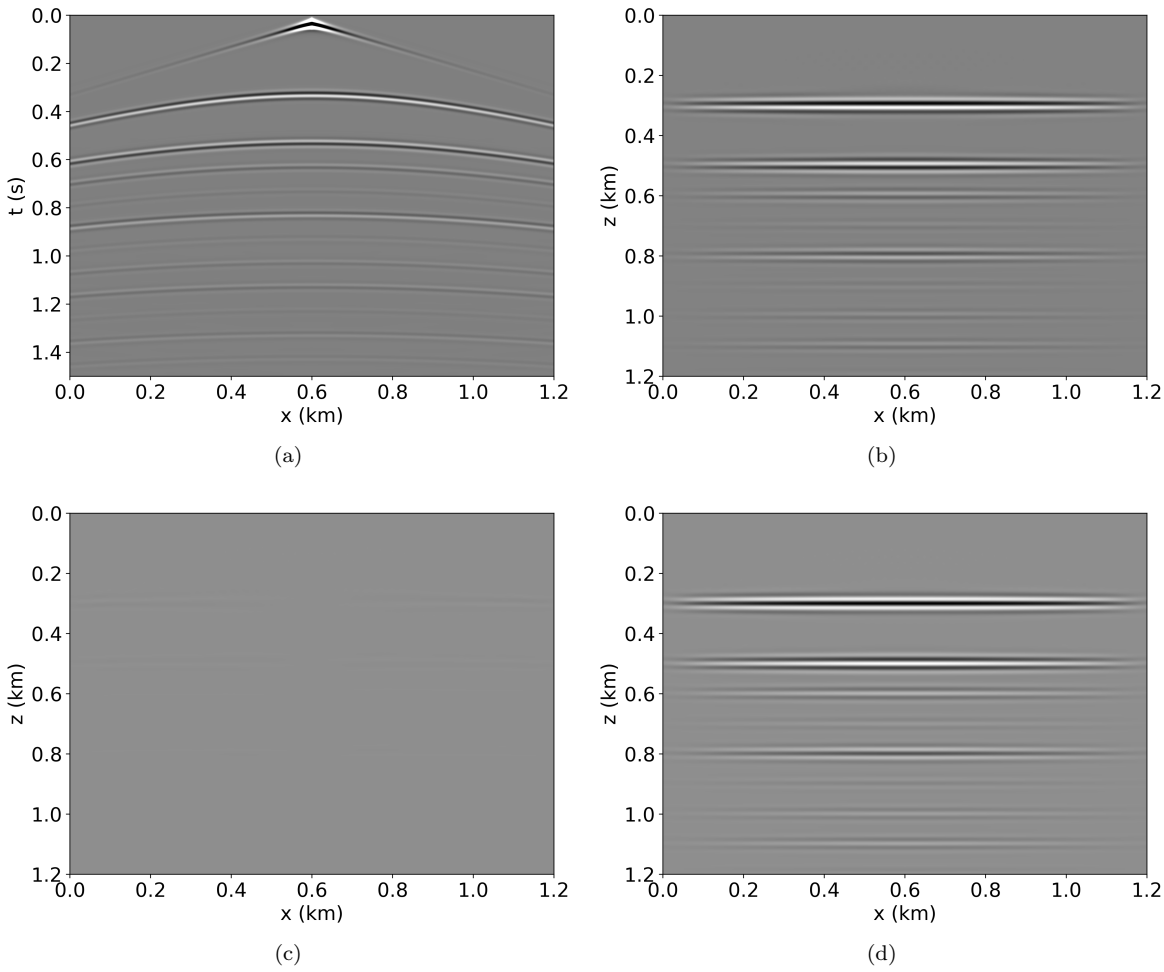
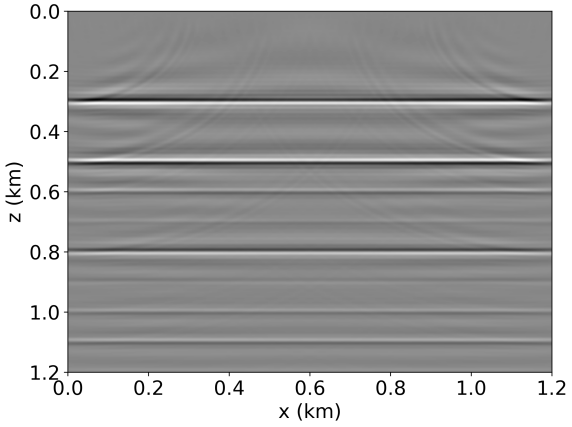
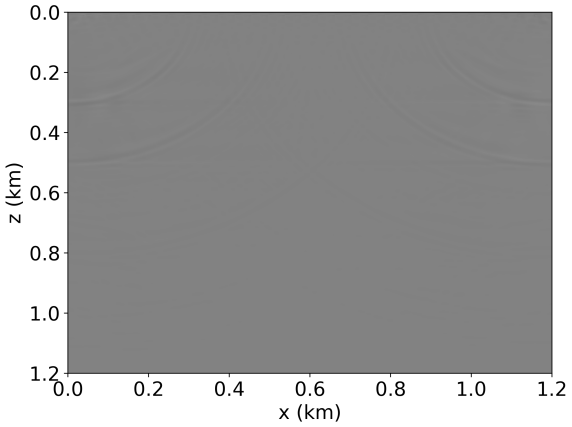


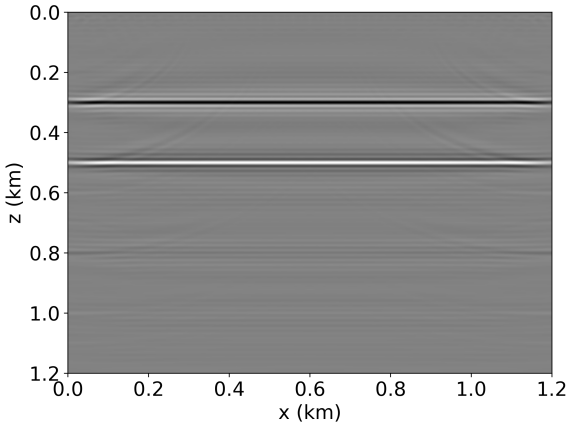
Figure 2.5 (a) Simulated shot gather at shot position x = 600 m and z = 10 m, (b) first iteration of LSRTM, which corresponds to migration, (c) and (d) are the components x and z of the first iteration of nonlinear imaging, respectively.



(a)



(b)



(c)

Figure 2.6 (a) LSRTM image, (b) x and (c) z components of the nonlinear imaging, after 35 iterations.

2.4.3 Marmousi Model

To further evaluate the nonlinear imaging in comparison with linear methods (RTM and LSRTM) in a complex setting, we experiment with the Marmousi model obtained from Operto (2023) and modified to vertically extend the water layer vertically. Figures Figure 2.7(a) and Figure 2.7(b) show the velocity and density models equally sampled at 4 m. The horizontal and vertical reflectivities in Figures Figure 2.8(a) and Figure 2.8(b) are computed by taking the gradient of the logarithm of the impedance formed by multiplying the density and velocity models. We model 50 shots to the maximum recording time of 4 s using equation 2.4, a Ricker wavelet with frequency peak at 20 Hz and the computed reflectivity (Figure 2.8) as input. The shots are uniformly distributed at z = 10 m, and the receivers are placed at every grid point at z = 10 m.

