Elastic Least-squares Reverse Time Migration and Elastic Gauss-Newton Full-waveform Inversion

by

Ke Chen

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Geophysics

Department of Physics

University of Alberta

Abstract

With the fast development of high-performance computing resources, imaging and inversion techniques in the exploration geophysics community are moving from simplified methods to more complex methods that honour as far as possible the physics of wave propagation. Multiparameter imaging and inversion based on two-way wave equation operators are becoming viable for estimating subsurface structures and media properties.

In this thesis, I present new time-domain methods for linearized and nonlinear inversion of elastic wavefields. First, I derive the elastic Born and reverse time migration (RTM) operators for the first-order velocity-stress isotropic elastic wave-equation system. I develop an elastic least-squares reverse-time migration (LSRTM) method with the elastic Born and RTM as forward and adjoint operators. I adopt the conjugate gradient least-squares (CGLS) algorithm to solve the least-squares optimization problem that only requires the action of forward (elastic Born) and adjoint operator (elastic RTM) applied "on the fly" to vectors. In this case, the Hessian operator of the problem is implicitly inverted via a matrix-free algorithm. The proposed elastic LSRTM can suppress the ubiquitous multiparameter crosstalk artifacts that arise in seismic elastic inversion.

I then draw connections between waveform linearized imaging and full-waveform inversion techniques. I point out that the elastic LSRTM can be viewed as one iteration of the elastic Gauss-Newton full-waveform inversion (FWI) algorithm. This framework offers considerable freedom to design and apply Gauss-Newton FWI algorithms to elastic wavefields. I develop a matrix-free elastic Gauss-Newton FWI method based on the elastic LSRTM code. It consists of two loops of iterations: outer Gauss-Newton nonlinear iterations and inner CGLS linear

iterations. The Hessian in each Gauss-Newton iteration is inverted in a matrix-free form that only requires the forward (elastic Born), and adjoint Fréchet derivative operator (elastic RTM) applied to vectors.

Numerical tests are utilized to demonstrate the ability of the proposed inversion techniques to effectively perform multiparameter seismic inversion in both the linear and non-linear regimes.

Preface

Chapter 3 of this thesis has been published as K. Chen and M. D. Sacchi, 2017, Elastic least-squares reverse time migration via linearized elastic full-waveform inversion with pseudo-Hessian preconditioning: Geophysics, 82, no. 5, S341-S358. I was responsible for designing the algorithms, programming the code, generating the numerical examples and writing the manuscript. Dr. M. Sacchi was the supervisory author and was involved with concept formulation and manuscript editing.

To my families and my teachers

Acknowledgements

I owe great thanks to Prof. Mauricio D. Sacchi for taking me on as a student and supervising my PhD study. I am always amazed by his ability to break down complex problems to concise concepts. His patience and encouragement are important for the completion of this thesis work. I would like to thank my thesis committee, Prof. Douglas R. Schmitt, Prof. Mirko van der Baan, Prof. Lawrence Le, Prof. Edmond Lou, and Prof. Alison Malcolm for insightful comments and suggestions to improve the thesis. I also thank all my colleagues at the Signal Analysis and Imaging Group (SAIG) for their inspiring discussions on research. I want to thank Dr. Edward Jenner, Dr. Christof Stork, Dr. John Novak, Dr. Greg Wimpey and Dr. Joseph Jennings for an internship at ION Geophysical in Denver in the summer of 2014; and Dr. Sven Treitel, Dr. Timothy Keho, Dr. Weichang Li, Dr. Houzhu Zhang, Dr. Youli Quan, Dr. Roy Burnstad, Dr. Yang Zhao, Dr. Hong Liang, Dr. Baek Hyoungsu at Saudi Aramco Houston Research Centre for an internship from Nov. 2015 to Jan. 2016. Finally, I thank my families for their encouragement and support during the journey of PhD study.

Contents

List of Figures

CHAPTER 1

Introduction

Seismic imaging and inversion methods infer subsurface information from observed seismic data on the surface of the earth. Due to the rapid advancement of computational power, seismic imaging and inversion techniques for resource exploration have seen a fast development from *migration* to *least-squares migration* (LSM) and *full-waveform inversion* (FWI). Two-way wave equation operators are more and more widely used for the development of imaging and inversion techniques. Researchers are also using more complex physical principles to approximate seismic wave propagation in realistic earth models. The latter entails, for instance, including elastic, viscoelastic and anisotropic modelling techniques. In this chapter, I will review the development of LSM and FWI techniques and introduce the language often utilized in seismic imaging.

1.1 Review of least-squares migration

Least-squares migration refers to methods that image the subsurface by solving a linear inverse problem. The Born approximation is adopted to derive forward and adjoint operators that are utilized by least-squares migration techniques. A numerical procedure for the inversion of a large linear system of equations is then used to retrieve subsurface images that honour preprocessed seismic data. The long-wavelength content of the subsurface model is assumed known. Least-squares migration methods are designed for estimating the short-wavelength content of the subsurface model.

Initial efforts in the area of least-squares migration adopted Kirchhoff operators to develop a family of methods referred as least-squares Kirchhoff migration (LSKM) (Keys and Weglein, 1983; Tarantola, 1984b). If the background medium is assumed to be homogenous, analytical Green's functions can be computed to design the forward operator and

the adjoint operator required by least-squares migration (Keys and Weglein, 1983; Tarantola, 1984b). Lambare et al. (1992), Schuster (1993) and Nemeth et al. (1999) utilize ray tracing techniques to numerically compute Green's functions that allows the application of LSKM to heterogeneous media. One-way wave equation propagators (Claerbout, 1970, 1971) lead to the development of least-squares one-way wave-equation migration (LSWEM) (Rickett, 2003). In Rickett (2003), the one-way wave-equation is solved using a Fourier finite-difference method (Tang, 2009). Recently, Lu et al. (2017) apply LSWEM on broadband data for high-resolution imaging. Also based on one-way wave-equation, but in the frequency domain, Kuehl and Sacchi (2003) propose a double-square-root least-squares split-step migration (LS-SSM). The latter is extended to 3D and applied to field data (Wang et al., 2005; Wang and Sacchi, 2007). Least-squares migration methods that adopt the two-way wave equation has lead to the so called least-squares reverse time migration (LSRTM) method. This technique has been initially proposed under the name "linearized inversion" by Bourgeois et al. (1989). In Bourgeois et al. (1989), the two-way wave-equation is solved using a time-domain finite-difference (FD) method and the P-wave impedance perturbation is the unknown of the problem. Ostmo et al. (2002) implement a LSRTM in frequency domain under the name of "linearized waveform inversion". In recent years, LSRTM has been revisited, and it has attracted considerable interest in both industrial and academic research circles (Ji, 2009; Dai et al., 2012; Duan et al., 2014; Wong et al., 2015; Chen et al., 2017; Xu and Sacchi, 2018). The fast development of computer hardware provides opportunities for computational-demanding techniques such as LSRTM. The advantage of LSRTM over other LSM methods is that it can handle complex geological structures. The drawback of LSRTM is its high computational cost.

A limited number of LSM studies have considered the elastic nature of seismic wave propagation in the subsurface. Beydoun and Mendes (1989) and Jin et al. (1992) implement an elastic least-squares Kirchhoff migration for multicomponent seismic data. Tura and Johnson (1993) propose an elastic least-squares migration in f-k domain for simple homogeneous background medium. Anikiev et al. (2013) investigate the decoupling of parameters for frequency-domain elastic LSRTM in a homogeneous model. Stanton and Sacchi (2015) utilize Helmholtz decomposition for elastic least-squares split-step migration. In Chapter 3 and 4 of this thesis, I propose elastic least-squares reverse time migration algorithms based on elastic Gauss-Newton full-waveform inversion kernels.

1.2 Review of full-waveform inversion

Full-waveform inversion consists of estimating subsurface Earth's parameters from observed seismic data on the surface of the earth (seismic reflection data) or in the borehole (VSP[1] data). The goal of FWI is ambitious because it does not assume a scale separation of the subsurface model. It aims at retrieving both the long wavelengths and short wavelengths of the model. FWI is a nonlinear inverse problem because observed waveforms are nonlinear functions of the earth model. Global stochastic optimization methods are far too expensive to solve practical imaging problems. Then, the nonlinear inverse problem is often solved by iterative methods. A smooth background model is needed to initiate the inversion (Virieux and Operto, 2009).

The mathematical basis of FWI can be traced back to the adjoint-state theory developed in the book "Optimal control of systems governed by partial differential equations" by French mathematician Jacques-Louis Lions (Lions, 1971). The development of FWI initiates at the beginning of the 1980s. Bamberger et al. (1979, 1982) apply the adjoint-state method to invert for the coefficients of the 1D acoustic wave equation from only boundary measurements. Essentially, it is a 1D full-waveform inversion problem. Tarantola and Valette (1982) define the generalized nonlinear inverse problem using the least-squares criterion. The latter is an extension of the linear inverse theory of Backus and Gilbert (1970) and Jackson (1979). Lailly (1983) and Tarantola (1984a) provide general formulations of time-domain acoustic FWI. The nonlinear inversion resembles an iterative prestack migration algorithm. Tarantola (1986) extends the FWI theory to an isotropic elastic earth model.

Gauthier et al. (1986) and Mora (1987b, 1988) numerically study the 2D acoustic FWI and isotropic elastic FWI using synthetic data. Pica et al. (1990) apply a 2D acoustic FWI on real data. The medium is simplified as a 1D layered model. Crase et al. (1990, 1992) apply 2D elastic FWI to real marine and land seismic reflection data. Time-domain FWI utilizes time-domain forward modelling to simulate the wavefield in each nonlinear waveform inversion iteration. The time-domain modelling technique often uses the explicit time-stepping scheme to solve the wave equation iteratively. FWI is computationally intensive. Even though using a CRAY supercomputer, the early studies of FWI utilize only a few numbers of shot gathers or a few numbers of iterations for the nonlinear inversion.

In the 1990s, the FWI is reformulated in frequency domain (Pratt et al., 1996). These techniques use frequency-domain forward modelling as an engine for nonlinear inversion. Instead of an explicit time-stepping scheme, the frequency-domain modelling formulates the Green's function in a matrix form and inverts the matrix using a direct method such as the

[1] Vertical Seismic Profile

LU decomposition. The frequency-domain modelling belongs to an implicit method for solving the wave equation. The frequency-domain FWI has some advantages over time-domain methods. First, in the frequency-domain forward solver, the LU decomposed matrices can be saved and reused for simulating different shot gathers. While, the time-domain forward solver never saves the Green's function in memory. The time-domain wave equation needs to be repeatedly solved for each shot. Second, only a few discrete frequencies of the source and data are needed to perform full-waveform inversion in the frequency domain. The latter is feasible given the wavenumber illumination redundancy that exists in wide-aperture seismic data (Sirgue and Pratt, 2004). Frequency-domain methods greatly reduce the computational cost of FWI. For these reasons, the 2D frequency-domain FWI has seen a fast development in the last 20 years (Pratt, 1999; Pratt and Shipp, 1999; Ravaut et al., 2004; Operto et al., 2006; Brenders and Pratt, 2007). 2D Frequency-domain isotropic elastic FWI has also been investigated (Gelis et al., 2007; Brossier et al., 2009b). With the development of high-speed computer hardware, 3D frequency-domain acoustic FWI is becoming an area of fast development as well (Ben-Hadj-Ali et al., 2008; Operto et al., 2015). However, the frequency-domain direct solver for Helmholtz equation becomes the principal bottleneck for frequency-domain 3D inversion because of its memory requirements.

At the same time, time-domain acoustic FWI has been extended to three-dimension (Vigh and Starr, 2008; Schiemenz and Igel, 2013). Recent years, 3D time-domain elastic FWI has been implemented on modern supercomputers, and they appear to be the method of choice for the development of practical algorithms for FWI (Epanomeritakis et al., 2008; Guasch et al., 2012; Vigh et al., 2014; Borisov and Singh, 2015; Albertin et al., 2016; Borisov et al., 2018; Oh et al., 2018). Besides the advantage with respect to memory, the time-domain method allows easier incorporation of elastic or anisotropic parameters in the inversion than frequency-domain method. Time-domain methods are becoming the trend for future development of FWI.

Because of its ability to estimate subsurface model, FWI has also been applied for time-lapse reservoir monitoring (Routh et al., 2012; Raknes and Arntsen, 2014; Yang et al., 2016a; Alemie and Sacchi, 2016; Kotsi and Malcolm, 2017; Kamei and Lumley, 2017).

When it was first proposed, FWI was viewed as an optimal control problem of a system governed by partial differential equations (PDEs) (Bamberger et al., 1979; Lions, 1971). In FWI problem, the PDEs are the elastic wave equations. In other words, FWI is a PDE-constrained optimization problem (Borzi and Schulz, 2012). Another example of PDE-constrained optimization problem is the optimal control of Navier-Stokes equations in fluid dynamics.

1.3 Full-waveform inversion: Challenges

In this section I enumerate the main challenges of FWI and current research that aims to combat them.

- **Computational cost**: FWI is a computationally demanding technique. It requires solving the numerical solution of the wave equation multiple times in each nonlinear iteration. The forward solver can be implemented via the finite-difference method, the finite-element method or via the spectral-element method. The forward simulation of the wave equation itself is expensive. FWI is normally performed on high-performance computing infrastructures. The computational cost can be partially reduced by data contraction strategies such as simultaneous source encoding (Mora, 1987b; Krebs et al., 2009; Ben-Hadj-Ali et al., 2011; Anagaw and Sacchi, 2014; Matharu and Sacchi, 2017), plane-wave synthesis (Vigh and Starr, 2008; Kwon et al., 2015) and stochastic source subsampling (van Leeuwen and Herrmann, 2013).
- **Convergence to local minima**: The cost function adopted in FWI suffers from spurious local minima. Optimization techniques used to solve the FWI problem could converge to a local minimum solution. The latter may happen when the modelled and observed seismograms have a time delay greater than half a period of the dominant frequency of the seismic source. In FWI, this phenomenon is called "cycle-skipping". The initial model of the iterative algorithm should be good enough to avoid the cycle-skipping problem. Several techniques can alleviate the problem above:
  - *Multiscale frequency continuation methods* (Kolb et al., 1986; Bunks et al., 1995). The nonlinearity of the cost function is more evident for high temporal frequencies than for low frequencies. In other words, the FWI objective function behaves like a quadratic function at lower frequencies. The frequency continuation entails solving the FWI from low to high frequencies to avoid convergence to local minima. The technique works only when the observed data contain sufficient low-frequency content to initialize the continuation process.
  - *FWI with a modified measure of data fit.* In this case, rather than using the classical least-squares misfit function to measure the difference between observed and modelled seismograms, one can adopt other measures of similarity such as the crosscorrelation (Luo and Schuster, 1991; Dahlen et al., 2000; Tromp et al., 2005; Luo et al., 2016), deltaness of a matching filter (Luo and Sava, 2011; Warner and Guasch, 2016), difference of envelopes (Mora, 1987b; Fichtner et al., 2008; Bozdag et al., 2011; Wu et al., 2014), and cost measures inspired on the optimal transport distance (Metivier et al., 2016; Yang et al., 2018).

  - *Reflection waveform inversion (RWI).* FWI updates the long and short wavelength components of the model simultaneously in the nonlinear iterations. The RWI approach assumes scale separation between a smooth velocity macromodel and a localized reflectivity model. RWI updates long and short wavelength components of the model alternatively in the nonlinear iterations (Mora, 1989; Chavent et al., 1994; Clement et al., 2001; Xu et al., 2012; Brossier et al., 2015; Zhou et al., 2015). The RWI uses imaged reflectors that act as secondary sources to produce transmission responses ("rabbit ears") which are used to update the macromodel.
  - *Extended waveform inversion (EWI).* In EWI, the model space is extended by a non-physical axis to obtain an objective function that approximate a quadratic cost. The model extension axis can be subsurface offset (Symes, 2008; Fu and Symes, 2017) or a time lag (Biondi and Almomin, 2014).

- **Multiparameter coupling**: Resolving more than one parameter such as compressional and shear-wave velocities and density is a major problem in FWI. Parameters are coupled and therefore, the data can be fitted by models where final multiparameter images are contaminated by parameter *cross-talk* artifacts. Besides, different classes of parameters have different magnitude orders or physical units. They have different strength of responses in the seismic wavefield that makes the inversion ill-conditioned. Through diffraction pattern analysis of different parameterizations, hierarchical inversion strategies were proposed to mitigate parameter cross-talk contamination (Tarantola, 1986; Sears et al., 2008). Alternatively, preconditioning operators can be applied in the multiparameter inversion to reduce parameter cross-talk. In this thesis, I investigate the role of the Hessian operator in decoupling the elastic parameters.