Figure 2.9 contains the result of the initial iteration (migration) of linear (LSRTM) and nonlinear imaging. The initial guess is zero for the horizontal and vertical components of the vector image in the L-BFGS optimization. We observe coherent events traversing actual reflectors, signifying crosstalk between events that should not be cross-correlated (Lu *et al.*, 2021). The phase distortion caused by the interference of ghosts is also visible, and is best seen at the water bottom.

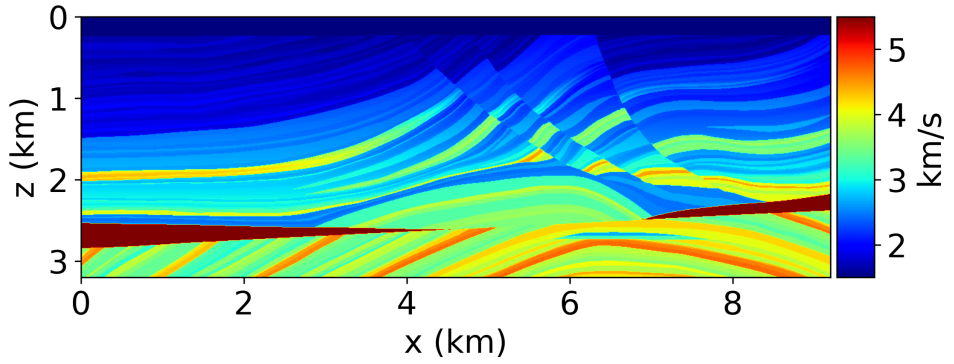
After 35 iterations, we obtain the images shown in Figure 2.10. The nonlinear imaging output provides clearer images (Figure 2.10(b) and Figure 2.10(c)) compared to its initial iteration, significantly reducing cross-talk noise in the horizontal and vertical components. In contrast, the LSRTM image in Figure 2.10(a) exhibits increased noise, as it attempts to fit the data containing multiples and ghosts using a forward modeling approach capable of simulating only primaries. In addition, the ghost effect becomes more evident in the LSRTM image, while it attenuates in the nonlinear case, where reflectors are represented by a symmetric pulse with reduced side lobes.

2.4.4 Misfit Convergence: Adjoint vs Time-reversal

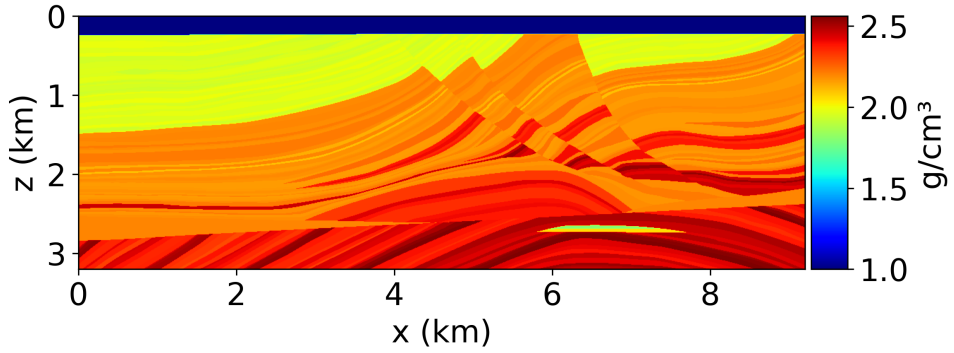
We compare the convergence of the data misfit objective function in the Marmousi example for the L-BFGS optimizer in two cases:

- The adjoint wavefield λ in equation 2.19 is calculated from the correct adjoint wave equation 2.16.
- The wave equation operator in equation2.11 is assumed to be self-adjoint, and the following equation instead is solved backward in time to obtain λ:

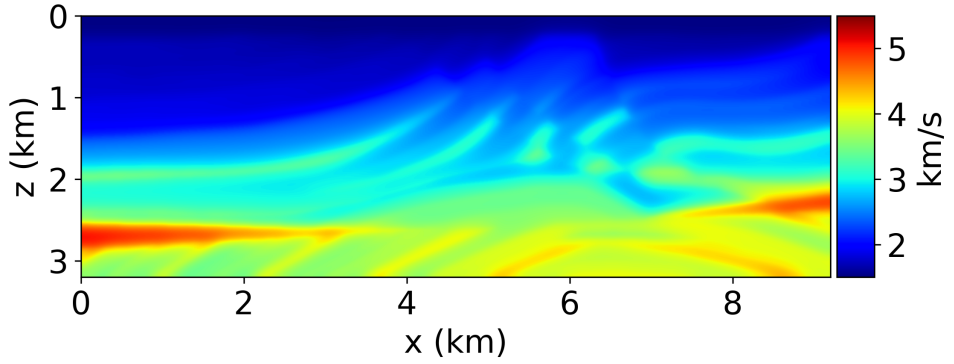
$$\frac{1}{v^2}\frac{\partial^2 \lambda}{\partial t^2} + \mathbf{m} \cdot \nabla u - \nabla^2 \lambda = s_a \,. \tag{2.21}$$



(a)

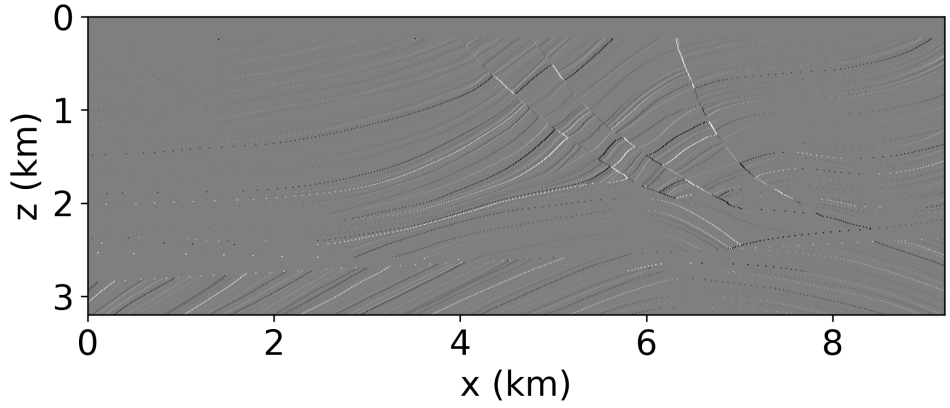


(b)

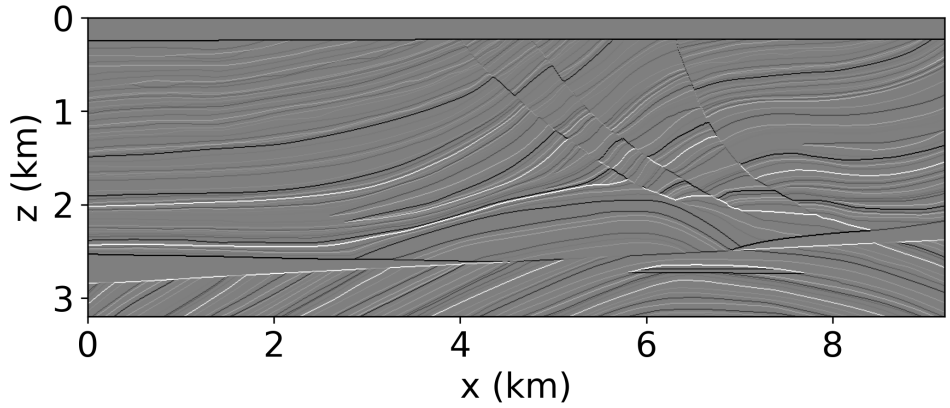


(c)

Figure 2.7 (a) Marmousi velocity model, (b) density model and (c) smooth version of (a) used in migrations (first iteration of inversion) and inversions.