1.4 Isotropic elastic parameter sensitivity analysis

The diffraction pattern or radiation pattern indicates the distribution of amplitude of scattered energy as a function of diffraction angle (scattering angle). It is essentially the derivative of the wave equation with respect to model parameters. The diffraction patterns for different parameters provide an insight on the cause of the multiparameter crosstalk problem. Under the ray + Born approximation, the analytical expressions for the elastic diffraction patterns of the Lamé parameters, and density perturbations $[\delta\lambda/\lambda, \delta\mu/\mu, \delta\rho/\rho]$ are given by

(Wu and Aki, 1985; Beylkin and Burridge, 1990; Forgues and Lambare, 1997)

$$\begin{aligned}\mathbf{W}_{\lambda,\mu,\rho} &= \begin{pmatrix} W_\lambda^{P\text{-}P} & W_\mu^{P\text{-}P} & W_\rho^{P\text{-}P} \\ W_\lambda^{P\text{-}SV} & W_\mu^{P\text{-}SV} & W_\rho^{P\text{-}SV} \\ W_\lambda^{SV\text{-}P} & W_\mu^{SV\text{-}P} & W_\rho^{SV\text{-}P} \\ W_\lambda^{SV\text{-}SV} & W_\mu^{SV\text{-}SV} & W_\rho^{SV\text{-}SV} \end{pmatrix} \\ &= \begin{pmatrix} (V_p^2 - 2V_s^2)/V_p^2 & (2V_s^2/V_p^2)\cos^2\theta^{P\text{-}P} & \cos\theta^{P\text{-}P} \\ 0 & (V_s/V_p)\sin 2\theta^{P\text{-}SV} & \sin\theta^{P\text{-}SV} \\ 0 & -(V_s/V_p)\sin 2\theta^{SV\text{-}P} & -\sin\theta^{SV\text{-}P} \\ 0 & \cos 2\theta^{SV\text{-}SV} & \cos\theta^{SV\text{-}SV} \end{pmatrix},\end{aligned} \tag{1.1}$$

where $W_\lambda^{P\text{-}P}$ represents the P- to P-wave diffraction pattern due to Lamé parameter perturbation $(\delta\lambda/\lambda)$, $\theta^{P\text{-}P}$ is the diffraction angle between the incident P-wave and scattered P-wave, V_p and V_s are the background P- and S-wave velocities. The analytical expressions for the diffraction patterns of the P- and S-wave impedances, and density perturbations $[\delta I_p/I_p, \delta I_s/I_s, \delta\rho/\rho]$ are given by (Wu and Aki, 1985; Beylkin and Burridge, 1990; Forgues and Lambare, 1997)

$$\begin{aligned}\mathbf{W}_{I_p,I_s,\rho} &= \begin{pmatrix} W_{I_p}^{P\text{-}P} & W_{I_s}^{P\text{-}P} & W_\rho^{P\text{-}P} \\ W_{I_p}^{P\text{-}SV} & W_{I_s}^{P\text{-}SV} & W_\rho^{P\text{-}SV} \\ W_{I_p}^{SV\text{-}P} & W_{I_s}^{SV\text{-}P} & W_\rho^{SV\text{-}P} \\ W_{I_p}^{SV\text{-}SV} & W_{I_s}^{SV\text{-}SV} & W_\rho^{SV\text{-}SV} \end{pmatrix} \\ &= \begin{pmatrix} 2 & -(4V_s^2/V_p^2)\sin^2\theta^{P\text{-}P} & (2V_s^2/V_p^2)\sin^2\theta^{P\text{-}P} - 2\sin^2(\theta^{P\text{-}P}/2) \\ 0 & (2V_s/V_p)\sin 2\theta^{P\text{-}SV} & -(V_s/V_p)\sin 2\theta^{P\text{-}SV} + \sin\theta^{P\text{-}SV} \\ 0 & -(2V_s/V_p)\sin 2\theta^{SV\text{-}P} & (V_s/V_p)\sin 2\theta^{SV\text{-}P} - \sin\theta^{SV\text{-}P} \\ 0 & 2\cos 2\theta^{SV\text{-}SV} & -\cos 2\theta^{SV\text{-}SV} + \cos\theta^{SV\text{-}SV} \end{pmatrix}.\end{aligned} \tag{1.2}$$

The analytical expressions for the diffraction patterns of the P- and S-wave velocities, and density perturbations $[\delta V_p/V_p, \delta V_s/V_s, \delta\rho/\rho]$ are given by (Wu and Aki, 1985; Beylkin and

Burridge, 1990; Forgues and Lambare, 1997)

$$\mathbf{W}_{V_p,V_s,\rho} = \begin{pmatrix} W_{V_p}^{P\text{-}P} & W_{V_s}^{P\text{-}P} & W_{\rho}^{P\text{-}P} \\ W_{V_p}^{P\text{-}SV} & W_{V_s}^{P\text{-}SV} & W_{\rho}^{P\text{-}SV} \\ W_{V_p}^{SV\text{-}P} & W_{V_s}^{SV\text{-}P} & W_{\rho}^{SV\text{-}P} \\ W_{V_p}^{SV\text{-}SV} & W_{V_s}^{SV\text{-}SV} & W_{\rho}^{SV\text{-}SV} \end{pmatrix}$$
$$= \begin{pmatrix} 2 & -(4V_s^2/V_p^2)\sin^2\theta^{P\text{-}P} & 2-(2V_s^2/V_p^2)\sin^2\theta^{P\text{-}P} - 2\sin^2(\theta^{P\text{-}P}/2) \\ 0 & (2V_s/V_p)\sin 2\theta^{P\text{-}SV} & (V_s/V_p)\sin 2\theta^{P\text{-}SV} + \sin\theta^{P\text{-}SV} \\ 0 & -(2V_s/V_p)\sin 2\theta^{SV\text{-}P} & -(V_s/V_p)\sin 2\theta^{SV\text{-}P} - \sin\theta^{SV\text{-}P} \\ 0 & 2\cos 2\theta^{SV\text{-}SV} & \cos 2\theta^{SV\text{-}SV} + \cos\theta^{SV\text{-}SV} \end{pmatrix}. \tag{1.3}$$

In Figures 1.1 - 1.3, I plot the theoretical amplitude diffraction patterns (equations 1.1 - 1.3) of model perturbations in homogenous media under the ray + Born approximation (Wu and Aki, 1985; Tarantola, 1986; Forgues and Lambare, 1997). The diffraction pattern for different parameterizations of perturbations is plotted in polar coordinates as a function of the diffraction angle. Diffraction patterns for different elastic parameters overlap. This overlap causes the so-called crosstalk artifacts in the inversion. In other words, the inability of the inversion algorithm to differentiate responses caused by different elastic parameters. If one adopts density and Lamé parameters to parameterize the inversion problem, the diffraction patterns become alike for moderate offsets (Figure 1.1). The P-P wave responses of the three perturbations are almost identical for small offsets. It indicates a strong coupling of the three parameters. Parameterization adopting density and Lamé parameters is not a good choice for both linear and nonlinear waveform inversion methods. If one adopts density, P- and S-wave impedances for parameterization, the diffraction patterns for a P-wave impedance perturbation and an S-wave impedance perturbation are very different (Figure 1.2). The density diffractor scatters waves forward. Therefore, it may be hard to resolve density using moderate-offset surface reflection data (Tarantola, 1986). Under density, P-and S-wave velocities parameterization, the diffraction patterns for P-wave velocity perturbation and S-wave velocity perturbation are also very different (Figure 1.3). While, the diffraction pattern of density diffractor shows ambiguity with the diffraction patterns of P- and S-wave velocity diffractors. The density may be difficult to resolve using only small offset P-P wave data (Tarantola, 1986). The diffraction pattern of P-wave velocity diffractor for P-P wave mode is homogenous at all angles. The P-wave velocity diffractor does not involve with other wave modes. Both short and long wavelengths of P-wave velocity can be reconstructed from wide-offset data. The diffraction pattern of S-wave velocity diffractor for P-P wave mode spans a range of intermediate scattering angles. The combination of the radiation patterns of the S-wave velocity diffractor for different wave modes spans a full range of scattering angles (Pageot et al., 2013). It indicates that both short and long wave-

lengths of S-wave velocity are able to be reconstructed using all the wave modes. However, the diffraction pattern of density diffractor has large amplitude at small scattering angle and small amplitude at large scattering angle. Only short-to-intermediate wavelengths of density can be retrieved from seismic data (Operto et al., 2013). From a physical point of view, the P- and S-wave velocity has influence on not only traveltimes but also amplitudes of waves, while the density mainly influences the amplitude of waves.

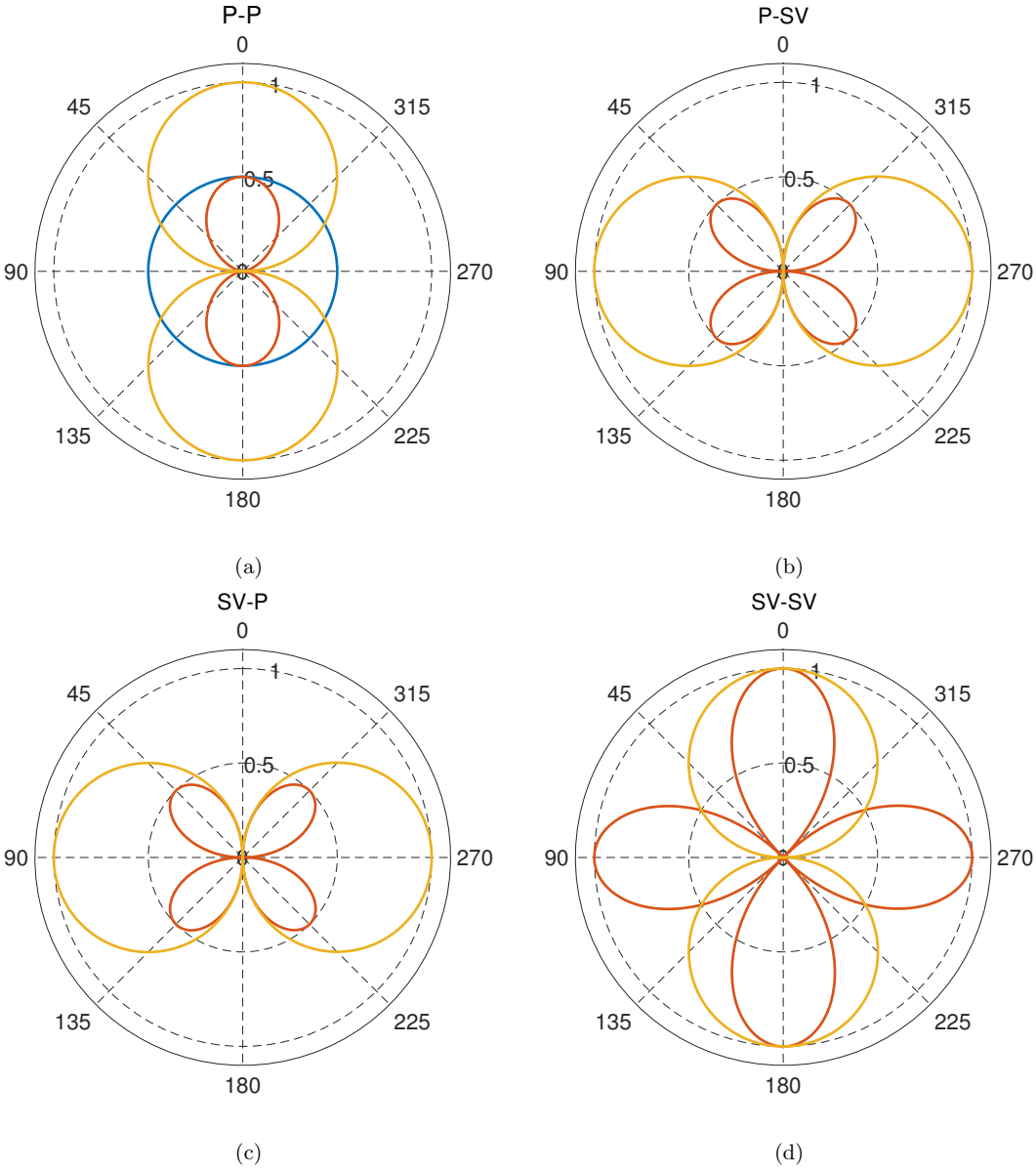


Figure 1.1: Elastic amplitude diffraction patterns for density and Lamé parameters parameterization. (a) P-P (Blue: $|W_{\lambda}^{P\text{-}P}|$, red: $|W_{\mu}^{P\text{-}P}|$, yellow: $|W_{\rho}^{P\text{-}P}|$). (b) P-SV (Red: $|W_{\mu}^{P\text{-}SV}|$, yellow: $|W_{\rho}^{P\text{-}SV}|$). (c) SV-P (Red: $|W_{\mu}^{SV\text{-}P}|$, yellow: $|W_{\rho}^{SV\text{-}P}|$). (d) SV-SV (Red: $|W_{\mu}^{SV\text{-}SV}|$, yellow: $|W_{\rho}^{SV\text{-}SV}|$). The diffraction pattern is plotted in polar coordinates as a function of diffraction angle.

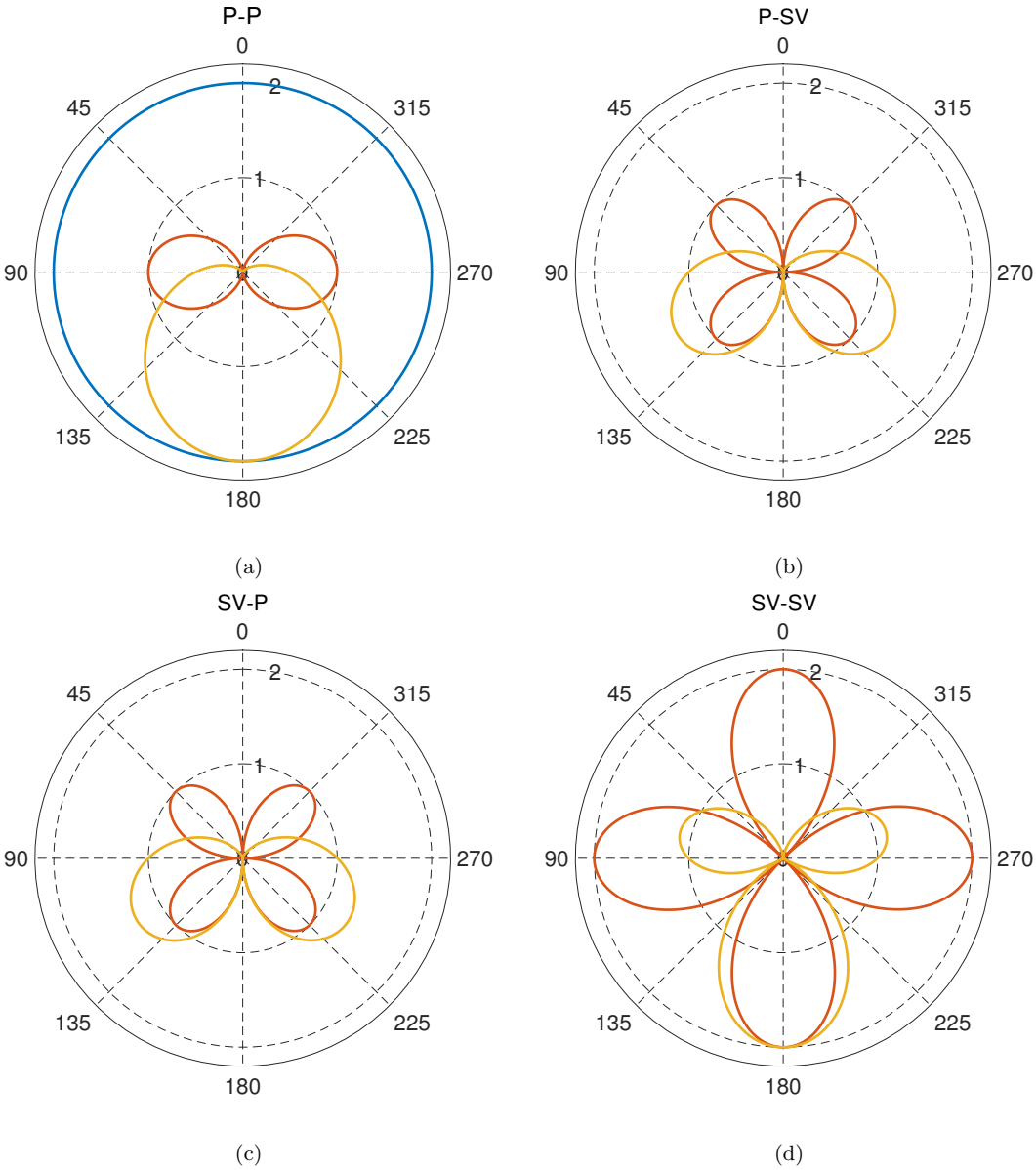


Figure 1.2: Elastic amplitude diffraction patterns for density, P- and S-wave impedances parameterization. (a) P-P (Blue: $|W_{I_p}^{P\text{-}P}|$, red: $|W_{I_s}^{P\text{-}P}|$, yellow: $|W_{\rho}^{P\text{-}P}|$). (b) P-SV (Red: $|W_{I_s}^{P\text{-}SV}|$, yellow: $|W_{\rho}^{P\text{-}SV}|$). (c) SV-P (Red: $|W_{I_s}^{SV\text{-}P}|$, yellow: $|W_{\rho}^{SV\text{-}P}|$). (d) SV-SV (Red: $|W_{I_s}^{SV\text{-}SV}|$, yellow: $|W_{\rho}^{SV\text{-}SV}|$). The diffraction pattern is plotted in polar coordinates as a function of diffraction angle.