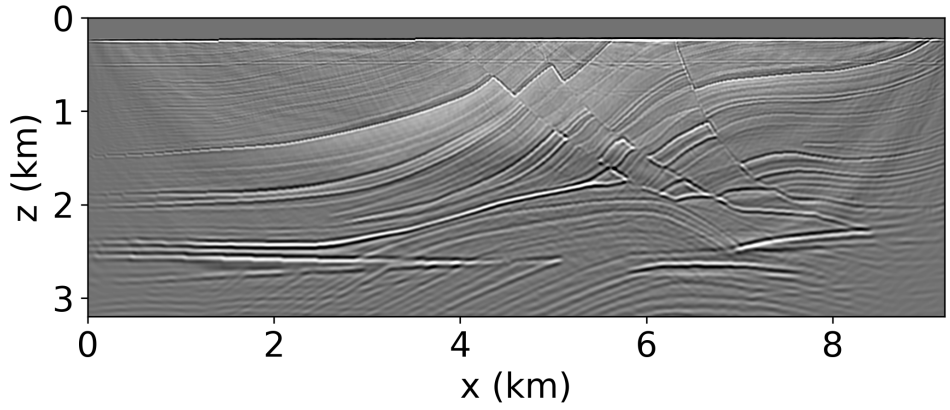


(a)

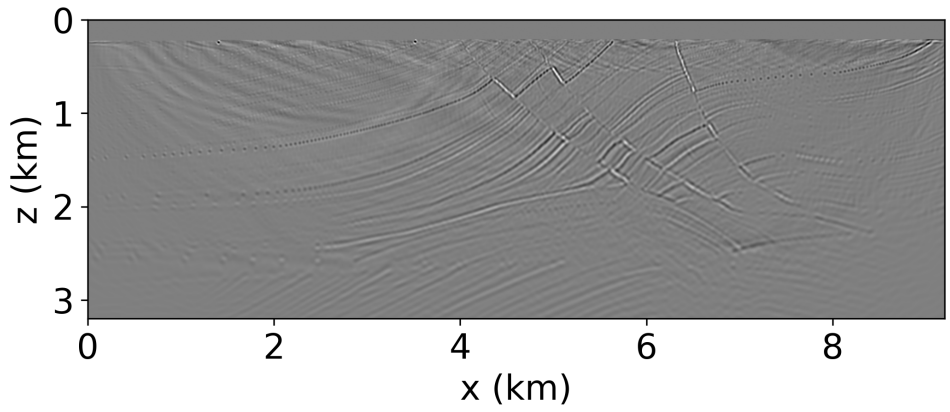


(b)

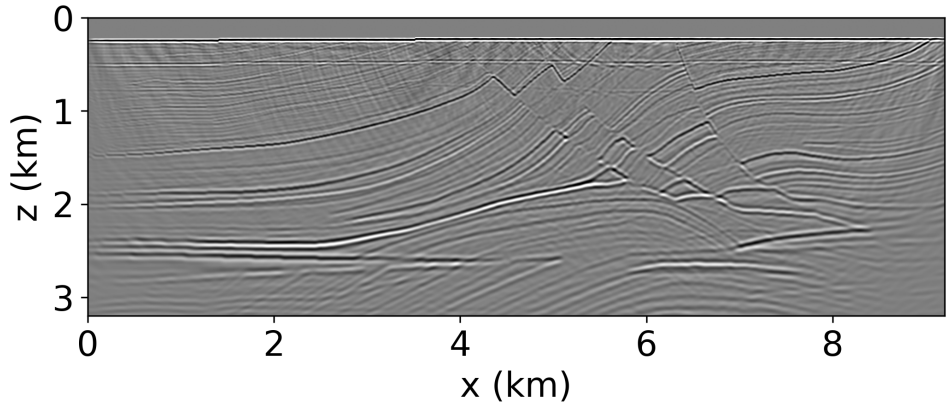
Figure 2.8 (a) Reflectivity component x and (b) reflectivity component z, obtained through $\frac{\nabla Z}{Z}$ using the velocity and density models in Figure 2.7(a) and Figure 2.7(b), respectively.



(a)

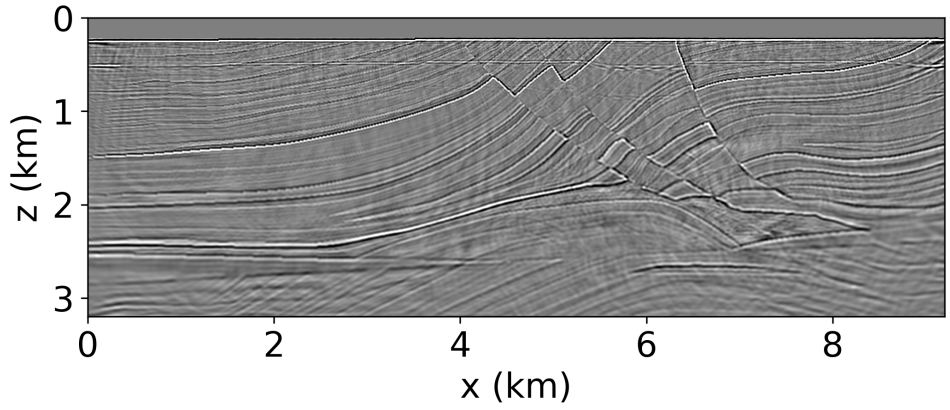


(b)

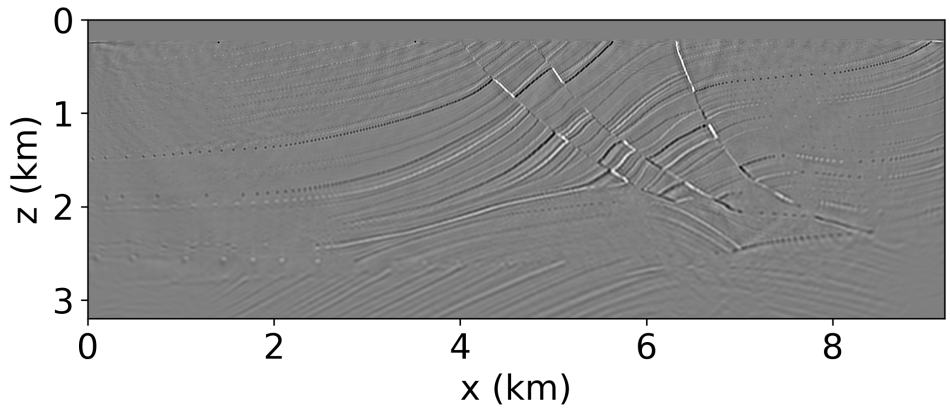


(c)