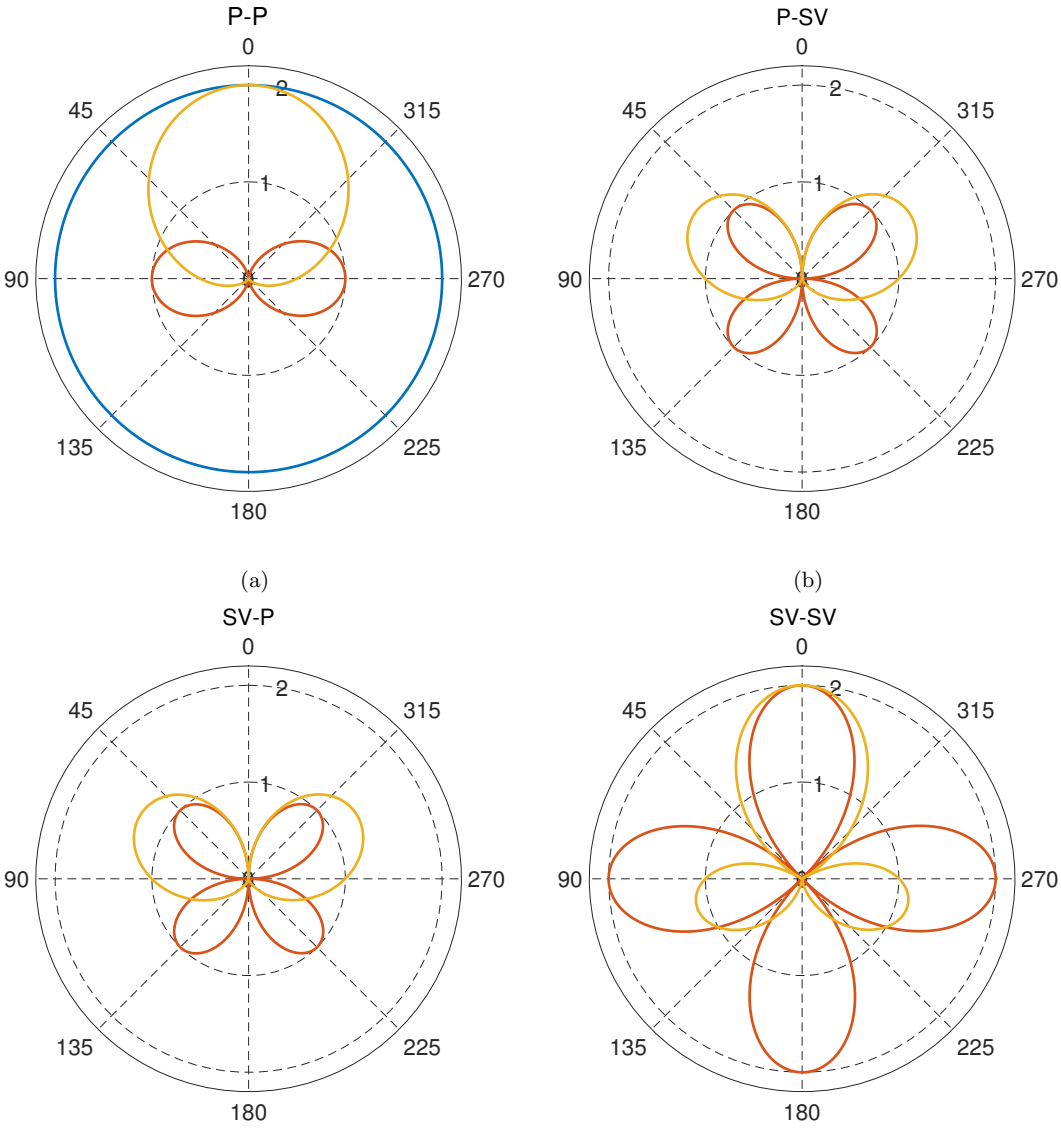


Figure 1.3: Elastic amplitude diffraction patterns for density, P- and S-wave velocities parameterization. (a) P-P (Blue: $|W_{V_p}^{P\text{-}P}|$, red: $|W_{V_s}^{P\text{-}P}|$, yellow: $|W_{\rho}^{P\text{-}P}|$). (b) P-SV (Red: $|W_{V_s}^{P\text{-}SV}|$, yellow: $|W_{\rho}^{P\text{-}SV}|$). (c) SV-P (Red: $|W_{V_s}^{SV\text{-}P}|$, yellow: $|W_{\rho}^{SV\text{-}P}|$). (d) SV-SV (Red: $|W_{V_s}^{SV\text{-}SV}|$, yellow: $|W_{\rho}^{SV\text{-}SV}|$). The diffraction pattern is plotted in polar coordinates as a function of diffraction angle.

1.5 Contributions of this thesis

This thesis investigates the role of the Hessian operator in both linearized and nonlinear elastic waveform inversion. In particular, I have examined theoretical aspects associated to the development of exact forward and adjoint operators for linearized elastic imaging and developed numerical solutions to solve both the linear and nonlinear elastic imaging problems. More specifically:

- I introduce a time domain elastic least-squares reverse time migration (LSRTM) algorithm. I formulate the elastic LSRTM as a linearized elastic full-waveform inversion problem. The proposed elastic LSRTM is able to reduce multiparameter cross-talk. Special attention was paid in deriving exact forward adjoint operators using the Born elastic approximation.
- I investigate the influence of density on crosstalk artifact and in the convergence of elastic LSRTM. For this purpose, I develop a three-parameter elastic LSRTM.
- I introduce a time-domain matrix-free elastic Gauss-Newton full-waveform inversion (FWI) algorithm. I point out that the elastic LSRTM can be viewed as one iteration of the elastic Gauss-Newton full-waveform inversion (FWI) algorithm. This framework offers great freedom to design and apply Gauss-Newton FWI algorithms to elastic wavefields. The proposed algorithm consists of two loops of iterations: the outer Gauss-Newton nonlinear iterations and the inner conjugate gradient least-squares (CGLS) linear iterations. The Hessian is implicitly inverted in a matrix-free fashion that only requires the forward and adjoint operators applied "on the fly" to vectors.

1.6 Organization of this thesis

- In **Chapter 2**, I draw connections between seismic wave forward modelling, Born approximation modelling, migration, least-squares migration and full-waveform inversion. I point out that the least-squares migration problem is equivalent to one iteration of Gauss-Newton full-waveform inversion problem.
- In **Chapter 3**, I propose a time-domain elastic LSRTM method. I derive the elastic Born approximation and elastic reverse time migration (RTM) operators using the time-domain continuous adjoint-state method. The state equation system used is the first-order velocity-stress isotropic elastic wave-equation system. I adopt the optimize-before-discretize (OBD) approach in which the operators and algorithms are derived

in continuous form and are discretized at the final stage. My discretized numerical versions of the elastic Born operator and its adjoint (elastic RTM operator) pass the dot-product test. I use the CGLS algorithm to solve the least-squares optimization problem. In other words, the Hessian operator for elastic LSRTM is implicitly inverted via a matrix-free algorithm that only requires the action of forward and adjoint operators on vectors. Numerical experiments show that the elastic LSRTM provides higher resolution images with fewer artifacts and a superior balance of amplitudes when compared to elastic RTM. The elastic LSRTM can reduce cross-talk artefacts between the P-wave and S-wave images.

- In **Chapter 4**, I investigate the influence of density on the generation of crosstalk and on the convergence of elastic LSRTM. I propose a time-domain three-parameter elastic LSRTM algorithm to simultaneously invert for density, P- and S-wave images. I then evaluate the proposed algorithm on two synthetic examples. The proposed three-parameter elastic LSRTM can suppress the multiparameter crosstalk among density, P- and S-wave images and it can correctly estimate the density perturbation image. Including the density image in the elastic LSRTM inversion also improves the convergence of least-squares inversion.

- In **Chapter 5**, I propose a time-domain matrix-free elastic Gauss-Newton FWI algorithm. I consider the elastic Gauss-Newton FWI as an iterative elastic LSRTM problem. The proposed algorithm consists of two loops of iterations: the outer Gauss-Newton nonlinear iterations and the inner CGLS linear iterations. The Gauss-Newton search direction in each outer FWI iteration is computed using the CGLS method. This step is equivalent to apply an elastic LSRTM on data residuals. The proposed algorithm is a matrix-free technique that only requires the forward and adjoint Fréchet derivative operators applied "on the fly" to vectors. The latter is achieved via the adjoint-state method. I use the proposed algorithm to simultaneously invert for P- and S-wave velocities. The proposed elastic Gauss-Newton FWI generates slightly better-inverted models than the nonlinear conjugate gradient (NLCG) method based elastic FWI. The elastic Gauss-Newton FWI converges faster than the elastic NLCG FWI. Because the proposed elastic Gauss-Newton FWI algorithm is matrix-free, its memory requirements are similar to those of the elastic NLCG FWI method.

CHAPTER 2

Migration, least-squares migration, and full-waveform inversion: Connections

In this chapter, I will draw the connection between migration and steepest-descent full-waveform inversion, and the connection between least-squares migration and Gauss-Newton full-waveform inversion.

2.1 Forward problem

The forward problem consists of predicting measurements using given the description of a physical system (Tarantola, 2005). Assuming that $\mathbf{m}$ represents the vector of parameters of the system and $\mathbf{d}$ the predicted measurements, the forward problem can be expressed as

$$\mathbf{d} = \boldsymbol{\phi}(\mathbf{m}), \tag{2.1}$$

where the operator $\boldsymbol{\phi}(\cdot)$ is called the forward operator. The forward problem can be descried by Figure 2.1.

The forward problem can be linear or nonlinear in its nature. If the forward problem is linear, equation 2.1 can be simplified as the following expression

$$\mathbf{d} = \boldsymbol{\Phi}\mathbf{m}, \tag{2.2}$$

where $\boldsymbol{\Phi}$ represents a general linear operator (or a matrix). In general, both linear and nonlinear forward problems are usually a well-posed problems.

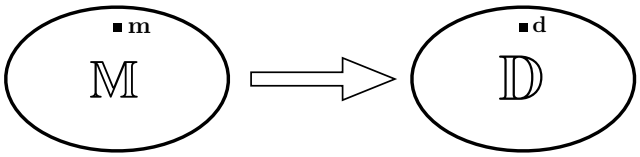


Figure 2.1: The sketch of general forward problem from model parameterized as $\mathbf{m}$ in model space $\mathbb{M}$ to the predicted measurement $\mathbf{d}$ in the data space $\mathbb{D}$.

2.1.1 Nonlinear forward problem

Examples of nonlinear forward problems include solving Schrödinger equation, Navier-Stokes equation and Eikonal equation. In this section, I will discuss the seismic wave simulation as an example which is a nonlinear problem in terms of the elastic parameters of the subsurface. The seismic wave equation is derived by considering a dynamical equation and a constitutive equation. The general heterogeneous elastic wave equation is defined as (Aki and Richards, 2002)

$$\begin{aligned} \rho(\mathbf{x})\dot{\mathbf{v}}(\mathbf{x},t) - \nabla\cdot\boldsymbol{\sigma}(\mathbf{x},t) &= \mathbf{f}(\mathbf{x},t), \\ \dot{\boldsymbol{\sigma}}(\mathbf{x},t) - \mathbf{C}(\mathbf{x}) : \nabla\mathbf{v}(\mathbf{x},t) &= \mathbf{0}, \\ \boldsymbol{\sigma}(\mathbf{x},t)\cdot\mathbf{n}(\mathbf{x})|_{\mathbf{x}\in\partial\Omega} &= 0, \\ \mathbf{v}(\mathbf{x},t)|_{t=0} &= \mathbf{0}, \\ \mathbf{x}\in\Omega\subset\Re^3, t\in\Re, & \end{aligned} \tag{2.3}$$

where $\mathbf{v}(\mathbf{x},t)$ is the particle velocity vector, $\boldsymbol{\sigma}(\mathbf{x},t)$ is the second-order stress tensor, $\rho(\mathbf{x})$ is mass density, $\mathbf{C}(\mathbf{x})$ is the fourth-order elastic tensor, Ω represents the computational domain and $\partial\Omega$ is the surface of the domain. The above wave equation fulfills a free surface boundary condition and a zero initial condition.

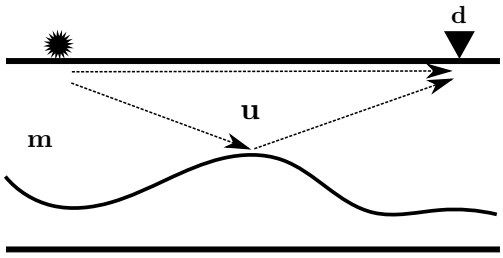


Figure 2.2: A sketch representing the seismic wave propagation problem $\mathbf{d} = \boldsymbol{\phi}(\mathbf{m})$. The star represents a source and triangles represent for receivers.

The seismic wave equation can be concisely written as

$$\mathbf{Su} = \mathbf{f}, \tag{2.4}$$

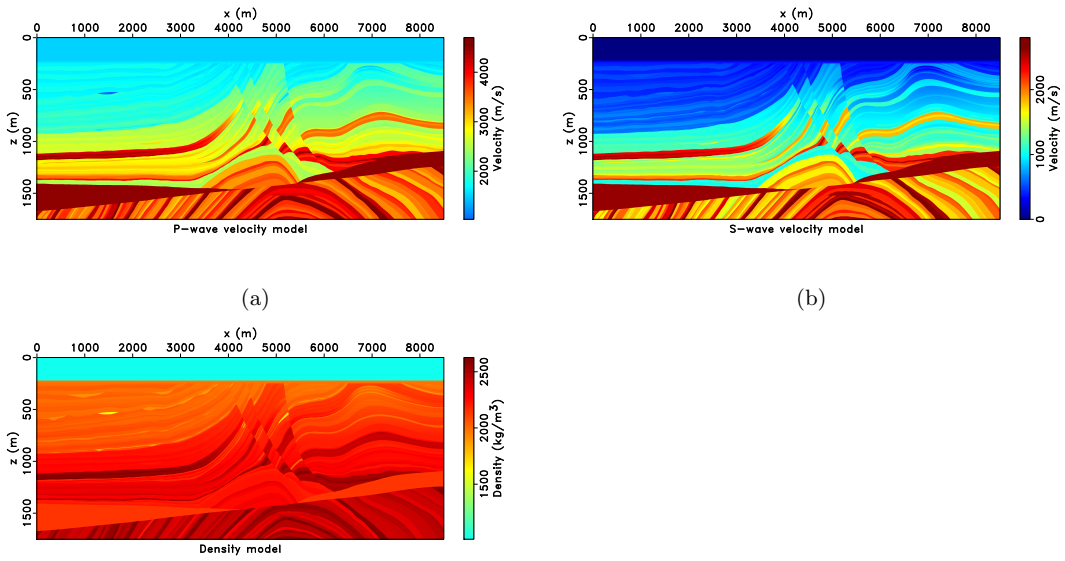
where $\mathbf{S} = \mathbf{S}(\mathbf{m})$ is the wave equation operator with $\mathbf{m}$ as the model parameter describing the subsurface, $\mathbf{u}$ is the wavefield and $\mathbf{f}$ is the source function. The solution of the wave equation 2.4 can be written using Green's operator

$$\mathbf{u} = \mathbf{S}^{-1}\mathbf{f} = \mathbf{Gf}, \tag{2.5}$$

where $\mathbf{G} = \mathbf{S}^{-1}$ is the Green's operator. The seismic data $\mathbf{d}$ are extracted at the receiver positions by the application of a sampling or restriction operator

$$\mathbf{d} = \phi(\mathbf{m}) = \mathbf{Ru}(\mathbf{m}) = \mathbf{RG}(\mathbf{m})\mathbf{f}. \tag{2.6}$$

The symbol $\mathbf{R}$ represents the sampling operator. Green's operator $\mathbf{G}$ is a nonlinear function of $\mathbf{m}$. Figure 2.2 shows the seismic wave propagation and data acquisition in a controlled source experiment.

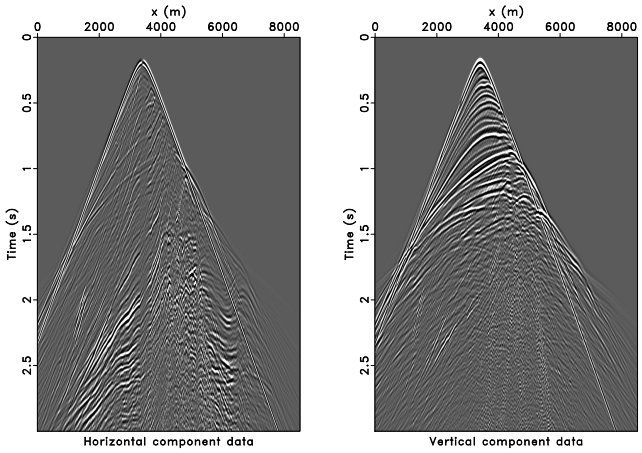


(a) (b)

(c)

Figure 2.3: The isotropic elastic Marmousi2 model. (a) P-wave velocity. (b) S-wave velocity. (c) Density.

For an heterogeneous subsurface model, there is no analytical solution for the Green's op-



(a) (b)

Figure 2.4: Simulated OBC seismic data for shot at lateral position $x = 3400$ m. (a) Horizontal particle velocity data. (b) Vertical particle velocity data. Simulation is done with a time-domain staggered-grid finite-difference scheme and an unsplit convolutional perfectly matched layer (C-PML) boundary.

erator. Numerical methods such as finite-difference, finite-element or spectral-element are needed to solve the elastic wave equation. Figure 2.3 presents the isotropic elastic Marmousi2 model. Figure 2.4 shows an ocean bottom cable (OBC) seismic experiment data simulated with the elastic Marmousi2 model.