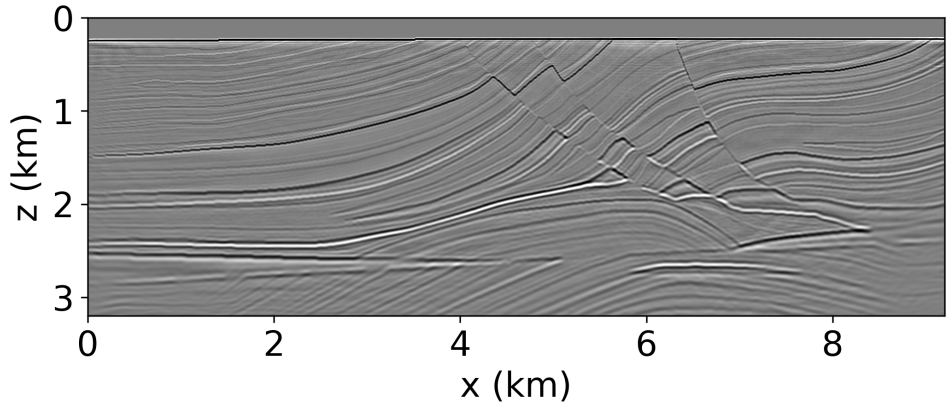
Figure 2.9 (a) First iteration of LSRTM, which is equivalent to an RTM (Reverse Time Migration), (b) x and (c) z components of the image vector obtained from the first iteration of the nonlinear imaging method.



(a)



(b)



(c)

Figure 2.10 (a) LSRTM result after 35 iterations, (b) x and (c) z components of the image vector after 35 iterations of the nonlinear imaging method.

Figure 2.11 illustrates that employing time-reversal to calculate λ requires 35% more iterations to achieve the same misfit reduction compared to using the adjoint wave equation. We note that this speed-up is specific to this model and wave equation parameterization. The impact of using the incorrect adjoint modeling operator varies based on the disparity between time-reversal and the actual adjoint solution. Since gradient-based optimizers rely on accurate gradients from the correct adjoint equation solution, it is essential to provide precise estimations for optimal results. In this experiment, only the adjoint wave equation was incorrect. Other operators crucial for the adjoint computation, like sampling $\mathbf{R}$ and injection $\mathbf{R}^\dagger$, use correct adjoint pairs.

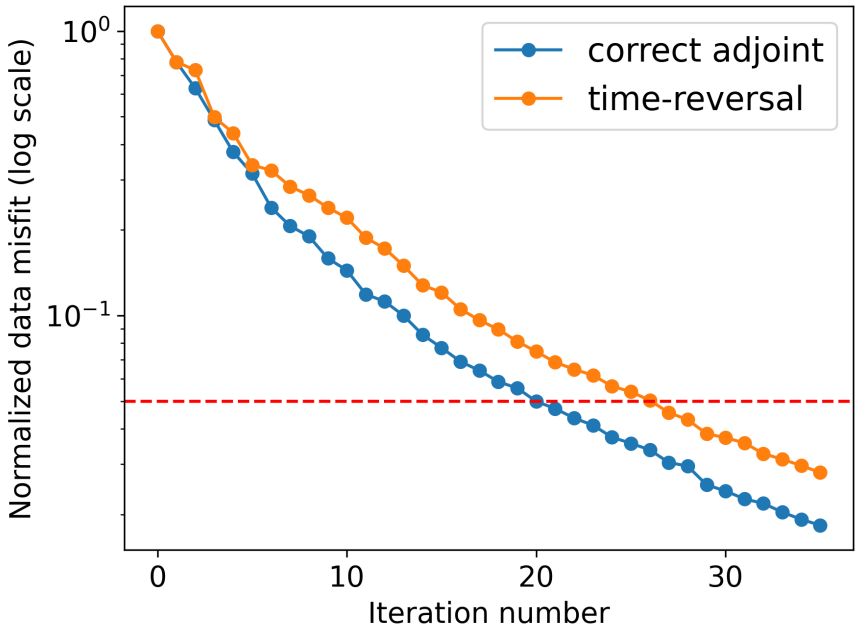


Figure 2.11 Comparison of the data misfit convergence for the nonlinear imaging method when the correct adjoint or the time-reversal only is used to compute the adjoint wavefield λ as part of the misfit gradient. The misfit level of 5% is indicated by the dashed red line.

2.5 Discussions

The numerical examples presented highlight the ability of the nonlinear imaging method to mitigate crosstalk arising in linear methods, specifically RTM and LSRTM when multiples are present in the data, besides compensating for uneven illumination due to acquisition footprint and increase in resolution. The crosstalk generated by linear methods manifests as noise in the image, underscoring the importance of adopting a nonlinear approach to enhance imaging quality and minimize artifacts. In addition, expressing the image as a vector facilitates a better representation of contrasts in the subsurface across distinct directions.

We emphasize the importance of the adjoint of the forward modeling as it plays a crucial role in computing the correct gradient necessary for gradient-based optimization methods. The paper illustrates that in examples like the Marmousi model, employing the correct adjoint leads to faster convergence compared to the time-reversal of the forward modeling.

The sensitivity kernel illustrated in Figure 2.1 for the nonlinear imaging includes low-wavenumber components, but does not contaminate the final inversion images as observed in Figures Figure 2.6 and Figure 2.10. The migration kernel has a higher amplitude compared to the tomographic kernel and dominates the inversion, making kernel separation unnecessary.

In the numerical examples presented in this paper, the wavelet is assumed to be known, but in practical applications, its accurate estimation influences the results. The waveform misfit objective function in Equation 2.7 is sensitive to differences in wavelet strength and phase mismatches caused by velocity errors, and is no different from LSRTM in this regard. To enhance the inversion robustness, exploring alternative objective functions becomes essential. For instance, Zhang *et al.* (2015) employ a zero-lag cross-correlation objective function to alleviate strength differences between observed and predicted data in LSRTM. This objective function can be interpreted as a cosine similarity measure (Xia *et al.*, 2015), evaluating the parallelism of observed and predicted data as vectors belonging to a multidimensional data space through the normalized dot product. When both data vectors point in the same direction, the cosine of the formed angle is 1, while the cosine for opposite directions is -1. The optimization goal is to maximize the cosine objective function. However, this approach also assumes that the velocity model is known or optimized to match the data travel times. In scenarios where small velocity errors are present, an alternative is to combine cosine similarity with the use of dynamic warping to match the phase of the data before comparing the amplitudes, as suggested by (Luo & Hale, 2014). This combined approach offers a potential solution to address phase discrepancies caused by velocity errors and differences in wavelet strength, further enhancing the reliability of the inversion process.

2.6 Conclusions

We formulate a nonlinear imaging method that proves to be a robust solution for overcoming crosstalk inherent in linear methods like RTM and LSRTM if the data are contaminated with multiples. This nonlinear approach not only mitigates such noise, but also addresses issues related to uneven illumination and ghosts, resulting in improved imaging quality. Like conventional LSRTM, the nonlinear imaging method relies on an accurate velocity model to match the data kinematics and avoid cycle skipping. Although we assume an isotropic acoustic approximation, our method can include anisotropy, extend to elastic media, and account for attenuation effects to match events observed in real data. Obtaining the adjoint wave equation as part

of the gradient calculation accelerates convergence, which is especially relevant when using wave equations encapsulating complex physical relationships.