2.1.2 Linear forward problem

Many geophysical problems rely on linear forward modelling. Examples of the latter are the inversion of density from gravity anomalies and post-stack seismic deconvolution. The Born approximation is often used to represent a nonlinear forward problem associated to wave propagation phenomena via a linear approximation. The Born approximation is the basis of seismic migration techniques. The subsurface model is assumed to be a perturbed model around a background model

$$\mathbf{m} = \mathbf{m}_0 + \delta\mathbf{m}\,. \tag{2.7a}$$

Similarly, the wavefield is a perturbation around background wavefield

$$\mathbf{u} = \mathbf{u}_0 + \delta\mathbf{u}. \tag{2.7b}$$

The wave equation 2.4 is now rewritten as follows

$$\mathbf{S}(\mathbf{m}_0 + \delta\mathbf{m})(\mathbf{u}_0 + \delta\mathbf{u}) = \mathbf{f}. \tag{2.8}$$

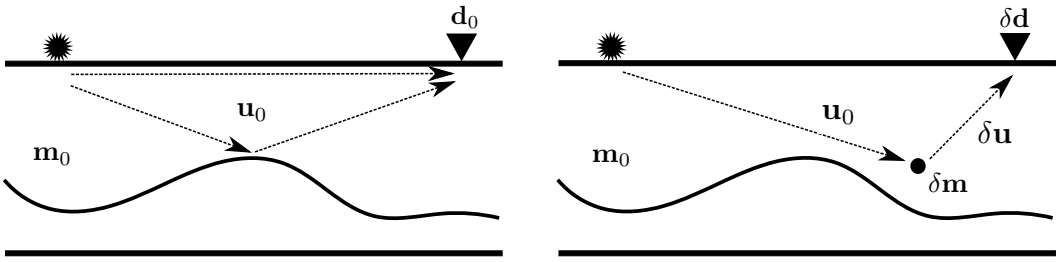
The above equation can be written as

$$(\mathbf{S}_0 + \delta\mathbf{S})(\mathbf{u}_0 + \delta\mathbf{u}) = \mathbf{f}, \tag{2.9}$$

where $\mathbf{S}_0$ is the background wave equation operator and $\delta\mathbf{S}$ is the perturbed wave equation operator. The propagation of seismic waves in the background model is described as

$$\mathbf{S}_0\mathbf{u}_0 = \mathbf{f}. \tag{2.10}$$

Figure 2.5 a shows the propagation of seismic waves in a background model.



(a) (b)

Figure 2.5: (a) The propagation of seismic wave in a background model $\mathbf{u}_0 = \mathbf{G}_0\mathbf{f}$. (b) The first-order Born approximation $\delta\mathbf{u} = -\mathbf{G}_0\delta\mathbf{S}\mathbf{u}_0$. The Born approximation takes the incident field $\mathbf{u}_0$ in place of the total field as the driving field at each scatterer. The incident field hits the scatterer and acts as secondary Born source. The black dot represents a scatter (model perturbation) $\delta\mathbf{m}$.

If the model perturbation is small, the second-order term of the perturbation is dropped, equation 2.9 can be simplified in the so-called first-order Born approximation

$$\mathbf{S}_0\delta\mathbf{u} = -\delta\mathbf{S}\mathbf{u}_0. \tag{2.11}$$

or

$$\delta\mathbf{u} = -\mathbf{S}_0^{-1}\delta\mathbf{S}\mathbf{u}_0 = -\mathbf{G}_0\delta\mathbf{S}\mathbf{u}_0 = -\mathbf{G}_0\frac{\partial\mathbf{S}}{\partial\mathbf{m}}\mathbf{u}_0\delta\mathbf{m}, \tag{2.12}$$

where $\mathbf{G}_0$ is the Green's operator in the background model. The scattered wavefield is

sampled at the receiver positions by the restriction operator $\mathbf{R}$

$$\underbrace{\delta\mathbf{d}}_{\text{data}} = \mathbf{R}\delta\mathbf{u} = -\mathbf{R}\mathbf{G}_0\frac{\partial\mathbf{S}}{\partial\mathbf{m}}\mathbf{u}_0\delta\mathbf{m} = \underbrace{-\mathbf{R}\mathbf{G}_0\frac{\partial\mathbf{S}}{\partial\mathbf{m}}\mathbf{G}_0\mathbf{f}}_{\text{Born operator}}\underbrace{\delta\mathbf{m}}_{\text{model}} . \tag{2.13}$$

The Green's operator $\mathbf{G}_0$ only depends on $\mathbf{m}_0$, it does not depends on $\delta\mathbf{m}$. Therefore, the Born approximation modelling operator is a linear forward problem of the form

$$\delta\mathbf{d} = \mathbf{\Phi}\delta\mathbf{m}, \tag{2.14}$$

where $\mathbf{\Phi}$ denotes the Born approximation modelling operator. Figure 2.5 b shows the first-order Born approximation. Reflectors with more complex structure can be regarded as a superposition of scatterers.

2.2 Inverse problem

The inverse problem consists of using the actual result of measurements to infer the values of the parameters that characterize the system (Tarantola, 2005). In our case, seismic observations are used to infer the elastic properties of the subsurface. As in many real scenarios, the data measurements are normally contaminated by noise. In general, noise can be assumed Gaussian and uncorrelated. The latter facilitates the design of cost function to solve the inverse problem. However, there are situation where one might prefer to adopt strategies that do not assume measurements contaminated with Gaussian noise.

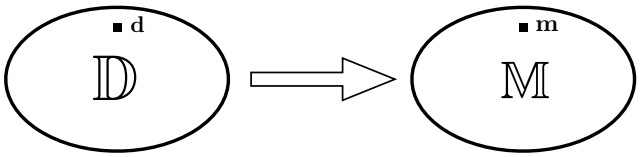


Figure 2.6: The sketch of general inverse problem from a measurement in data space $\mathbb{D}$ to a model parameter in model space $\mathbb{M}$.

Different criteria can be used for solving inverse problems (Claerbout and Muir, 1973; Huber, 1981; Chen and Sacchi, 2015, 2017b). The least-squares criterion is often adopted because it leads to linear system of equations when the problem is linear.

2.2.1 Linear inverse problem

Many geophysical inverse problems can be formulated as a linear least-squares inverse (LLSI) problem. Examples of the latter include deconvolution, linearized AVO inversion, least-squares migration (LSM), Radon de-multiple, etc. Linear inverse theory is popular because of its simplicity. In this section, I will discuss LSM as an example. The simplest approximation to the "inverse" solution of the Born forward modelling equation 2.14 can be computed via the adjoint (or transpose) of the linear operator

$$\delta\mathbf{m}^* = \mathbf{\Phi}^\dagger \delta\mathbf{d}, \tag{2.15}$$

where † denotes the adjoint of the Born operator. Applying the adjoint operator to obtain a distorted image $\delta\mathbf{m}^*$ is called seismic migration or seismic imaging in exploration seismology (Claerbout, 1970; French, 1975; Loewenthal and Mufti, 1983). Migration assumes that the smooth background model ($\mathbf{m}_0$) is known. Migration locates the singularities of subsurface structures ($\delta\mathbf{m}$). The adjoint process is simple but suffers from artifacts, low-resolution and it does not honour the data perturbation $\delta\mathbf{d}$. The LSM problem (Tarantola, 1984b; Bourgeois et al., 1989; Kuehl and Sacchi, 2003; Dai et al., 2012) consists of estimating the least-squares inverse solution of equation 2.14

$$\text{minimizes} \quad J(\delta\mathbf{m}) = \frac{1}{2}\|\mathbf{\Phi}\delta\mathbf{m} - \delta\mathbf{d}\|_2^2, \tag{2.16}$$

where $\delta\mathbf{d}$ is the observed scattered data and J is a quadratic function of $\delta\mathbf{m}$. Figure 2.7 shows a simple 2D quadratic function.

The gradient of the objective function 2.16 is given by

$$\frac{\partial J(\delta\mathbf{m})}{\partial \delta\mathbf{m}^\dagger} = \mathbf{\Phi}^\dagger(\mathbf{\Phi}\delta\mathbf{m} - \delta\mathbf{d}). \tag{2.17}$$

The LLSI problem has a unique local minimum which is also the global minimum. The minimum is at the point where the gradient vanishes $[\partial J(\delta\mathbf{m})]/\partial\delta\mathbf{m}^\dagger = 0$

$$\mathbf{\Phi}^\dagger\mathbf{\Phi}\delta\mathbf{m} = \mathbf{\Phi}^\dagger\delta\mathbf{d}. \tag{2.18}$$

The direct analytical solution of LLSI problem 2.16 is

$$\delta\hat{\mathbf{m}} = \left(\mathbf{\Phi}^\dagger\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^\dagger\delta\mathbf{d}. \tag{2.19}$$

The size of operator $\mathbf{\Phi}^\dagger\mathbf{\Phi}$ can be large. The direct solution is impractical for large-scale problems. Linear iterative algorithms are usually used to solve the linear inverse problem. For instance, one can adopt the conjugate gradient least-squares (CGLS) method or the

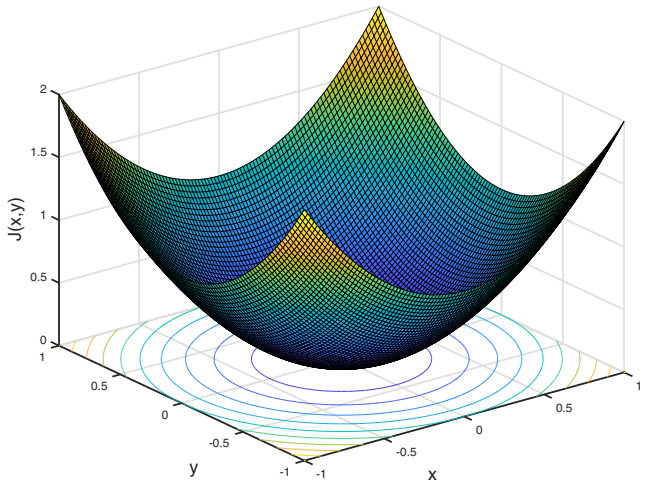


Figure 2.7: An example of 2D quadratic function $J(x, y) = x^2 + y^2$.

LSQR algorithm (Paige and Saunders, 1982) to iteratively solve the inverse problem without assembling the operator $\mathbf{\Phi}^\dagger\mathbf{\Phi}$.

In Chapters 3 and 4 of this thesis, I discuss the linear inverse problem and I propose elastic least-squares reverse time migration (LSRTM) algorithms for elastic seismic imaging of subsurface structures. At this point, it is important to mention that the operator $\mathbf{\Phi}$ is never stored as a matrix. In my research, $\mathbf{\Phi}$ and $\mathbf{\Phi}^\dagger$ are functions (subroutines) in C language. This point will become clear in Chapters 3 and 4.

2.2.2 Nonlinear inverse problem

The nonlinear least-squares inverse (NLSI) problem consists of searching a model that fits the observed data with a nonlinear forward operator in a least-squares sense. Geophysical inverse problems such as inversion of the Knott-Zoeppritz equations and full-waveform inversion (FWI) are nonlinear inverse problem. In this section, I will discuss FWI as an introductory example. FWI estimates the coefficients of the wave equation given the seismic data observed on Earth's surface (boundary values). FWI does not assume a scale separation of the subsurface model. It aims at retrieving both the long wavelengths and short wavelengths of the model. The FWI problem is usually formulated via the following problem

$$\text{minimizes} \quad J(\mathbf{m}) = \frac{1}{2}\|\boldsymbol{\phi}(\mathbf{m}) - \mathbf{d}\|_2^2, \tag{2.20}$$

where the function J is non-quadratic function of $\mathbf{m}$. The operator $\boldsymbol{\phi}(\cdot)$ indicates the nonlinear forward modelling. The NLSI problem normally has a large number of local minima. Figure 2.8 shows a simple example of a 2D non-quadratic function.

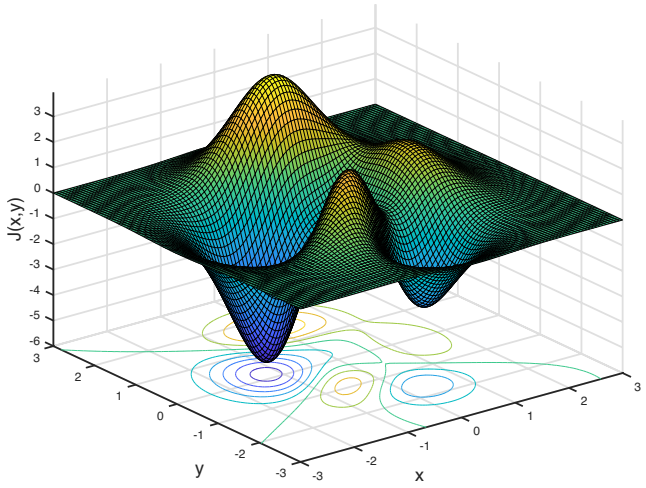


Figure 2.8: An example of 2D non-quadratic function.

In general, there is no analytical solution for the NLSI problem 2.20. The dimension of the model space is usually large. Global optimization methods, such as Monte-Carlo search, simulated annealing and genetic algorithms, are too expensive for solving large-scale problems. Local optimization iterative algorithms are preferred for FWI. The minimum of the objective function J is sought in the vicinity of current model $\mathbf{m}$. The updated model is a perturbation of the available model $\mathbf{m} \rightarrow \mathbf{m} + \delta\mathbf{m}$. The objective function in the vicinity of $\mathbf{m}$ can be expanded using Taylor series

$$J(\mathbf{m} + \delta\mathbf{m}) = J(\mathbf{m}) + \delta\mathbf{m}^\dagger \frac{\partial J(\mathbf{m})}{\partial \mathbf{m}} + \frac{1}{2}\delta\mathbf{m}^\dagger \frac{\partial^2 J(\mathbf{m})}{\partial \mathbf{m}^2}\delta\mathbf{m} + \mathcal{O}(\delta\mathbf{m}^3). \tag{2.21}$$

Discarding the third and higher-order perturbation terms, the derivative of objective function with respect to model parameters is given by

$$\frac{\partial J(\mathbf{m} + \delta\mathbf{m})}{\partial \delta\mathbf{m}} = \frac{\partial J(\mathbf{m})}{\partial \mathbf{m}} + \frac{\partial^2 J(\mathbf{m})}{\partial \mathbf{m}^2}\delta\mathbf{m}. \tag{2.22}$$

The minimum of the objective function in the vicinity of model $\mathbf{m}$ is obtained when the derivative vanishes. The latter leads to the optimal model update (Virieux and Operto,

2009)

$$\begin{aligned}\delta\mathbf{m} &= -\left[\frac{\partial^2 J(\mathbf{m})}{\partial \mathbf{m}^2}\right]^{-1}\frac{\partial J(\mathbf{m})}{\partial \mathbf{m}} \\ &= -\left[\left(\frac{\partial \boldsymbol{\phi}(\mathbf{m})}{\partial \mathbf{m}}\right)^{\dagger}\left(\frac{\partial \boldsymbol{\phi}(\mathbf{m})}{\partial \mathbf{m}}\right)+\left(\frac{\partial^2 \boldsymbol{\phi}(\mathbf{m})}{\partial \mathbf{m}^2}\right)^{\dagger}(\boldsymbol{\phi}(\mathbf{m})-\mathbf{d})\right]^{-1}\left(\frac{\partial \boldsymbol{\phi}(\mathbf{m})}{\partial \mathbf{m}}\right)^{\dagger}(\boldsymbol{\phi}(\mathbf{m})-\mathbf{d}) \\ &= \left[\boldsymbol{\Phi}^{\dagger}\boldsymbol{\Phi}-\left(\frac{\partial \boldsymbol{\Phi}}{\partial \mathbf{m}}\right)^{\dagger}\delta\mathbf{d}\right]^{-1}\boldsymbol{\Phi}^{\dagger}\delta\mathbf{d},\end{aligned} \tag{2.23}$$

where $\boldsymbol{\Phi} = \partial\boldsymbol{\phi}(\mathbf{m})/\partial\mathbf{m}$ is the Fréchet derivative of the nonlinear operator and $\delta\mathbf{d} = \mathbf{d} - \boldsymbol{\phi}(\mathbf{m})$. The Fréchet derivative is given by

$$\boldsymbol{\Phi} = \frac{\partial \boldsymbol{\phi}(\mathbf{m})}{\partial \mathbf{m}} = \mathbf{R}\frac{\partial \mathbf{u}}{\partial \mathbf{m}}, \tag{2.24}$$

where $\mathbf{u}$ is the wavefield and $\mathbf{R}$ is the sampling operator. The derivative $\partial\mathbf{u}/\partial\mathbf{m}$ can be obtained by differentiating the wave equation 2.4 with respect to $\mathbf{m}$

$$\frac{\partial \mathbf{S}}{\partial \mathbf{m}}\mathbf{u} + \mathbf{S}\frac{\partial \mathbf{u}}{\partial \mathbf{m}} = \mathbf{0}. \tag{2.25}$$

The Fréchet derivative $\boldsymbol{\Phi}$ becomes

$$\boldsymbol{\Phi} = \mathbf{R}\frac{\partial \mathbf{u}}{\partial \mathbf{m}} = -\mathbf{R}\mathbf{S}^{-1}\frac{\partial \mathbf{S}}{\partial \mathbf{m}}\mathbf{u} = -\mathbf{R}\mathbf{G}\frac{\partial \mathbf{S}}{\partial \mathbf{m}}\mathbf{G}\mathbf{f}, \tag{2.26}$$

where $\mathbf{G}$ is the Green's operator in the current model solution $\mathbf{m}$. It is easy to recognize that the Fréchet derivative operator $\boldsymbol{\Phi}$ is the Born approximation operator 2.13.

The term in the squares brackets in equation 2.23 is named the full Hessian of the FWI problem. The computation of the action of second-order derivative on a vector involves the second-order adjoint-state method (Santosa and Symes, 1988; Fichtner and Trampert, 2011). Discarding this term leads to the Gauss-Newton update

$$\delta\mathbf{m} = (\boldsymbol{\Phi}^{\dagger}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\dagger}\delta\mathbf{d}. \tag{2.27}$$

I recognize that the Gauss-Newton model update expression is equivalent to the LSM formulation 2.19. The fixed-point iteration for the Gauss-Newton FWI algorithm is given by

$$\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} + \eta^{(k)}\delta\mathbf{m}^{(k)} = \mathbf{m}^{(k)} + \eta^{(k)}({\boldsymbol{\Phi}^{(k)}}^{\dagger}\boldsymbol{\Phi}^{(k)})^{-1}{\boldsymbol{\Phi}^{(k)}}^{\dagger}\delta\mathbf{d}^{(k)} \tag{2.28}$$

where $\mathbf{m}^{(k)}$ is the model at the kth iteration, $\eta^{(k)}$ is the step size at kth iteration and $\delta\mathbf{m}^{(k)}$ is the Gauss-Newton search direction at kth iteration. Therefore, the Gauss-Newton FWI

method is an *iterative least-squares migration problem*. In the conventional steepest-descent FWI, the approximated Hessian $\mathbf{\Phi}^{(k)^\dagger}\mathbf{\Phi}^{(k)}$ is ignored to save computational cost

$$\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} + \eta^{(k)}\mathbf{\Phi}^{(k)^\dagger}\delta\mathbf{d}^{(k)}. \tag{2.29}$$

The model update direction in each iteration of steepest-descent FWI is the same with the migration formulation 2.15. The steepest-descent FWI can be regarded as an *iterative migration problem*.

In Chapter 5, I will discuss the nonlinear inverse problem. For this purpose, I will propose a matrix-free elastic Gauss-Newton FWI algorithm for elastic parameters inversion.

CHAPTER 3

Elastic least-squares reverse time migration [1]

Time domain elastic least-squares reverse time migration (LSRTM) is formulated as a linearized elastic full waveform inversion (FWI) problem. The elastic Born approximation and elastic reverse time migration (RTM) operators are derived from the time-domain continuous adjoint-state method. Our approach defines P-wave and S-wave impedance perturbations as unknown elastic images. Our algorithm is obtained using continuous functional analysis where the problem is discretized at the final stage (optimize-before-discretize approach). The discretized numerical versions of the elastic Born operator and its adjoint (elastic RTM operator) pass the dot product test. The conjugate gradient least squares (CGLS) method is used to solve the least-squares migration quadratic optimization problem. In other words, the Hessian operator for elastic LSRTM is implicitly inverted via a matrix-free algorithm that only requires the action of forward and adjoint operators on vectors. The diagonal of the pseudo-Hessian operator is used to design a preconditioning operator to accelerate the convergence of the elastic LSRTM. The elastic LSRTM provides higher resolution images with fewer artifacts and a superior balance of amplitudes when compared to elastic RTM. More important, elastic LSRTM can mitigate cross-talk between the P-wave and S-wave impedance perturbations given that the off-diagonal elements of the Hessian are attenuated via the inversion.

3.1 Introduction

Acoustic reverse time migration (RTM) was initially proposed for poststack seismic data migration (McMechan, 1983; Baysal et al., 1983; Whitmore, 1983). Prestack RTM was

[1] A version of this chapter has been published in Chen, K., and M. D. Sacchi, 2017, Elastic least-squares reverse time migration via linearized elastic full-waveform inversion with pseudo-Hessian preconditioning: Geophysics, **82**, no. 5, S341-S358.

implemented in shot-profile domain by either applying the excitation-time imaging condition (Chang and McMechan, 1986) or the crosscorrelation imaging condition (Etgen, 1986). RTM utilizes the two-way wave equation for extrapolating wavefields into the interior of the earth. RTM can handle steep and complex geological structures such as sedimentary areas with salt inclusions (Etgen et al., 2009). Lailly (1983) pointed out that computing the gradient in full waveform inversion (FWI) is equivalent to apply the prestack RTM operator on data residuals. From then, RTM has been connected to the adjoint-state method that is utilized in FWI (Tromp et al., 2005; Plessix, 2006; Douma et al., 2010).

Acoustic methods approximate the elastic solid earth by a fluid. However, shear waves also convey subsurface rock property information. Shear waves can be recorded by multicomponent sensors (Hardage et al., 2011). The multicomponent seismic data can be used for elastic seismic imaging. Elastic RTM was developed by Sun and McMechan (1986) and Chang and McMechan (1987) with the excitation-time imaging condition. The elastic data record is back-extrapolated using the elastic wave equation and the image time is computed by ray tracing using a P-wave velocity model. In these methods, P- and S-images are not explicitly separated. Instead, they compute the so-called vertical and horizontal images. The latter could impede the proper interpretation of multicomponent images. Sun and McMechan (2001) and Sun et al. (2004) proposed another method where the elastic data are first back-propagated by the elastic wave equation and then a Helmholtz decomposition (Dellinger and Etgen, 1990) is used to separate P- and S-wave modes at a predefined datum. The separated P- and S-wave data are extrapolated upwards to the surface of the earth via the acoustic wave equation. Then, the separated P- and S-wave data are injected into two acoustic RTM algorithms that adopt an excitation-time imaging condition. In a similar vein, Yan and Sava (2008) proposed to perform forward and backward extrapolation of wavefields via elastic wave equation and to apply Helmholtz decomposition to separate P- and S-wave fields followed by the crosscorrelation imaging condition (Lu et al., 2009). One drawback of methods based on the Helmholtz wave-mode decomposition is that the P-S images will have polarity reversals. The latter will prevent stacking of individual shot contributions. Polarity reversal correction strategies for elastic RTM methods that adopt Helmholtz decomposition have been proposed (Du et al., 2012; Duan and Sava, 2015). Early work in the field of FWI (Tarantola, 1986) recognized that the gradient in each iteration of an elastic FWI algorithm is indeed a prestack elastic RTM operator applied to data residuals. This idea was adopted by Luo et al. (2009) and Zhu et al. (2009) who proposed elastic RTM algorithms based on FWI sensitivity kernels. In these methods, the elastic model perturbations on the background model are defined as "elastic images".

A migration algorithm can be described as the adjoint of the forward Born modelling operator. In this case, the migrated image can be considered a blurred version of subsurface model perturbations. The deblurring operator is the inverse of the Hessian of the imaging problem

which is defined, for linearized inversion, as the cascade of the forward and adjoint operators. Several strategies have been developed to approximate the Hessian to deblur seismic images. For instance, prestack least-squares Kirchhoff migration was initially formulated to retrieve acoustic velocity perturbation given a background velocity model (Tarantola, 1984b; Lambare et al., 1992). Nemeth et al. (1999) implemented least-squares Kirchhoff migration for migrating incomplete reflection seismic data. Least-squares Fourier finite-difference one-way wave-equation migration was formulated by Rickett (2003) and then utilized by Tang (2009) for blended seismic data migration. Least-squares reverse time migration (LSRTM) was investigated for inverting P-wave impedance perturbation and P-wave velocity perturbation (Bourgeois et al., 1989) under the name "linearized inversion". Ostmo et al. (2002) implemented acoustic LSRTM in frequency domain under the name of "linearized waveform inversion". In recent years, LSRTM was further developed for high-resolution true amplitude imaging (Dong et al., 2012; Zhang et al., 2015; Yao and Jakubowicz, 2016), migration of multi-source blended seismic data (Dai et al., 2012; Xue et al., 2016), and for imaging with multiples (Zhang and Schuster, 2014; Wong et al., 2015). The acoustic slowness perturbation or velocity perturbation represents the "image" or reflectivity model. We would like to classify the above-mentioned least-squares migration techniques as linearized waveform inversion. These methods, in general, invert an image that is proportional to an averaged subsurface reflectivity. Least-squares migration can also be implemented in an extended-domain to produce an image volume that depends on redundant parameters (Symes, 2008). This idea was proposed and implemented by Kuehl and Sacchi (2003) using the survey-sinking approach (Claerbout, 1985). The technique was modified to process 3D field data via the constant azimuth approximation by Wang et al. (2005) and to include sparsity constraints to increase vertical resolution by Wang and Sacchi (2007). Moreover, Kaplan et al. (2010) derived least-squares split-step migration for extended shot domain image inversion. The latter was also applied for the migration of blended seismic data (Cheng et al., 2016). Similarly, Dai and Schuster (2013) implemented LSRTM in extended plane wave domain for blended seismic data. Finally, we also mention that Hou and Symes (2016) and Huang et al. (2016) implemented LSRTM in extended subsurface offset domain and extended shot domain, respectively. We can classify this type of least-squares migration algorithms as extended least-squares migration (Symes, 2008) where an extended reflectivity volume is inverted. Least-squares migration has also been formulated in image domain that the inverse of Hessian is approximated via various strategies (Rickett, 2003; Guitton, 2004; Fletcher et al., 2016; Wang et al., 2016). The image domain least-squares migration requires lower computational cost than data domain least-squares migration methods.

The aforementioned least-squares migration methods are based on the acoustic approximation. Land data and ocean bottom data record both P- and S-waves. The geophysical community has investigated several elastic least-squares migration algorithms. Elastic least-

squares ray-Born migration/inversion was implemented by Beydoun and Mendes (1989) and Jin et al. (1992) in heterogeneous media for multicomponent seismic data. Tura and Johnson (1993) discussed an elastic least-squares migration/inversion method in the f-k domain for homogeneous background media. Anikiev et al. (2013) investigated the decoupling of parameters for frequency domain elastic LSRTM for the case of a point scatter in a homogeneous background model. In these studies, the elastic parameter perturbations are inverted and defined as "elastic images". Recently, Stanton and Sacchi (2015, 2017) and Xu et al. (2016) utilized Helmholtz decomposition (Dellinger and Etgen, 1990) for elastic least-squares split-step and reverse time migration for the inversion of elastic reflectivity volumes in extended domain, respectively.

We propose to formulate time-domain elastic LSRTM as a linearized elastic FWI problem (Chen and Sacchi, 2016). The discussion of the relationship between LSRTM and Gauss-Newton FWI can be found in Chapter 2. We derive the elastic Born approximation and elastic RTM operators via a time-domain continuous adjoint-state method for the first-order velocity-stress elastic wave equation system. The adjoint-state equation system is equivalent to the state equation system after a variable transformation (Vigh et al., 2014). This allows us to reuse our forward modelling code to compute the receiver side wavefield (adjoint-state variable). The differences are the replacement of the explosive source by an adjoint source and the back-propagation of the adjoint source from final time to zero time. In our work, P-wave and S-wave impedance perturbations are defined as elastic images. The two terms "model perturbation" and "image" are used interchangeably. All derivations are in continuous functional form, the problem is discretized after developing the algorithm. This is the so-called optimize-before-discretize approach (Borzi and Schulz, 2012). We carefully discretize the continuous forms of the elastic Born operator and the elastic RTM operator to guarantee that the operators pass the dot product test (Mora, 1987a; Claerbout, 1992). The latter allows the use of the conjugate gradient least squares (CGLS) algorithm (Hestenes and Stiefel, 1952; Paige and Saunders, 1982) to solve the least-squares migration optimization problem. In other words, the Hessian operator is implicitly inverted via a matrix-free algorithm that requires only the forward and adjoint operator applied to vectors. Another advantage of the CGLS algorithm is that the step size of the method is analytically calculated. One does not need to compute the step size via line search (Dong et al., 2012) as it often done when the forward and adjoint operators do not satisfy the numerical condition for adjointness. We investigate the structure of the multiparameter Hessian operator for elastic LSRTM to design a preconditioning strategy. We adapt the pseudo-Hessian (Shin et al., 2001a) to the time-domain elastic case and derive the equations of the Hessian and pseudo-Hessian for elastic parameters. We adopt the diagonal of the pseudo-Hessian to precondition the CGLS algorithm. Our elastic LSRTM yields higher resolution images with fewer artifacts and more balanced amplitudes than

elastic RTM. Moreover, elastic LSRTM can reduce the multiparameter cross-talk given that the off-diagonal terms of the Hessian operator are attenuated via the least-squares inversion. Finally, the adopted pseudo-Hessian preconditioning strategy accelerates the convergence of our algorithm and improves the amplitude responses of both the P-wave and S-wave impedance perturbation images.

This article is organized as follows. First, we describe the system of equations that we have adopted to forward model elastic wavefields. Then, we introduce linearized forward modelling by adopting the elastic Born approximation. Subsequently, we derive the adjoint operator (elastic RTM operator) of the linearized forward modelling operator. We discuss the numerical adjointness of forward and adjoint operators and propose to solve the elastic least-squares reverse time migration via the conjugate gradients least squares (CGLS) algorithm. Furthermore, we discuss preconditioning as a strategy to accelerate the convergence of the CGLS algorithm. In the last section, we provide numerical examples that permit us to evaluate the performance of the proposed algorithm. The first numerical example is a simple elastic Camembert model. Our second numerical example entails adopting the elastic Marmousi2 model for additional tests.

3.2 Theory

3.2.1 Heterogeneous, isotropic elastic wave equation

In this study, we assume a 2D heterogeneous, isotropic elastic earth media. The propagation of P-SV wave is governed by first-order partial differential equations (Virieux, 1986; Vigh et al., 2014)

$$\begin{pmatrix} \rho\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \frac{\partial}{\partial t}\mathbf{u} - \begin{pmatrix} \mathbf{0} & \mathbf{D} \\ \mathbf{C}\mathbf{D}^T & \mathbf{0} \end{pmatrix} \mathbf{u} = \mathbf{f}, \tag{3.1}$$

where

$$\mathbf{u} = \begin{pmatrix} \mathbf{v} \\ \boldsymbol{\sigma} \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} v_x \\ v_z \end{pmatrix}, \qquad \boldsymbol{\sigma} = \begin{pmatrix} \sigma_{xx} \\ \sigma_{zz} \\ \sigma_{xz} \end{pmatrix}, \qquad \mathbf{f} = \begin{pmatrix} \mathbf{0} \\ \mathbf{f}_\sigma \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} \lambda + 2\mu & \lambda & 0 \\ \lambda & \lambda + 2\mu & 0 \\ 0 & 0 & \mu \end{pmatrix}, \qquad \mathbf{D} = \begin{pmatrix} \dfrac{\partial}{\partial x} & 0 & \dfrac{\partial}{\partial z} \\ 0 & \dfrac{\partial}{\partial z} & \dfrac{\partial}{\partial x} \end{pmatrix},$$

with zero initial condition $\mathbf{u}\,|_{t=0} = \mathbf{0}$ and appropriate boundary conditions. The vector $\mathbf{u}$ denotes the wavefield where $\mathbf{v}$ is the vector particle velocity field and $\boldsymbol{\sigma}$ is the stress vector.

Similarly, ρ indicates density, $\mathbf{C}$ is the isotropic elastic tensor in Voigt notation with λ and μ the Lamé parameters. The matrix $\mathbf{D}$ is a collection of spatial differential operators and $\mathbf{f}_\sigma$ is the explosive source term. Finally, $\mathbf{I}$ is the identity matrix. In the wave equation, we dropped the dependence on spatial and temporal coordinates $\mathbf{x}$ and t of our variables to make the notations concise but we understand that $\mathbf{v} = \mathbf{v}(\mathbf{x}, t)$, $\lambda = \lambda(\mathbf{x})$, etc. The elastic wave equation is the state equation of the elastic parameter inversion problem when it is regarded as optimal control problem (Lions, 1971; Plessix, 2006). Abstractly, the elastic wave equation 3.1 can be written in functional form as follows

$$\mathbf{S}\mathbf{u} = \mathbf{f}, \tag{3.2}$$

where $\mathbf{u}$ is the wavefield vector in space $\mathbb{U}$, $\mathbf{f}$ is the source vector in space $\mathbb{F}$, $\mathbf{S} = \mathbf{S}(\mathbf{m})$ ($\mathbf{S} : \mathbb{U} \to \mathbb{F}$) is the wave equation operator with initial conditions and boundary conditions, and $\mathbf{m} = (\rho, \lambda, \mu)^T$ denotes the model parameter vector in space $\mathbb{M}$. The solution of the wave equation can be abstractly written as

$$\mathbf{u} = \mathbf{S}^{-1}\mathbf{f} = \mathbf{G}\mathbf{f}, \tag{3.3}$$

where $\mathbf{G} = \mathbf{S}^{-1}$ is the inverse of wave equation operator $\mathbf{S}$ named the Green's operator (Tarantola, 1988). The Green's operator is an integral operator with the integration kernel given by the Green's function of the wave equation (Tarantola, 1988). The wavefield $\mathbf{u}$ is linear in the source term $\mathbf{f}$ but is nonlinear in the model $\mathbf{m}$. If the source term is assumed known, $\mathbf{u}$ can be regarded as a nonlinear function of the model parameters $\mathbf{m}$

$$\mathbf{u} = \mathbf{u}(\mathbf{m}). \tag{3.4}$$

In a general heterogeneous media, there is no analytic solution for $\mathbf{u}$ given $\mathbf{m}$. A numerical method must be used to solve the forward problem. In this article, a time domain staggered-grid finite-difference (FD) scheme (Virieux, 1986; Levander, 1988) is utilized to discretize the continuous form elastic wave equation 3.1. The unsplit Convolutional Perfectly Matched Layer (C-PML) method is used to absorb incident waves on artificial computational boundaries (Komatitsch and Martin, 2007). In our FD code, we adopted a second-order centred difference scheme in time and a selectable order staggered difference scheme in space. In seismic exploration, the wavefield is sampled at the surface of the earth by a finite number of receivers

$$\mathbf{d}(\mathbf{m}) = \mathbf{R}\mathbf{u}(\mathbf{m}), \tag{3.5}$$

where $\mathbf{d}$ is recorded full waveform seismic data and operator $\mathbf{R}$ represents the sampling operator that extracts the wavefield at receivers positions.