2.7 Acknowledgments

The first author thanks Petrobras for the sponsorship that made this research possible, and Alan Souza for useful discussions. We thank the sponsors of the Center for Wave Phenomena at the Colorado School of Mines. The domain-specific language (DSL) Devito (`https://www.devitoproject.org/`) was used to implement the method and generate the numerical examples.

2.8 Appendix A: Adjoint Wave Equation Operator

The objective of this section is to derive the adjoint operator for the wave equations 2.3 and 2.4. The latter is a specialized form of the former, obtained by excluding the velocity gradient term. Obtaining the correct adjoint operator is crucial for accurately calculating the gradient of the objective function through the adjoint state method.

Consider two wavefields $u(\mathbf{x}, t)$ and $\lambda(\mathbf{x}, t)$ in the function space $\mathcal{U}$ of all possible wavefields and the linear operator

$$\mathbf{A} = \frac{1}{v^2}\frac{\partial^2}{\partial t^2} - \frac{\nabla v}{v} \cdot \nabla + \mathbf{m} \cdot \nabla - \nabla^2, \tag{2.22}$$

or

$$\mathbf{A} = \frac{1}{v^2}\frac{\partial^2}{\partial t^2} - \mathbf{S}\mathbf{G}_v^\dagger\mathbf{G} + \mathbf{M}^\dagger\mathbf{G} - \mathbf{L} \tag{2.23}$$

where we use the linear operators

$$\begin{cases} \mathbf{S} = diag\left(\frac{1}{v}\right) \\ \mathbf{G}_v = \nabla v \\ \mathbf{G} = \nabla \\ \mathbf{M} = diag(\mathbf{m}) \\ \mathbf{L} = \nabla^2 . \end{cases} \tag{2.24}$$

The term $diag(\cdot)$ generates a diagonal matrix operator from the argument values. If the argument is a vector, such as $\mathbf{m}$, the $diag()$ function generates a diagonal matrix for each vector component, resulting in a columnar block matrix. These matrices are not explicitly formed but are implemented as linear operators representing matrix operations (Claerbout, 2014).

The adjoint operator $\mathbf{A}^\dagger$ satisfies the following property for all vectors u and λ in the function space $\mathcal{U}$ (Strang, 2007):

$$\langle \mathbf{A}u, \lambda \rangle_{\mathcal{U}} = \langle u, \mathbf{A}^\dagger\lambda \rangle_{\mathcal{U}}, \tag{2.25}$$

where $\langle \cdot, \cdot \rangle_{\mathcal{U}}$ represents the inner product in the space $\mathcal{U}$ and is defined as

$$\begin{aligned} \langle \mathbf{A}u, \lambda \rangle_{\mathcal{U}} &= \int_0^T \int_\Omega \mathbf{A}u\lambda \, d\mathbf{x} \, dt, \\ &= \int_0^T \langle \mathbf{A}u, \lambda \rangle_{\mathbf{x}} \, dt \end{aligned} \tag{2.26}$$

where T is the maximum time, Ω is the spatial domain of integration and $\langle \cdot, \cdot \rangle_{\mathbf{x}}$ with subscript $\mathbf{x}$ denotes integration over space (Plessix, 2006).

The inner product is linear in the first argument (Axler, 2015), consequently we can write

$$\begin{aligned} \langle \mathbf{A}u, \lambda \rangle_{\mathcal{U}} &= \left\langle \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} - \mathbf{S}\mathbf{G}_v^\dagger \mathbf{G}u + \mathbf{M}^\dagger \mathbf{G}u - \mathbf{L}u, \lambda \right\rangle_{\mathcal{U}} \\ &= \left\langle \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2}, \lambda \right\rangle_{\mathcal{U}} + \left\langle -\mathbf{S}\mathbf{G}_v^\dagger \mathbf{G}u + \mathbf{M}^\dagger \mathbf{G}u - \mathbf{L}u, \lambda \right\rangle_{\mathcal{U}} \end{aligned} . \tag{2.27}$$

With the goal of transferring the second-order derivative operator to the adjoint variable λ, we follow Plessix (2006) and integrate by parts in time the first term in equation 2.27

$$\begin{aligned} \int_0^T \left\langle \frac{1}{v^2}\frac{\partial^2 u}{\partial^2 t}, \lambda \right\rangle_{\mathbf{x}} dt &= \int_0^T \left\langle u, \frac{1}{v^2}\frac{\partial^2 \lambda}{\partial^2 t} \right\rangle_{\mathbf{x}} dt + \left\langle \frac{1}{v^2}\frac{\partial u(T)}{\partial t}, \lambda(T) \right\rangle_{\mathbf{x}} \\ - \left\langle \frac{1}{v^2}\frac{\partial u(0)}{\partial t}, \lambda(0) \right\rangle_{\mathbf{x}} &- \left\langle u(T), \frac{1}{v^2}\frac{\partial \lambda(T)}{\partial t} \right\rangle_{\mathbf{x}} + \left\langle u(0), \frac{1}{v^2}\frac{\partial \lambda(0)}{\partial t} \right\rangle_{\mathbf{x}} . \end{aligned} \tag{2.28}$$

Using the wavefield u as the solution of the forward wave equation, which has homogeneous initial conditions, $u(0) = 0$ and $\frac{\partial u(0)}{\partial t} = 0$, we eliminate the third and fifth term on the right-hand side of equation 2.28. As we are only interested in the values of $\lambda(\mathbf{x}, t)$ for $0 \leq t \leq T$, we can set the final conditions $\lambda(T) = 0$ and $\frac{\partial \lambda(T)}{\partial t} = 0$ (Demanet, 2021), which means that when solving the adjoint wave equation it has to be solved backward in time. With these final conditions, we further simplify 2.28 to:

$$\int_0^T \left\langle \frac{1}{v^2}\frac{\partial^2 u}{\partial^2 t}, \lambda \right\rangle_{\mathbf{x}} dt = \int_0^T \left\langle u, \frac{1}{v^2}\frac{\partial^2 \lambda}{\partial^2 t} \right\rangle_{\mathbf{x}} dt \tag{2.29}$$

If we apply the adjoint definition in equation 2.25 to the second term on the right-hand side of equation 2.27, we obtain

$$\langle -\mathbf{S}\mathbf{G}_v^\dagger \mathbf{G}u + \mathbf{M}^\dagger \mathbf{G}u - \mathbf{L}u, \lambda \rangle_{\mathcal{U}} = \langle u, (-\mathbf{S}\mathbf{G}_v^\dagger \mathbf{G} + \mathbf{M}^\dagger \mathbf{G} - \mathbf{L})^\dagger \lambda \rangle_{\mathcal{U}} \tag{2.30}$$

or, using the property that adjoint of a composition of operators is the same as taking the adjoint of each operator separately and then composing them in the reverse order,

$$\langle -\mathbf{S}\mathbf{G}_v^\dagger \mathbf{G}u + \mathbf{M}^\dagger \mathbf{G}u - \mathbf{L}u, \lambda \rangle_{\mathcal{U}} = \langle u, -\mathbf{G}^\dagger \mathbf{G}_v \mathbf{S}^\dagger \lambda + \mathbf{G}^\dagger \mathbf{M}\lambda - \mathbf{L}^\dagger \lambda \rangle_{\mathcal{U}} . \tag{2.31}$$

As demonstrated in Claerbout (2014) and Strang (2007), the Laplacian operator is self-adjoint $\mathbf{L} = \mathbf{L}^\dagger$ and the adjoint of the gradient operator is the negative of the divergence operator $\mathbf{D} = \nabla \cdot$, i.e., $\mathbf{G}^\dagger = -\mathbf{D}$.

Additionally, $\mathbf{S} = \mathbf{S}^\dagger$ because it represents a diagonal operator. Thus, equation 2.31 simplifies to

$$\langle -\mathbf{S}\mathbf{G}_v^\dagger \mathbf{G}u + \mathbf{M}^\dagger \mathbf{G}u - \mathbf{L}u, \lambda \rangle_{\mathcal{U}} = \langle u, \mathbf{D}\mathbf{G}_v \mathbf{S}\lambda - \mathbf{D}\mathbf{M}\lambda - \mathbf{L}\lambda \rangle_{\mathcal{U}} \tag{2.32}$$

Replacing the results of equations 2.29 and 2.32 in equation 2.27

$$\langle \mathbf{A}u, \lambda \rangle_{\mathcal{U}} = \left\langle u, \frac{1}{v^2}\frac{\partial^2 \lambda}{\partial t^2} + \mathbf{D}\mathbf{G}_v \mathbf{S}\lambda - \mathbf{D}\mathbf{M}\lambda - \mathbf{L}\lambda \right\rangle_{\mathcal{U}} = \langle u, \mathbf{A}^\dagger \lambda \rangle_{\mathcal{U}}\,, \tag{2.33}$$

leads to

$$\mathbf{A}^\dagger = \frac{1}{v^2}\frac{\partial^2}{\partial t^2} + \mathbf{D}\mathbf{G}_v \mathbf{S} - \mathbf{D}\mathbf{M} - \mathbf{L} \tag{2.34}$$

or

$$\mathbf{A}^\dagger = \frac{1}{v^2}\frac{\partial^2}{\partial t^2} + \nabla \cdot \frac{\nabla v}{v} - \nabla \cdot \mathbf{m} - \nabla^2\,. \tag{2.35}$$

Comparing the adjoint operator $\mathbf{A}^\dagger$ in equation 2.35 with the operator $\mathbf{A}$ in equation 2.22, we note that: while in the second and third terms of 2.22 we have a dot product between vectors and gradient operators, in 2.35 these terms have opposite signs and correspond to the divergence operator acting on vectors. When acting on a wavefield, the adjoint operator runs 2.22 forward in time and the operator 2.35 runs backward in time, as a consequence of the homogeneous terminal conditions imposed in equation 2.28. If v is smooth, the second term on the right-hand side of equation 2.35 can be ignored and the resulting operator leads to the adjoint wave equation 2.16.

CHAPTER 3

CONCLUSIONS

In this thesis, I propose an isotropic nonlinear acoustic imaging method that takes into consideration ghosts and multiples in addition to the primaries present in the input data. The inverse problem, parameterized by a vector quantity representing the image, has the objective to minimize the mismatch between observed and predicted data. I test the method using numerical examples and compare the resulting images with those obtained by conventional techniques based on the Born approximation to underscore the ability of my nonlinear approach to handle multiply scattered events. I derive the gradient of the formulated objective function, and the adjoint of the wave equation operator it depends on, and compare the convergence of the data misfit objective function relative to approximate adjoints obtained by simple time-reversal of the forward wave equation. The numerical comparison illustrates the effectiveness of correct adjoint operator to achieve faster convergence.

Future work could extend this method to elastic and anisotropic media. Although the acoustic approximation of the Earth has proven successful in various applications, such as locating oil and gas reservoirs, identifying and mapping hydrological and geothermal sources and others, the Earth's medium is characterized by more complex physics. As the demand for more information increases, a more realistic description of wave phenomena is necessary. However, these additional complexities require higher computational cost. For example, the elastic approach requires finer sampling to accommodate the lower S-wave wavelength resulting from smaller velocities. Moreover, the correct adjoint is key to avoiding further cost increases due to slow convergence.

To compute the gradient of the objective function, I utilize all seismic shot gathers in the data, implying that the gradient is proportional to the number of shots. An alternative approach may use a subset of randomly sampled shots to alleviate computational load, a technique known as Stochastic Gradient Descent (Bishop & Nasrabadi, 2006), commonly used in machine learning for high-dimensional optimization problems to reduce the costs. However, this approximation requires a detailed investigation of the number of shots needed in the subset, and the statistical distribution used to sample the shots space. Future work could also explore shots selection guided by a criterion that minimizes angular differences between their respective gradients, to reduce information redundancy and improve convergence while reducing cost (van Herwaarden *et al.*, 2020).

Another opportunity for future work is to formulate and solve the nonlinear seismic imaging inverse problem in the image domain, minimizing the objective function

$$J = \frac{1}{2}\|\mathbf{H}\mathbf{m} - \mathbf{m}_{mig}\|^2 \,, \tag{3.1}$$

where $\mathbf{m}_{mig}$ is the migrated image obtained by applying the adjoint of the modeling operator $\mathbf{A}^\dagger$ to the data $\mathbf{d}$

$$\mathbf{m}_{mig} = \mathbf{A}^\dagger \mathbf{d} \,. \tag{3.2}$$

The operator $\mathbf{A}^\dagger\mathbf{A}$ represents image blurring and the migrated image $\mathbf{m}_{mig}$ is a blurred version of the image $\mathbf{m}$ we seek to recover by solving

$$\mathbf{m} = \mathbf{H}^{-1}\mathbf{m}_{mig} = [\mathbf{A}^\dagger\mathbf{A}]^{-1}\mathbf{m}_{mig}, \,. \tag{3.3}$$

Equation 3.3 suggests that the image $\mathbf{m}$ can be obtained by applying the inverse of the operator $\mathbf{H}$, in a process called migration deconvolution (Hu *et al.*, 2001), which brings computational advantages compared to the formulation in the data domain (Yu & Schuster, 2003).

Equation 3.1 assumes that the operator $\mathbf{H}$ is linear, but it can be adapted to the nonlinear case by assuming $\mathbf{H} = \mathbf{A}^\dagger\mathbf{A}$, where $\mathbf{A}$ is the wave equation operator given by equation 2.11, which depends on $\mathbf{m}$. The blurring operator depends on $\mathbf{m}$, $\mathbf{H} = \mathbf{H}(\mathbf{m})$, and thus equation 3.1 becomes

$$J = \frac{1}{2}\|\mathbf{H}(\mathbf{m}) - \mathbf{m}_{mig}\|^2 \,. \tag{3.4}$$

Minimizing 3.4 using a gradient based optimization method involves updating not only the image $\mathbf{m}$ at each iteration, but also the operator $\mathbf{H}$, since it depends on $\mathbf{m}$. I do not expect savings in computation compared to implementation in the data domain to be as significant as in the linear case, where the operator $\mathbf{H}$ is kept constant through the iterations. In the linear case, only one modeling and one migration are performed to build $\mathbf{H}$, after which, the optimization is performed in the image domain only. In the nonlinear situation, it is worth investigating if solving the problem directly in the image domain offers computational benefits. This could help avoid expensive line searches (Nocedal & Wright, 1999) in the data domain where the objective function needs repeated evaluations through the solution of the wave equation. In addition, it is important to assess potential improvements in quality or stability of the solution when changing to the model domain approach.