3.2.2 The linearized forward problem: elastic Born approximation

The relationship between the seismic data $\mathbf{d}$ and model parameters $\mathbf{m}$ is nonlinear as discussed above. An expansion in terms of Taylor series can be used to linearize the nonlinear forward problem $\mathbf{d} = \mathbf{d}(\mathbf{m})$. A perturbation of the model parameters $\mathbf{m} \rightarrow \mathbf{m} + \delta\mathbf{m}$ leads to a perturbation of the seismic data $\mathbf{d}(\mathbf{m}) \rightarrow \mathbf{d}(\mathbf{m} + \delta\mathbf{m})$

$$\mathbf{d}(\mathbf{m} + \delta\mathbf{m}) = \mathbf{d}(\mathbf{m}) + \frac{\partial \mathbf{d}}{\partial \mathbf{m}}\delta\mathbf{m} + \mathcal{O}(\|\delta\mathbf{m}\|^2), \tag{3.6}$$

where $\mathbf{m}$ is the background model, $\delta\mathbf{m}$ is the model perturbation, $\mathbf{d}(\mathbf{m})$ is the seismic data associated to propagation in the background model, $\mathbf{d}(\mathbf{m}+\delta\mathbf{m})$ is the seismic data associated to propagation in the perturbed model and the linear operator $\frac{\partial \mathbf{d}}{\partial \mathbf{m}} = \left(\frac{\partial \mathbf{d}}{\partial \rho}, \frac{\partial \mathbf{d}}{\partial \lambda}, \frac{\partial \mathbf{d}}{\partial \mu}\right)$ is the Fréchet derivative or Jacobian matrix of $\mathbf{d}$. The second and higher order terms in the Taylor series are dropped resulting in the first-order Born approximation

$$\delta\mathbf{d} = \mathbf{L}\delta\mathbf{m} = \frac{\partial \mathbf{d}}{\partial \mathbf{m}}\delta\mathbf{m} = \mathbf{R}\frac{\partial \mathbf{u}}{\partial \mathbf{m}}\delta\mathbf{m}, \tag{3.7}$$

where Fréchet derivative $\mathbf{L} = \frac{\partial \mathbf{d}}{\partial \mathbf{m}}$ is the Born modelling operator, $\delta\mathbf{d}$ is the first-order scattered seismic data $\delta\mathbf{d} \approx \mathbf{d}(\mathbf{m} + \delta\mathbf{m}) - \mathbf{d}(\mathbf{m})$, and the linear operator $\frac{\partial \mathbf{u}}{\partial \mathbf{m}}$ is the Fréchet derivative or Jacobian matrix of $\mathbf{u}$. The Fréchet derivative is prohibitively expensive to compute explicitly. Alternatively, the adjoint-state method is used to compute the action of the Fréchet derivative on vectors. For this purpose, we first differentiate the wave equation 3.2 with respect to $\mathbf{m}$ (Fichtner, 2010; Fichtner and Trampert, 2011)

$$\frac{\partial \mathbf{S}}{\partial \mathbf{m}}\mathbf{u} + \mathbf{S}\frac{\partial \mathbf{u}}{\partial \mathbf{m}} = \mathbf{0}, \tag{3.8}$$

where linear operator $\frac{\partial \mathbf{S}}{\partial \mathbf{m}} = \left(\frac{\partial \mathbf{S}}{\partial \rho}, \frac{\partial \mathbf{S}}{\partial \lambda}, \frac{\partial \mathbf{S}}{\partial \mu}\right)$. The right side of equation becomes zero because the source term does not depend on the model. Then, the Fréchet derivative of $\mathbf{u}$ can be expressed as follows

$$\frac{\partial \mathbf{u}}{\partial \mathbf{m}} = -\mathbf{S}^{-1}\frac{\partial \mathbf{S}}{\partial \mathbf{m}}\mathbf{u} = -\mathbf{S}^{-1}\frac{\partial \mathbf{S}}{\partial \mathbf{m}}\mathbf{S}^{-1}\mathbf{f}, \tag{3.9}$$

where the multiplication of two operators follows the rule $\mathbf{ABv} = \mathbf{A}(\mathbf{Bv})$ (Chen and Lee, 2015). The term $-\frac{\partial \mathbf{S}}{\partial \mathbf{m}}\mathbf{u}$ is the so-called "virtual secondary source" which is the product of the incident wavefield and $\frac{\partial \mathbf{S}}{\partial \mathbf{m}}$. The operator $\frac{\partial \mathbf{S}}{\partial \mathbf{m}}$ represents the radiation pattern of the virtual secondary source (Pageot et al., 2013). Inserting the expression of Fréchet derivative of $\mathbf{u}$ into equation 3.7, the Born approximation can be written as

$$\delta\mathbf{d} = \mathbf{L}\delta\mathbf{m} = \frac{\partial \mathbf{d}}{\partial \mathbf{m}}\delta\mathbf{m} = -\mathbf{R}\mathbf{S}^{-1}\frac{\partial \mathbf{S}}{\partial \mathbf{m}}\mathbf{u}\delta\mathbf{m}. \tag{3.10}$$

Equation 3.10 indicates that the incident wavefield $\mathbf{u}$ hits the scatters $\delta\mathbf{m}$, acts as a secondary source and generates the scattered wavefield. The latter is sampled at the surface of the earth by receivers and generates the scattered seismic data. The linearized Born modelling operator $\mathbf{L}$ only depends on the smooth background model $\mathbf{m}$ and the acquisition geometry.

One can apply the abstract linearized Born approximation equation 3.10 directly to the first-order velocity-stress elastic wave equation system 3.1. Alternatively, perturbing the elastic wave equation will lead to the same result. A perturbation of the model parameters

$$\rho \rightarrow \rho + \delta\rho, \tag{3.11a}$$

$$\lambda \rightarrow \lambda + \delta\lambda, \tag{3.11b}$$

$$\mu \rightarrow \mu + \delta\mu, \tag{3.11c}$$

leads to a perturbation of the wavefield

$$\mathbf{u} \rightarrow \mathbf{u} + \delta\mathbf{u}. \tag{3.11d}$$

Inserting equation 3.11 into equation 3.1, subtracting equation 3.1, and dropping second and higher order terms leads to the Born approximation for the first-order velocity stress elastic wave equation system

$$\begin{pmatrix} \rho\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \frac{\partial}{\partial t}\delta\mathbf{u} - \begin{pmatrix} \mathbf{0} & \mathbf{D} \\ \mathbf{C}\mathbf{D}^T & \mathbf{0} \end{pmatrix} \delta\mathbf{u} = \begin{pmatrix} -\delta\rho\dfrac{\partial}{\partial t}\mathbf{v} \\ \delta\mathbf{C}\mathbf{D}^T\mathbf{v} \end{pmatrix}, \tag{3.12}$$

where

$$\delta\mathbf{u} = \begin{pmatrix} \delta\mathbf{v} \\ \delta\boldsymbol{\sigma} \end{pmatrix}, \quad \delta\mathbf{v} = \begin{pmatrix} \delta v_x \\ \delta v_z \end{pmatrix}, \quad \delta\boldsymbol{\sigma} = \begin{pmatrix} \delta\sigma_{xx} \\ \delta\sigma_{zz} \\ \delta\sigma_{xz} \end{pmatrix}, \quad \delta\mathbf{C} = \begin{pmatrix} \delta\lambda + 2\delta\mu & \delta\lambda & 0 \\ \delta\lambda & \delta\lambda + 2\delta\mu & 0 \\ 0 & 0 & \delta\mu \end{pmatrix},$$

with zero initial condition $\delta\mathbf{u}\,|_{t=0} = \mathbf{0}$ and appropriate boundary conditions. The vector $\delta\mathbf{u}$ is the scattered wavefield, with $\delta\mathbf{v}$ and $\delta\boldsymbol{\sigma}$ the scattered particle velocity field and scattered stress field due to model perturbations $\delta\rho$, $\delta\lambda$ and $\delta\mu$, and $\mathbf{v}$ is the incident particle velocity field. The right side of equation 3.12 is the so-called "secondary source". The scattered wavefield can be computed using the same finite-difference code that is adopted to compute the source side incident wavefield in equation 3.1. The scattered data is obtained by sampling the scattered wavefield at the receiver positions $\delta\mathbf{d} = \mathbf{R}\delta\mathbf{u}$.

3.2.3 The linearized adjoint problem: elastic reverse time migration

The migration operator is the adjoint of the Born modelling operator that maps from reflection data to model perturbation or image. The adjoint operator of the Born operator satisfies

$$\langle \delta\mathbf{d}, \mathbf{L}\delta\mathbf{m}\rangle_{\mathbb{D}} = \langle \mathbf{L}^{\dagger}\delta\mathbf{d}, \delta\mathbf{m}\rangle_{\mathbb{M}}, \tag{3.13}$$

where $\dagger$ denotes the adjoint of an operator, $\langle\cdot,\cdot\rangle_{\mathbb{D}}$ and $\langle\cdot,\cdot\rangle_{\mathbb{M}}$ denotes inner products in data domain and model domain, respectively. The adjoint of Born operator (equation 3.10) applied to reflection data can be expressed as (Tarantola, 1984a)

$$\delta\mathbf{m}^{*} = \mathbf{L}^{\dagger}\delta\mathbf{d} = \left(\frac{\partial\mathbf{d}}{\partial\mathbf{m}}\right)^{\dagger}\delta\mathbf{d} = -\left(\frac{\partial\mathbf{S}}{\partial\mathbf{m}}\mathbf{u}\right)^{\dagger}\left(\mathbf{S}^{-1}\right)^{\dagger}\mathbf{R}^{\dagger}\delta\mathbf{d} = -\left(\frac{\partial\mathbf{S}}{\partial\mathbf{m}}\mathbf{u}\right)^{\dagger}\left(\mathbf{S}^{\dagger}\right)^{-1}\mathbf{R}^{\dagger}\delta\mathbf{d}, \tag{3.14}$$

where $\mathbf{R}^{\dagger}$ is the adjoint of sampling operator $\mathbf{R}$. Notice that we used the property $\left(\mathbf{S}^{-1}\right)^{\dagger} = \left(\mathbf{S}^{\dagger}\right)^{-1}$ in equation 3.14 (Tarantola, 1988). We adopted the symbol $\delta\mathbf{m}^{*}$ to represent the model perturbation that one can obtain by applying the adjoint operator to data perturbation $\delta\mathbf{d}$. Evidently, the adjoint operator is not equal to the inverse of the linearized forward operator and therefore, $\delta\mathbf{m}^{*} \neq \delta\mathbf{m}$. To continue with our analysis, we now introduce the adjoint-state variable $\mathbf{p} = \left(\mathbf{S}^{\dagger}\right)^{-1}\mathbf{R}^{\dagger}\delta\mathbf{d}$. The latter satisfies the "adjoint-state equation" corresponding to the state equation 3.2

$$\mathbf{S}^{\dagger}\mathbf{p} = \mathbf{R}^{\dagger}\delta\mathbf{d}, \tag{3.15}$$

where $\mathbf{p}$ is the adjoint-state variable of the state variable $\mathbf{u}$, $\mathbf{S}^{\dagger}$ is the adjoint wave equation operator and $\mathbf{R}^{\dagger}\delta\mathbf{d}$ is the adjoint source. The model perturbations or images (equation 3.14) can be expressed as follows

$$\delta\mathbf{m}^{*} = -\left(\frac{\partial\mathbf{S}}{\partial\mathbf{m}}\mathbf{u}\right)^{\dagger}\mathbf{p}, \tag{3.16}$$

or rewritten the implicit inner product over time explicitly

$$\delta\mathbf{m}(\mathbf{x})^{*} = -\int\left(\frac{\partial\mathbf{S}(\mathbf{x},t)}{\partial\mathbf{m}(\mathbf{x})}\mathbf{u}(\mathbf{x},t)\right)^{\dagger}\mathbf{p}(\mathbf{x},t)dt = -\int\mathbf{p}(\mathbf{x},t)^{T}\left(\frac{\partial\mathbf{S}(\mathbf{x},t)}{\partial\mathbf{m}(\mathbf{x})}\mathbf{u}(\mathbf{x},t)\right)dt, \tag{3.17}$$

where $\mathbf{p}(\mathbf{x},t)^{\dagger} = \mathbf{p}(\mathbf{x},t)^{T}$ was applied. This is the formulation of reverse time migration with the adjoint-state method (Lions, 1971; Tarantola, 1984a, 1988; Tromp et al., 2005). The estimated images for different shots are usually stacked to form a stacked image.

The abstract form adjoint-state equation 3.15 can be applied to the first-order velocity-stress

elastic wave equation system 3.1

$$\begin{pmatrix} \rho\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}^{\dagger} \left(\frac{\partial}{\partial t}\right)^{\dagger} \mathbf{p} - \begin{pmatrix} \mathbf{0} & \mathbf{D} \\ \mathbf{C}\mathbf{D}^T & \mathbf{0} \end{pmatrix}^{\dagger} \mathbf{p} = \mathbf{R}^{\dagger}\delta\mathbf{d}, \tag{3.18}$$

with zero final condition $\mathbf{p}\,|_{t=T} = \mathbf{0}$ and appropriate boundary conditions. The vector $\mathbf{p} = (\boldsymbol{\upsilon}, \boldsymbol{\varsigma})^T$, where $\boldsymbol{\upsilon} = (\upsilon_x, \upsilon_z)^T$ is the adjoint-state particle velocity field and $\boldsymbol{\varsigma} = (\varsigma_{xx}, \varsigma_{zz}, \varsigma_{xz})^T$ is the adjoint-state stress field. The reflection data $\delta\mathbf{d} = (\delta\mathbf{d}_v, \delta\mathbf{d}_\sigma)^T$ acts as the adjoint source to the adjoint-state equation. The differential operator is anti-self-adjoint $\left(\frac{\partial}{\partial x}\right)^{\dagger} = -\left(\frac{\partial}{\partial x}\right)$. Consequently, one can write the expression $\mathbf{D}^{\dagger} = -\mathbf{D}^T$. Finally, the adjoint-state equation can be rewritten as follows

$$\begin{pmatrix} \rho\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \left(-\frac{\partial}{\partial t}\right) \mathbf{p} + \begin{pmatrix} \mathbf{0} & \mathbf{D}\mathbf{C} \\ \mathbf{D}^T & \mathbf{0} \end{pmatrix} \mathbf{p} = \mathbf{R}^{\dagger}\delta\mathbf{d}. \tag{3.19}$$

The structure of the adjoint-state equation is slightly different to the structure of the state equation 3.1. However, the adjoint-state equation can be redefined into a form that resembles the state equation by adopting a transformation of variables (Vigh et al., 2014)

$$\tilde{\mathbf{p}} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix} \mathbf{p},$$

where $\tilde{\mathbf{p}} = (\tilde{\boldsymbol{\upsilon}}, \tilde{\boldsymbol{\varsigma}})^T$ is the transformed adjoint-state variable, $\tilde{\boldsymbol{\upsilon}} = (\tilde{\upsilon}_x, \tilde{\upsilon}_z)^T$ and $\tilde{\boldsymbol{\varsigma}} = (\tilde{\varsigma}_{xx}, \tilde{\varsigma}_{zz}, \tilde{\varsigma}_{xz})^T$. If we multiply both side of the adjoint-state equation 3.19 by the transformation matrix, the adjoint-state equation can be rewritten as follows

$$\begin{pmatrix} \rho\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \left(-\frac{\partial}{\partial t}\right) \tilde{\mathbf{p}} + \begin{pmatrix} \mathbf{0} & \mathbf{D} \\ \mathbf{C}\mathbf{D}^T & \mathbf{0} \end{pmatrix} \tilde{\mathbf{p}} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix} \mathbf{R}^{\dagger}\delta\mathbf{d}, \tag{3.20}$$

where the adjoint-state equation now has the same structure as the state equation 3.1. Consequently, the finite-difference code adopted to solve the forward equation system 3.1 and the Born modelling equation system 3.12 can be reused to compute the adjoint wavefield in equation 3.20. The only difference is that the source term is replaced by an elastic tensor scaled adjoint source and the finite-difference steps are in time reversal mode. After computing the transformed adjoint-state variable $\tilde{\mathbf{p}}$, the original adjoint-state variable $\mathbf{p}$ can be retrieved by the inverse transformation

$$\mathbf{p} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^{-1} \end{pmatrix} \tilde{\mathbf{p}} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^{-1} \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{\upsilon}} \\ \tilde{\boldsymbol{\varsigma}} \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{\upsilon}} \\ \mathbf{C}^{-1}\tilde{\boldsymbol{\varsigma}} \end{pmatrix}, \tag{3.21}$$

with

$$\mathbf{C}^{-1} = \begin{pmatrix} \frac{\lambda+2\mu}{4\lambda\mu+4\mu^2} & -\frac{\lambda}{4\lambda\mu+4\mu^2} & 0 \\ -\frac{\lambda}{4\lambda\mu+4\mu^2} & \frac{\lambda+2\mu}{4\lambda\mu+4\mu^2} & 0 \\ 0 & 0 & \frac{1}{\mu} \end{pmatrix}.$$

Using the migration operator formulation given by equation 3.17, elastic wave equation 3.1 and the adjoint wave equation 3.20, and variable transformation equation 3.21, the adjoint model perturbations or images ($\delta\mathbf{m}^* = (\delta\rho^*, \delta\lambda^*, \delta\mu^*)^T$) for a single shot can be expressed as

$$\begin{aligned}
\delta\rho^* &= -\int \dot{\mathbf{v}} \cdot \boldsymbol{\upsilon} dt = -\int \dot{\mathbf{v}} \cdot \tilde{\boldsymbol{\upsilon}} dt = -\int (\dot{v}_x \tilde{v}_x + \dot{v}_z \tilde{v}_z) dt \\
\delta\lambda^* &= \int \left(\frac{\partial \mathbf{C}}{\partial \lambda} \mathbf{D}^T \mathbf{v} \right) \cdot \boldsymbol{\varsigma} dt = \int \left(\frac{\partial \mathbf{C}}{\partial \lambda} \mathbf{C}^{-1} \dot{\boldsymbol{\sigma}} \right) \cdot \left(\mathbf{C}^{-1} \tilde{\boldsymbol{\varsigma}} \right) dt = \int \frac{(\dot{\sigma}_{xx} + \dot{\sigma}_{zz})(\tilde{\varsigma}_{xx} + \tilde{\varsigma}_{zz})}{4(\lambda+\mu)^2} dt \\
\delta\mu^* &= \int \left(\frac{\partial \mathbf{C}}{\partial \mu} \mathbf{D}^T \mathbf{v} \right) \cdot \boldsymbol{\varsigma} dt = \int \left(\frac{\partial \mathbf{C}}{\partial \mu} \mathbf{C}^{-1} \dot{\boldsymbol{\sigma}} \right) \cdot \left(\mathbf{C}^{-1} \tilde{\boldsymbol{\varsigma}} \right) dt \\
&= \int \left[\frac{\dot{\sigma}_{xz} \tilde{\varsigma}_{xz}}{\mu^2} + \frac{(\dot{\sigma}_{xx} + \dot{\sigma}_{zz})(\tilde{\varsigma}_{xx} + \tilde{\varsigma}_{zz})}{4(\lambda+\mu)^2} + \frac{(\dot{\sigma}_{xx} - \dot{\sigma}_{zz})(\tilde{\varsigma}_{xx} - \tilde{\varsigma}_{zz})}{4\mu^2} \right] dt,
\end{aligned} \tag{3.22}$$

where the over-dot means time derivative. The interaction of the forward wavefield (state variable) and backward wavefield (adjoint-state variable) requires access to the two wavefields at same time step. However, these two wavefields are computed in the reverse time direction. Naive methods such as saving either forward or backward wavefield to disk can be utilized. However, I/O can degrade computational performance for large-scale problems. In our work, we have adopted the source wavefield reconstruction method (Gauthier et al., 1986; Dussaud et al., 2008). During the forward simulation of state (source side) wavefield, only the wavefield within the depth of half of the spatial finite difference operator length on boundaries and the final time snapshots are saved in memory. Then, the state (source side) wavefield is recomputed from the saved wavefield by backward propagation while simultaneously computing the backward adjoint (receiver side) wavefield. This completes the derivation of the elastic Born operator and elastic RTM operator using continuous functional analysis and the time domain adjoint-state method.

3.2.4 Numerical adjointness of the elastic Born and RTM programs

We derived the elastic Born operator and elastic RTM operator using continuous functional analysis first and then discretize them. This is the so-called optimize-before-discretize approach (Borzi and Schulz, 2012). One must be careful when discretizing the forward and adjoint operators to guarantee that they truly behave like a forward and adjoint pair. In

particular, one needs to be attentive to scaling terms, source injection strategies, and rules for updating particle velocities and stresses. We adopted the dot product test to numerically evaluate how close the discretize adjoint operator is to the true adjoint of the forward operator (Mora, 1987a; Claerbout, 1992; Le, 2016). This implies generating data and model vectors of random numbers $\delta\mathbf{d}_1$ and $\delta\mathbf{m}_2$. Then, we evaluate $\delta\mathbf{d}_2 = \mathbf{L}\delta\mathbf{m}_2$ and $\delta\mathbf{m}_1 = \mathbf{L}^\dagger\delta\mathbf{d}_1$ and finally, we compute the closeness of inner products via the following expression

$$err = |\frac{\langle\delta\mathbf{d}_1, \delta\mathbf{d}_2\rangle_{\mathbb{D}} - \langle\delta\mathbf{m}_1, \delta\mathbf{m}_2\rangle_{\mathbb{M}}}{\langle\delta\mathbf{d}_1, \delta\mathbf{d}_2\rangle_{\mathbb{D}} + \langle\delta\mathbf{m}_1, \delta\mathbf{m}_2\rangle_{\mathbb{M}}}|\,.$$

Our code was written in single-precision *float* data type in C. The relative error of our dot product test is 10^{-3} for a model that consists of 500×500 samples in space and a single shot that consists of 5000 samples in time and 500 receivers.

3.2.5 Elastic least-squares reverse time migration

CGLS with adjoint-state method

From the above derivation, a properly designed elastic RTM code can be considered equivalent to the adjoint operator of the elastic Born forward modelling operator. The adjoint operator is an approximation to the inverse operator where the Hessian of the linearized inversion problem is replaced by an identity matrix. In other words, the migrated image obtained via the adjoint operator is a blurred version of the true subsurface image. The migrated image, in general, suffers from relative low resolution, unbalanced amplitudes due to geometric spreading and acquisition footprint. Moreover, multiparameter elastic migration will generate cross-talk among different components because different parameters are coupled. In order to estimate higher resolution images with properly balanced amplitudes and fewer cross-talk and artifacts, the elastic least-squares reverse time migration is formulated as a quadratic optimization problem where one minimizes the following cost function

$$J(\delta\mathbf{m}) = \frac{1}{2}\sum_{i=1}^{N_s} \|\mathbf{L}_i\delta\mathbf{m} - \delta\mathbf{d}_i\|_2^2, \tag{3.23}$$

where $\mathbf{L}_i$ is the Born approximation operator for the ith shot, $\delta\mathbf{d}_i$ is the reflection data associated to the ith shot gather, $\delta\mathbf{m}$ denotes model perturbation (elastic images), N_s indicates the number of shots and $\|\cdot\|_2$ indicates the ℓ_2 norm of vector. The optimal solution satisfies the condition $\frac{\partial J(\delta\mathbf{m})}{\partial\delta\mathbf{m}} = 0$. The latter leads to the normal equations

$$\left(\sum_{i=1}^{N_s}\mathbf{L}_i^\dagger\mathbf{L}_i\right)\delta\mathbf{m} = \sum_{i=1}^{N_s}\mathbf{L}_i^\dagger\delta\mathbf{d}_i, \tag{3.24}$$

where $\sum_{i=1}^{N_s} \mathbf{L}_i^\dagger \delta\mathbf{d}_i$ is the reverse time migrated image and $\sum_{i=1}^{N_s} \mathbf{L}_i^\dagger \mathbf{L}_i$ is the Hessian operator for least-squares migration. Last expression indicates that the migrated image is a blurred version of the true parameter perturbations. By applying the inverse of the Hessian operator, the raw migrated images can be deblurred. Explicit forming and inverting the Hessian is prohibitive expensive in terms of computational cost and memory requirements. Instead, we adopt an iterative method: the conjugate gradient least squares (CGLS) (Hestenes and Stiefel, 1952; Paige and Saunders, 1982). The CGLS only requires two operators $\mathbf{L}_i$ and $\mathbf{L}_i^\dagger$ that are applied "on the fly" to vectors. The operators are applied on vectors efficiently via the adjoint-state method. The CGLS algorithm can be safely used because our discretized numerical versions of $\mathbf{L}_i$ and $\mathbf{L}_i^\dagger$ passed the dot product test. A series of steps (CGLS iterations) are required to solve the quadratic optimization problem given by equation 3.23. The CGLS algorithm for elastic LSRTM is summarized as Algorithm 1.

Algorithm 1 CGLS algorithm

Initialize

$\delta\mathbf{m}^{(0)} = \mathbf{0}$

$\mathbf{r}_i^{(0)} = \delta\mathbf{d}_i,\ i = 1, \cdots, N_s$

$\mathbf{s}^{(0)} = \sum_{i=1}^{N_s} \mathbf{L}_i^\dagger \mathbf{r}_i^{(0)}$ `//compute the gradient`

$\mathbf{p}^{(0)} = \mathbf{s}^{(0)}$

$\gamma^{(0)} = \|\mathbf{s}^{(0)}\|_2^2$

for $k = 0, 1, \cdots$ **while** not converge **do**

$\mathbf{q}_i^{(k)} = \mathbf{L}_i \mathbf{p}^{(k)},\ i = 1, \cdots, N_s$ `//Born forward modeling`

$\delta^{(k)} = \sum_{i=1}^{N_s} \|\mathbf{q}_i^{(k)}\|_2^2$

$\alpha^{(k)} = \gamma^{(k)} / \delta^{(k)}$ `//calculate the step size`

$\delta\mathbf{m}^{(k+1)} = \delta\mathbf{m}^{(k)} + \alpha^{(k)} \mathbf{p}^{(k)}$ `//update the model`

$\mathbf{r}_i^{(k+1)} = \mathbf{r}_i^{(k)} - \alpha^{(k)} \mathbf{q}_i^{(k)},\ i = 1, \cdots, N_s$ `//compute data residuals`

$\mathbf{s}^{(k+1)} = \sum_{i=1}^{N_s} \mathbf{L}_i^\dagger \mathbf{r}_i^{(k+1)}$ `//compute the gradient`

$\gamma^{(k+1)} = \|\mathbf{s}^{(k+1)}\|_2^2$

$\beta^{(k)} = \gamma^{(k+1)} / \gamma^{(k)}$

$\mathbf{p}^{(k+1)} = \mathbf{s}^{(k+1)} + \beta^{(k)} \mathbf{p}^{(k)}$ `//compute the conjugate direction`

end

Parameterization

For reflection data full waveform inversion, P-wave impedance, S-wave impedance and density are the most suitable parameters to invert. The latter was confirmed by a detailed radiation pattern analysis for different parameterizations of elastic FWI (Tarantola, 1986). We parameterized our elastic LSRTM in terms of P-wave impedance perturbation δI_p and S-wave impedance perturbation δI_s. We have preferred to omit the inversion of the density

perturbation. We use the following relationships between elastic parameters

$$\lambda = \rho V_p^2 - 2\rho V_s^2, \quad \mu = \rho V_s^2,$$
$$I_p = \rho V_p, \quad I_s = \rho V_s,$$

in conjunction with total derivatives to write down the following parameter perturbation transformation

$$\begin{pmatrix} \delta\lambda \\ \delta\mu \end{pmatrix} = \begin{pmatrix} 2V_p & -4V_s \\ 0 & 2V_s \end{pmatrix} \begin{pmatrix} \delta I_p \\ \delta I_s \end{pmatrix}. \tag{3.25}$$

We insert the above parameter transformation into the elastic LSRTM formulation (equation 3.23). The parameter transformation matrix and its adjoint are incorporated into the CGLS algorithm (Algorithm 1). The change of parameters is similar to adding preconditioning to our system of equations. In our experience, inverting for impedance perturbations leads to an algorithm with faster convergence than inverting the Lamé parameter perturbations.

3.2.6 Preconditioning using multiparameter Hessian

The elastic LSRTM implicitly inverts the Hessian operator via CGLS iterations combined with adjoint-state method, i.e. apply Fréchet derivative and its adjoint on vectors via adjoint-state method on the fly in each CGLS iteration. It does not need to form the Fréchet derivative or Hessian explicitly. However, the CGLS algorithm may need a relatively large number of iterations to converge to an optimal solution. Preconditioning of gradients is important to accelerate the convergence of CGLS and saving computational resources. We investigate the structure of the Hessian operator of the elastic LSRTM and utilize the diagonal of pseudo-Hessian for preconditioning (Shin et al., 2001a). The Hessian operator of elastic LSRTM problem can be expressed as

$$\mathbf{H} = \sum_{i=1}^{N_s} \mathbf{L}_i^\dagger \mathbf{L}_i = \sum_{i=1}^{N_s} \left(\mathbf{R}\frac{\partial \mathbf{u}_i}{\partial \mathbf{m}}\right)^\dagger \left(\mathbf{R}\frac{\partial \mathbf{u}_i}{\partial \mathbf{m}}\right) = \sum_{i=1}^{N_s} \left(\mathbf{R}\mathbf{S}^{-1}\frac{\partial \mathbf{S}}{\partial \mathbf{m}}\mathbf{u}_i\right)^\dagger \left(\mathbf{R}\mathbf{S}^{-1}\frac{\partial \mathbf{S}}{\partial \mathbf{m}}\mathbf{u}_i\right), \tag{3.26}$$

where $\mathbf{u}_i$ is the source side wavefield for ith shot, and recall that $\mathbf{S}^{-1}$ is the Green's operator. If the Green's functions are not saved, explicitly computing the Hessian needs $N_s \cdot N_m$ forward simulations with N_m as the number of model grid points. Using the reciprocity of the Green's function (Tarantola, 1988), the number of forward simulations needed for explicitly computing the Hessian reduces to $N_s \cdot N_g$ where N_g denotes the number of receivers. The diagonal element of the Hessian is the zero-lag autocorrelation of the Fréchet derivative. It accounts for the geometric spreading effect (Shin et al., 2001b). Explicitly computing the diagonal of the Hessian also requires $N_s \cdot N_g$ forward simulations. Shin et al. (2001a) proposed to neglect the receiver Green's function to save computation cost. Under this

assumption, the Hessian can be simplified to the so-called "pseudo-Hessian"

$$\mathbb{H} = \sum_{i=1}^{N_s} \left(\frac{\partial \mathbf{S}}{\partial \mathbf{m}} \mathbf{u}_i \right)^{\dagger} \left(\frac{\partial \mathbf{S}}{\partial \mathbf{m}} \mathbf{u}_i \right). \tag{3.27}$$

For multiparameter problem, the Hessian and pseudo-Hessian are blockwise. The pseudo-Hessian for Lamé parameters is given by

$$\mathbb{H}_L = \begin{pmatrix} \mathbb{H}^{\lambda\lambda} & \mathbb{H}^{\lambda\mu} \\ \mathbb{H}^{\mu\lambda} & \mathbb{H}^{\mu\mu} \end{pmatrix}. \tag{3.28}$$

Using pseudo-Hessian operator expression (equation 3.27) and the elastic wave equation 3.1, the diagonal blocks can be expressed as

$$\begin{aligned} \mathbb{H}^{\lambda\lambda}(\mathbf{x}, \mathbf{x}') &= \sum_{i=1}^{N_s} \int \left(\frac{\partial \mathbf{C}}{\partial \lambda} \mathbf{D}^T \mathbf{v}(\mathbf{x}) \right) \cdot \left(\frac{\partial \mathbf{C}}{\partial \lambda} \mathbf{D}^T \mathbf{v}(\mathbf{x}') \right) dt, \\ \mathbb{H}^{\mu\mu}(\mathbf{x}, \mathbf{x}') &= \sum_{i=1}^{N_s} \int \left(\frac{\partial \mathbf{C}}{\partial \mu} \mathbf{D}^T \mathbf{v}(\mathbf{x}) \right) \cdot \left(\frac{\partial \mathbf{C}}{\partial \mu} \mathbf{D}^T \mathbf{v}(\mathbf{x}') \right) dt, \end{aligned} \tag{3.29}$$

where the dependence on shot index i of particle velocity field is omitted to avoid notation clutter. The diagonal terms of the pseudo-Hessian ($\mathbf{x} = \mathbf{x}'$) for Lamé parameters can be expressed as follows

$$\begin{aligned} \mathbb{H}^{\lambda\lambda}(\mathbf{x}, \mathbf{x}) &= \sum_{i=1}^{N_s} \int \frac{(\dot{\sigma}_{xx} + \dot{\sigma}_{zz})^2}{2(\lambda + \mu)^2} dt, \\ \mathbb{H}^{\mu\mu}(\mathbf{x}, \mathbf{x}) &= \sum_{i=1}^{N_s} \int \left[\frac{\dot{\sigma}_{xz}^2}{\mu^2} + \frac{(\dot{\sigma}_{xx} + \dot{\sigma}_{zz})^2}{2(\lambda + \mu)^2} + \frac{(\dot{\sigma}_{xx} - \dot{\sigma}_{zz})^2}{2\mu^2} \right] dt. \end{aligned} \tag{3.30}$$

Using the relationships between elastic parameters and chain rule, the pseudo-Hessian for P- and S- wave impedances is given by

$$\mathbb{H}_I = \begin{pmatrix} \mathbb{H}^{I_p I_p} & \mathbb{H}^{I_p I_s} \\ \mathbb{H}^{I_s I_p} & \mathbb{H}^{I_s I_s} \end{pmatrix} = \begin{pmatrix} 2V_p & 0 \\ -4V_s & 2V_s \end{pmatrix} \begin{pmatrix} \mathbb{H}^{\lambda\lambda} & \mathbb{H}^{\lambda\mu} \\ \mathbb{H}^{\mu\lambda} & \mathbb{H}^{\mu\mu} \end{pmatrix} \begin{pmatrix} 2V_p & -4V_s \\ 0 & 2V_s \end{pmatrix}. \tag{3.31}$$

We use the inverse of the diagonal of the pseudo-Hessian for preconditioning. The diagonal

terms of the pseudo-Hessian for P- and S-wave impedances can be expressed as follows

$$\begin{aligned}\mathbb{H}^{I_pI_p}(\mathbf{x},\mathbf{x}) &= \sum_{i=1}^{N_s}\int 2V_p^2\frac{(\dot{\sigma}_{xx}+\dot{\sigma}_{zz})^2}{(\lambda+\mu)^2}dt,\\ \mathbb{H}^{I_sI_s}(\mathbf{x},\mathbf{x}) &= \sum_{i=1}^{N_s}\int 4V_s^2\left[\frac{\dot{\sigma}_{xz}^2}{\mu^2}+\frac{(\dot{\sigma}_{xx}+\dot{\sigma}_{zz})^2}{2(\lambda+\mu)^2}+\frac{(\dot{\sigma}_{xx}-\dot{\sigma}_{zz})^2}{2\mu^2}\right]dt.\end{aligned} \tag{3.32}$$

The preconditioned version of elastic LSRTM minimizes

$$J(\delta\tilde{\mathbf{m}}) = \frac{1}{2}\sum_{i=1}^{N_s}\|\mathbf{L}_i\mathbf{T}\mathbf{P}\delta\tilde{\mathbf{m}}-\delta\mathbf{d}_i\|_2^2, \tag{3.33}$$

where $\mathbf{P}$ denotes the inverse of the diagonal of pseudo-Hessian for P- and S-wave impedances, $\mathbf{T}$ denotes the parameter transformation matrix in equation 3.25. As discussed in last section, the parameter transformation matrix $\mathbf{T}$ and its adjoint also play the role of preconditioning to our system of equations. The preconditioned conjugate gradient least squares (PCGLS) algorithm (Bjorck, 1996) can be summarized as Algorithm 2. The output of Algorithm 2 is the inverted Lamé parameter perturbations $\delta\hat{\mathbf{m}} = (\delta\hat{\lambda},\delta\hat{\mu})^T = \mathbf{T}\mathbf{P}\delta\hat{\tilde{\mathbf{m}}}$. And the inverted P- and S-wave impedance perturbations can be retrieved by $(\delta\hat{I}_p,\delta\hat{I}_s)^T = \mathbf{T}^{-1}(\delta\hat{\lambda},\delta\hat{\mu})^T$.