The nonlinear inverse imaging problem formulated in this thesis does not incorporate regularization, although regularization provides additional information to choose a single solution from infinitely many that also match the data (Oldenburg & Li, 2005). In Figure 2.6(c), the recovered image shows "ringing" artifacts surrounding the reflectors that are not present in the true reflectivity (Figure 2.4(d)), which have the

potential to be misinterpreted as thin layering. As the observed data do not fully constrain the solution due to its limited frequency bandwidth and aperture, the part of the solution that does not affect the data (null space) is filled with such spurious information. To mitigate this problem, complementary information should be provided, e.g., by augmenting the objective function with a term that penalizes unwanted characteristics in the solution. The regularization penalty is controlled by a regularization parameter that balances the data objective function (misfit) and the model objective function. These terms of the objective function, have different units (data versus reflectivity), such that picking a regularization parameter that provides the right balance is not straightforward. Some techniques such as the L-curve and generalized cross-validation methods serve this purpose (Oldenburg & Li, 2005). However, it implies solving an already costly inverse problem multiple times for a range of regularization parameter values to choose one according to certain criteria. In addition, a decision has to be made on the interval of regularization parameters to explore. For example, in the L-curve method, the chosen interval might not include the searched "knee" of the L-curve. Future research could also explore this discussion on what regularization to use in the nonlinear inverse problem and how to determine its amount in a cost-effective way.

While my focus is on seismic imaging, the nonlinear imaging method can also be applied to other fields, such as ultrasound imaging. In ultrasound, the existence of reverberations creates artifacts that mimic false body structures in the image. Considering these reverberations during the imaging process can not only prevent such artifacts, but also enhance the representation of the actual structure. This application has the potential to reduce uncertainties in medical diagnoses.

REFERENCES

Axler, Sheldon. 2015. *Linear algebra done right*. Springer.

Baysal, Edip, Kosloff, Dan D, & Sherwood, John WC. 1983. Reverse time migration. *Geophysics*, **48**(11), 1514–1524.

Berkhout, AJ. 2014. Review paper: An outlook on the future of seismic imaging. *Part III: Joint migration inversion: Geophysical Prospecting*, **62**, 950–971.

Berkhout, AJ, & Verschuur, Dirk J. 1994. Multiple technology: Part 2, migration of multiple reflections. *Pages 1497–1500 of: SEG Technical Program Expanded Abstracts 1994*. Society of Exploration Geophysicists.

Berkhout, AJ, & Verschuur, DJ. 2011. Full wavefield migration, utilizing surface and internal multiple scattering. *Pages 3212–3216 of: SEG Technical Program Expanded Abstracts 2011*. Society of Exploration Geophysicists.

Berkhout, Augustinus Johannes. 2012. *Seismic Migration: Imaging of Acoustic Energy by Wave Field Extrapolation: Imaging of Acoustic Energy by Wave Field Extrapolation*. Elsevier.

Biondi, Biondo L. 2006. *3D seismic imaging*. Society of Exploration Geophysicists.

Bishop, Christopher M, & Nasrabadi, Nasser M. 2006. *Pattern recognition and machine learning*. Vol. 4. Springer.

Bleistein, Norman, Cohen, Jack K, Stockwell Jr, John W, & Berryman, JG. 2001. Mathematics of multidimensional seismic imaging, migration, and inversion. interdisciplinary applied mathematics, vol 13. *Appl. Mech. Rev.*, **54**(5), B94–B96.

Bradley, Andrew. 2019. *PDE-constrained optimization and the adjoint method*. https://cs.stanford.edu/~ambrad/adjoint_tutorial.pdf. Accessed: September 2, 2023.

Cervenỳ, Vlastislav. 2001. *Seismic ray theory*. Vol. 110. Cambridge university press Cambridge.

Claerbout, Jon. 1992. *Earth soundings analysis: Processing versus Inversion*. Vol. 6. Blackwell Scientific Publications London.

Claerbout, Jon. 2014. *Geophysical image estimation by example*. Lulu.com.

Claerbout, Jon F. 1971. Toward a unified theory of reflector mapping. *Geophysics*, **36**(3), 467–481.

Claerbout, Jon F. 1985. *Imaging the Earth's Interior*. Vol. 1. Blackwell scientific publications, Oxford.

Dai, Wei, & Schuster, Gerard T. 2013. Plane-wave least-squares reverse-time migration. *Geophysics*, **78**(4), S165–S177.

Davydenko, Mikhail, & Verschuur, DJ. 2017. Full-wavefield migration: using surface and internal multiples in imaging. *Geophysical Prospecting*, **65**(1), 7–21.

Davydenko, Mikhail, & Verschuur, Eric. 2021. Full wavefield least-squares reverse time migration. *Geophysics*, **86**(5), WC67–WC74.

Demanet, Laurent. 2021. *Waves and Imaging class notes*. https://math.mit.edu/icg/resources/teaching/18.367/notes367.pdf. Accessed: October 1, 2023.

Dutta, Gaurav, Huang, Hui, Kanakamedala, Karthik, Deng, Bin, & Wang, Ping. 2019. Practical strategies for interbed multiple attenuation. *Page D033S088R004 of: SEG International Exposition and Annual Meeting*. SEG.

Dwivedi, Sandeep Kumar, Vishwakarma, Manish, & Soni, Akhilesh. 2018. Advances and researches on non destructive testing: A review. *Materials Today: Proceedings*, **5**(2), 3690–3698.

Hu, Jianxing, Schuster, Gerard T, & Valasek, Paul A. 2001. Poststack migration deconvolution. *Geophysics*, **66**(3), 939–952.

Hussain, Shah, Mubeen, Iqra, Ullah, Niamat, Shah, Syed Shahab Ud Din, Khan, Bakhtawar Abduljalil, Zahoor, Muhammad, Ullah, Riaz, Khan, Farhat Ali, & Sultan, Mujeeb A. 2022. Modern diagnostic imaging technique applications and risk factors in the medical field: A review. *BioMed Research International*, **2022**.

Jannane, M, Beydoun, W, Crase, E, Cao, D, Koren, Z, Landa, E, Mendes, M, Pica, A, Noble, M, Roeth, G, *et al.* 1989. Wavelengths of earth structures that can be resolved from seismic reflection data. *Geophysics*, **54**(7), 906–910.

Kosloff, Dan D, & Baysal, Edip. 1982. Forward modeling by a Fourier method. *Geophysics*, **47**(10), 1402–1412.

Liberzon, Daniel. 2011. *Calculus of variations and optimal control theory: a concise introduction*. Princeton University Press.

Lu, Shaoping, Whitmore, Dan N, Valenciano, Alejandro A, & Chemingui, Nizar. 2015. Separated-wavefield imaging using primary and multiple energy. *The Leading Edge*, **34**(7), 770–778.

Lu, Shaoping, Liu, Faqi, Chemingui, Nizar, Valenciano, Alejandro, & Long, Andrew. 2018a. Least-squares full-wavefield migration. *The Leading Edge*, **37**(1), 46–51.

Lu, Shaoping, Liu, Faqi, Chemingui, Nizar, Valenciano, Alejandro, & Long, Andrew. 2018b. Least-squares full-wavefield migration. *The Leading Edge*, **37**(1), 46–51.

Lu, Shaoping, Qiu, Lingyun, & Li, Xiang. 2021. Addressing the crosstalk issue in imaging using seismic multiple wavefields. *Geophysics*, **86**(3), S235–S245.

Luo, Simon, & Hale, Dave. 2014. Least-squares migration in the presence of velocity errors. *Geophysics*, **79**(4), S153–S161.

McLeman, James, Burgess, Tim, Sinha, Mrinal, Hampson, Gary, & Thompson, Troy. 2021. Reflection FWI with an augmented wave equation and quasi-Newton adaptive gradient scheme. *Pages 667–671 of: First International Meeting for Applied Geoscience & Energy*. Society of Exploration Geophysicists.

Mulder, Wim A, & Plessix, R-E. 2004. A comparison between one-way and two-way wave-equation migration. *Geophysics*, **69**(6), 1491–1504.

Nemeth, Tamas, Wu, Chengjun, & Schuster, Gerard T. 1999. Least-squares migration of incomplete reflection data. *Geophysics*, **64**(1), 208–221.

Nocedal, Jorge, & Wright, Stephen J. 1999. *Numerical optimization*. Springer.

Oldenburg, Douglas W, & Li, Yaoguo. 2005. Inversion for applied geophysics: A tutorial. *Near-surface geophysics*, 89–150.

Operto, Stéphane. 2023. *Marmousi model*. https://www.geoazur.fr/WIND/bin/view/Main/Data/Marmousi. Accessed: May 1, 2023.

Plessix, R-E. 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, **167**(2), 495–503.

Sava, Paul. 2014. *A comparative review of wavefield tomography methods*. https://cwp.mines.edu/wp-content/uploads/sites/112/2018/08/cwp802.pdf. Accessed: August 10, 2023.

Sava, Paul, & Hill, Stephen J. 2009. Overview and classification of wavefield seismic imaging methods. *The Leading Edge*, **28**(2), 170–183.

Scales, John A. 1995. *Theory of Seismic Imaging*. Vol. 2. Springer Berlin, Heidelberg.

Schuster, Gerard T. 2017. *Seismic Inversion*. Society of Exploration Geophysicists.

Sheriff, Robert E, & Geldart, Lloyd P. 1995. *Exploration Seismology*. Cambridge University Press.

Shuey, RT. 1985. A simplification of the Zoeppritz equations. *Geophysics*, **50**(4), 609–614.

Soubaras, Robert. 2010. Deghosting by joint deconvolution of a migration and a mirror migration. *Pages SEG–2010 of: SEG International Exposition and Annual Meeting*. SEG.

Stolt, Robert H.., & Weglein, Arthur B.. 2012. *Seismic Imaging and Inversion: Application of Linear Inverse Theory*. Cambridge University Press.

Strang, G. 2007. *Computational Science and Engineering*. Wellesley-Cambridge Press.

Talagrand, Olivier, & Courtier, Philippe. 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quarterly Journal of the Royal Meteorological Society*, **113**(478), 1311–1328.

Tarantola, Albert. 1984. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, **49**(8), 1259–1266.

Tarantola, Albert. 1986. A strategy for nonlinear elastic inversion of seismic reflection data. *Geophysics*, **51**(10), 1893–1903.

Valenciano, Alejandro A, Biondi, Biondo L, & Clapp, Robert G. 2009. Imaging by target-oriented wave-equation inversion. *Geophysics*, **74**(6), WCA109–WCA120.

van Herwaarden, Dirk Philip, Boehm, Christian, Afanasiev, Michael, Thrastarson, Solvi, Krischer, Lion, Trampert, Jeannot, & Fichtner, Andreas. 2020. Accelerated full-waveform inversion using dynamic mini-batches. *Geophysical Journal International*, **221**(2), 1427–1438.

Verschuur, Dirk J, Berkhout, AJ, & Wapenaar, CPA. 1992. Adaptive surface-related multiple elimination. *Geophysics*, **57**(9), 1166–1177.

Wang, Ping, Huang, Shouting, & Wang, Ming. 2017. Least-squares RTM: Theory and applications. *Pages 1225–1229 of: 15th International Congress of the Brazilian Geophysical Society & EXPOGEF, Rio de Janeiro, Brazil, 31 July-3 August 2017*. Brazilian Geophysical Society.

Wang, Ping, Zhang, Zhigang, Mei, Jiawei, Lin, Feng, & Huang, Rongxin. 2019. Full-waveform inversion for salt: A coming of age. *The Leading Edge*, **38**(3), 204–213.

Weglein, Arthur B, Gasparotto, Fernanda Araújo, Carvalho, Paulo M, & Stolt, Robert H. 1997. An inverse-scattering series method for attenuating multiples in seismic reflection data. *Geophysics*, **62**(6), 1975–1989.

Whitmore, ND, Valenciano, AA, Sollner, Walter, & Lu, Shaoping. 2010. Imaging of primaries and multiples using a dual-sensor towed streamer. *Pages 3187–3192 of: SEG Technical Program Expanded Abstracts 2010*. Society of Exploration Geophysicists.

Whitmore, ND, Ramos-Martinez, J, Yang, Y, & Valenciano, A. 2020. Full wavefield modeling with vector reflectivity. *Pages 1–5 of: EAGE 2020 Annual Conference & Exhibition Online*, vol. 2020. European Association of Geoscientists & Engineers.

Wu, Han, & Lu, Shaoping. 2023. Nonlinear least-squares reverse time migration using a modeling engine with vector reflectivity. *In: International Meeting for Applied Geoscience & Energy*.

Wu, Zedong, & Alkhalifah, Tariq. 2017. Efficient scattering-angle enrichment for a nonlinear inversion of the background and perturbations components of a velocity model. *Geophysical Journal International*, **210**(3), 1981–1992.

Xia, Peipei, Zhang, Li, & Li, Fanzhang. 2015. Learning similarity with cosine similarity ensemble. *Information sciences*, **307**, 39–52.

Yang, Y, Ramos-Martinez, J, Whitmore, D, Valenciano, A, & Chemingui, N. 2020. Full waveform inversion using wave equation reflectivity modeling. *Pages 1–5 of: EAGE 2020 Annual Conference & Exhibition Online*, vol. 2020. European Association of Geoscientists & Engineers.

Yang, Yang, Ramos-Martinez, Jaime, Whitmore, Dan, Huang, Guanghui, & Chemingui, Nizar. 2021. Simultaneous inversion of velocity and reflectivity. *First Break*, **39**(12), 55–59.

Yao, G, Warner, M, & Silverton, A. 2014. Reflection FWI for both reflectivity and background velocity. *Pages 1–5 of: 76th EAGE Conference and Exhibition 2014*, vol. 2014. European Association of Geoscientists & Engineers.

Yao, Gang, Wu, Di, & Wang, Shang-Xu. 2020. A review on reflection-waveform inversion. *Petroleum Science*, **17**, 334–351.

Yilmaz, Öz. 2001. *Seismic data analysis: Processing, inversion, and interpretation of seismic data*. Society of Exploration Geophysicists.

Yu, J, & Schuster, T. 2003. Migration deconvolution vs. least squares migration: Presented at the 74th Annual International Meeting, SEG. *In: Expanded Abstracts*.

Zhang, Yu, Duan, Lian, & Xie, Yi. 2015. A stable and practical implementation of least-squares reverse time migration. *Geophysics*, **80**(1), V23–V31.