Algorithm 2 Preconditioned CGLS algorithm

Initialize

$\delta\mathbf{m}^{(0)} = \mathbf{0}$

$\mathbf{r}_i^{(0)} = \delta\mathbf{d}_i,\ i = 1,\cdots,N_s$

$\mathbf{s}^{(0)} = \mathbf{P}^\dagger\mathbf{T}^\dagger\left(\sum_{i=1}^{N_s}\mathbf{L}_i^\dagger\mathbf{r}_i^{(0)}\right)$ `//compute the preconditioned gradient`

$\mathbf{p}^{(0)} = \mathbf{s}^{(0)}$

$\gamma^{(0)} = \|\mathbf{s}^{(0)}\|_2^2$

for $k = 0,1,\cdots$ **while** not converge **do**

$\mathbf{t}^{(k)} = \mathbf{T}\mathbf{P}\mathbf{p}^{(k)}$,

$\mathbf{q}_i^{(k)} = \mathbf{L}_i\mathbf{t}^{(k)},\ i = 1,\cdots,N_s$ `//Born forward modeling`

$\delta^{(k)} = \sum_{i=1}^{N_s}\|\mathbf{q}_i^{(k)}\|_2^2$

$\alpha^{(k)} = \gamma^{(k)}/\delta^{(k)}$ `//calculate the step size`

$\delta\mathbf{m}^{(k+1)} = \delta\mathbf{m}^{(k)} + \alpha^{(k)}\mathbf{t}^{(k)}$ `//update the model`

$\mathbf{r}_i^{(k+1)} = \mathbf{r}_i^{(k)} - \alpha^{(k)}\mathbf{q}_i^{(k)},\ i = 1,\cdots,N_s$ `//compute data residuals`

$\mathbf{s}^{(k+1)} = \mathbf{P}^\dagger\mathbf{T}^\dagger\left(\sum_{i=1}^{N_s}\mathbf{L}_i^\dagger\mathbf{r}_i^{(k+1)}\right)$ `//compute the preconditioned gradient`

$\gamma^{(k+1)} = \|\mathbf{s}^{(k+1)}\|_2^2$

$\beta^{(k)} = \gamma^{(k+1)}/\gamma^{(k)}$

$\mathbf{p}^{(k+1)} = \mathbf{s}^{(k+1)} + \beta^{(k)}\mathbf{p}^{(k)}$ `//compute the conjugate direction`

end

3.3 Examples

The proposed method was tested on two synthetic models: the elastic version of the Camembert model (Gauthier et al., 1986) and the elastic Marmousi2 model (Martin et al., 2006). All the "observed data" are generated with time-domain elastic staggered-grid finite-difference method. In other words, we have NOT committed the so-called "inverse crime" that entails using the linearized Born modelling operator to generate data to test least squares migration. The same staggered-grid finite-difference code was used for the elastic LSRTM inversion. The C-PML boundary condition was applied on the top of the model. The observed data were assumed to be vector particle velocity fields. Only the direct wave was muted from the data. The data contain internal multiples that are not honoured by the linearized Born modelling. In real data applications, the multiples can be attenuated from the data to only keep first-order scattered energy. The code for our numerical examples was written in C and parallelized with Message Passing Interface (MPI).

3.3.1 Elastic Camembert model

The elastic LSRTM is tested on a synthetic elastic version of Camembert model. This test shows that elastic LSRTM can attenuate cross-talk between P-wave and S-wave impedance perturbations. Figure 3.1 a and b show the true P- and S-wave velocity models. The velocity anomalies are embedded in two layered models. The velocity anomalies for P and S are in different positions. Density is assumed to be constant (1500 kg/m^3). The model has a dimension of 2.5 km in horizontal axis and 1.5 km in depth with 501×301 grid points. There are 101 shots and 501 receivers that simulate a fixed-spread survey geometry. The shot interval is 25 m and receiver interval is 5 m. The shot depth is 5 m and the receiver depth is 10 m. A 20 Hz central frequency Ricker wavelet is used to simulate an explosive source. The "observed data" are simulated using our elastic finite-difference code. No other pre-processing was applied to the data except for muting the direct wave. The "observed data" are shown in Figure 3.2. Figure 3.1 c and d show the smoothed background velocity models for elastic RTM and elastic LSRTM. Smoothed models were obtained by convolving the true models with a 2D Gaussian function of 50 m width with standard deviation as half the width. The width of the 2D Gaussian function is approximately equal to the shortest P-wave wavelength. Figure 3.1 e and f show the true P-wave and S-wave impedance perturbations with respect to the background models.

The results of elastic RTM are shown in Figure 3.3 a and b. The elastic RTM operator generates high-amplitude low-frequency artifacts caused by the cross-correlation of head wave, diving wave and backscattered internal reflections. A Laplacian filter (Youn and

Zhou, 2001) was used to attenuate the artifacts. As expected, there is cross-talk between P-wave and S-wave impedance perturbations in the elastic RTM images. Elastic LSRTM (Figure 3.3 c and d) not only reduces multiparameter cross-talk but also displays fewer artifacts, properly balanced amplitudes and higher resolution. To make a fair comparison with the elastic RTM images (Figure 3.3 a and b), the least-squares inverted images were post-processed by Laplacian filtering. No filters were applied during the inversion process. These results were computed after 82 iterations of elastic LSRTM. The relative data misfit percentage reduces to about 6%. The relative data misfit is defined as $\frac{\sum_{i=1}^{N_s} \|\mathbf{L}_i \delta\hat{\mathbf{m}} - \delta\mathbf{d}_i\|_2^2}{\sum_{i=1}^{N_s} \|\delta\mathbf{d}_i\|_2^2}$.

Figure 3.3 e and f show the pseudo-Hessian preconditioned elastic LSRTM after 20 iterations. The relative data misfit also reduces to about 6%. The results are similar to unpreconditioned elastic LSRTM. Figure 3.4 compares the convergence curves of the unpreconditioned and preconditioned version of elastic LSRTM. From this figure one can observe that preconditioned elastic LSRTM converges much faster than unpreconditioned LSRTM.

3.3.2 Sensitivity to background model error

We tested the proposed elastic LSRTM using background models with different degree of spatial smoothing. The setup is equivalent to the setup adopted in the last section (elastic Camembert model test). We have run our algorithm using background models with increasing degree of smoothing. The background velocity models were smoothed via 2D Gaussian functions of width (W) 50 m, 100 m, 150 m and 200 m. The standard derivation of the 2D Gaussian function equals to its half width. Results for $W = 50$ m smoothing have already been shown in the previous section. We only plot the models and results for smoothing with 100 m, 150 m and 200 m widths. Figure 3.5 shows background models using different levels of smoothing. The comparison of the data misfit convergence curves of elastic LSRTM using the four background models is shown in Figure 3.6. We observe that the level of smoothing of the background models influences the data misfit. The elastic LSRTM for different background models converge to different levels of data misfit. We have compared the results at fixed number of iterations (82 iterations) that all misfit curves have converged. The inverted results are shown in Figure 3.7. The quality of inverted images degrades and more artifacts are present in the images as the level of smoothing increases.

3.3.3 Elastic Marmousi2 model

The proposed method was also tested on a complex elastic model. To this end, we adopted the elastic Marmousi2 model (Martin et al., 2006). The model consists of a total of 199 layers with a steep anticline fault zone. The size of the original model was reduced to 1001×426

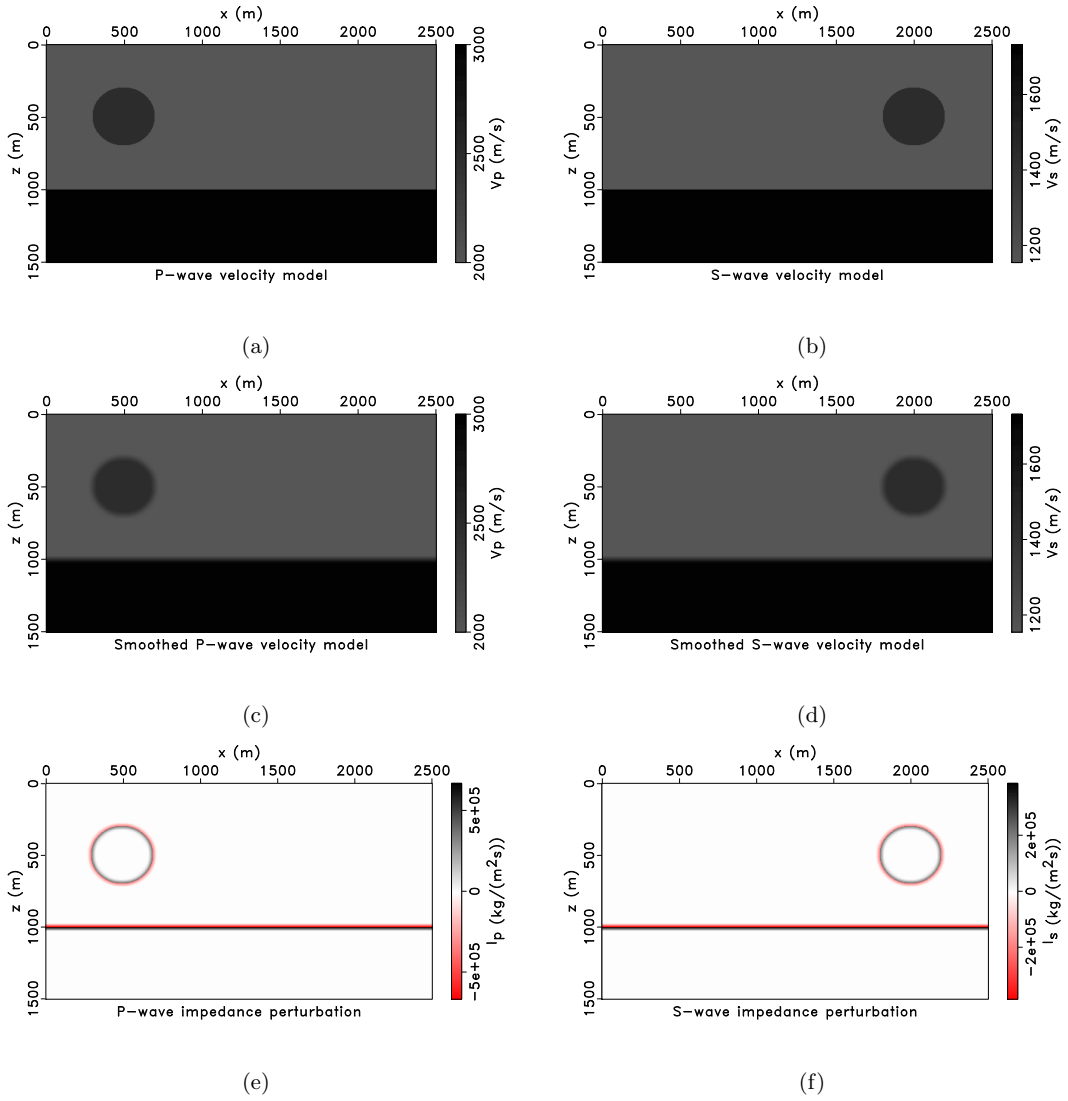
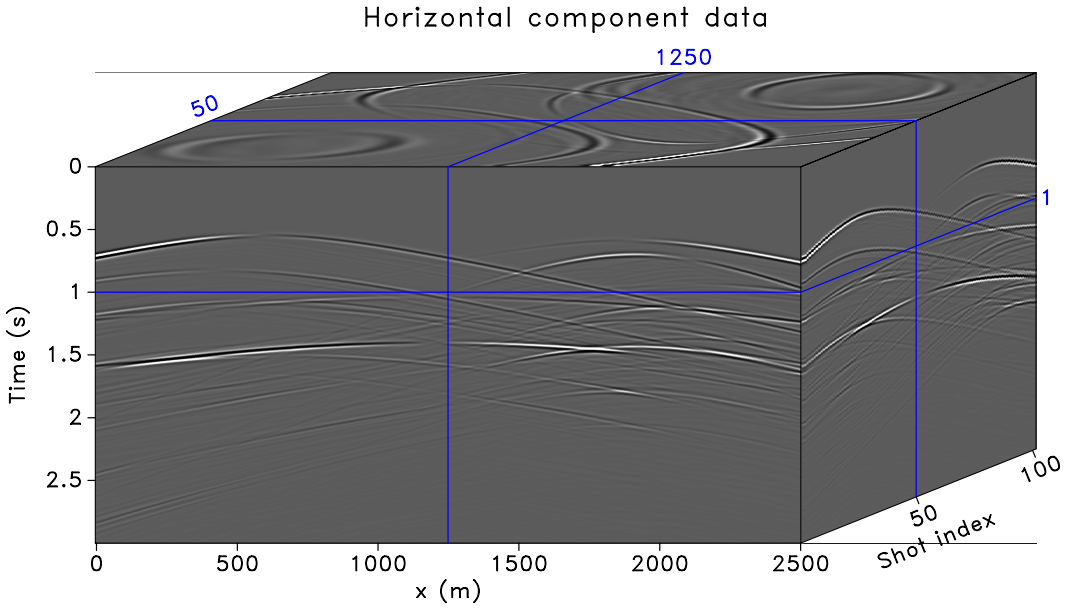
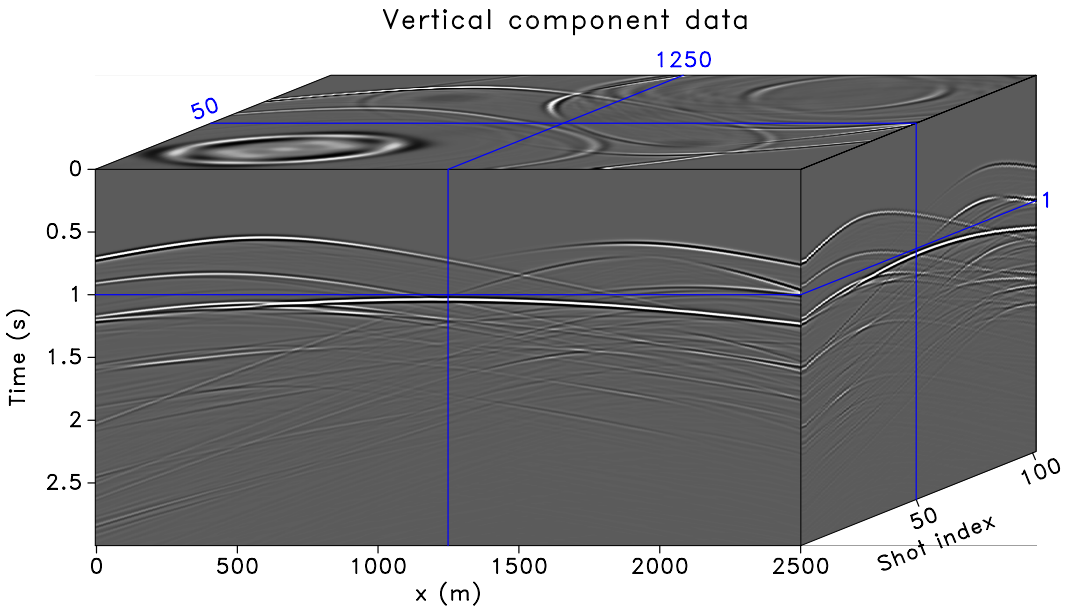


Figure 3.1: Elastic Camembert model. (a) P-wave velocity model. (b) S-wave velocity model. (c) Smoothed P-wave velocity model. (d) Smoothed S-wave velocity model. (e) True P-wave impedance perturbation. (f) True S-wave impedance perturbation.



(a)



(b)

Figure 3.2: Prestack multicomponent data for elastic Camembert model. (a) Horizontal particle velocity data. (b) Vertical particle velocity data.

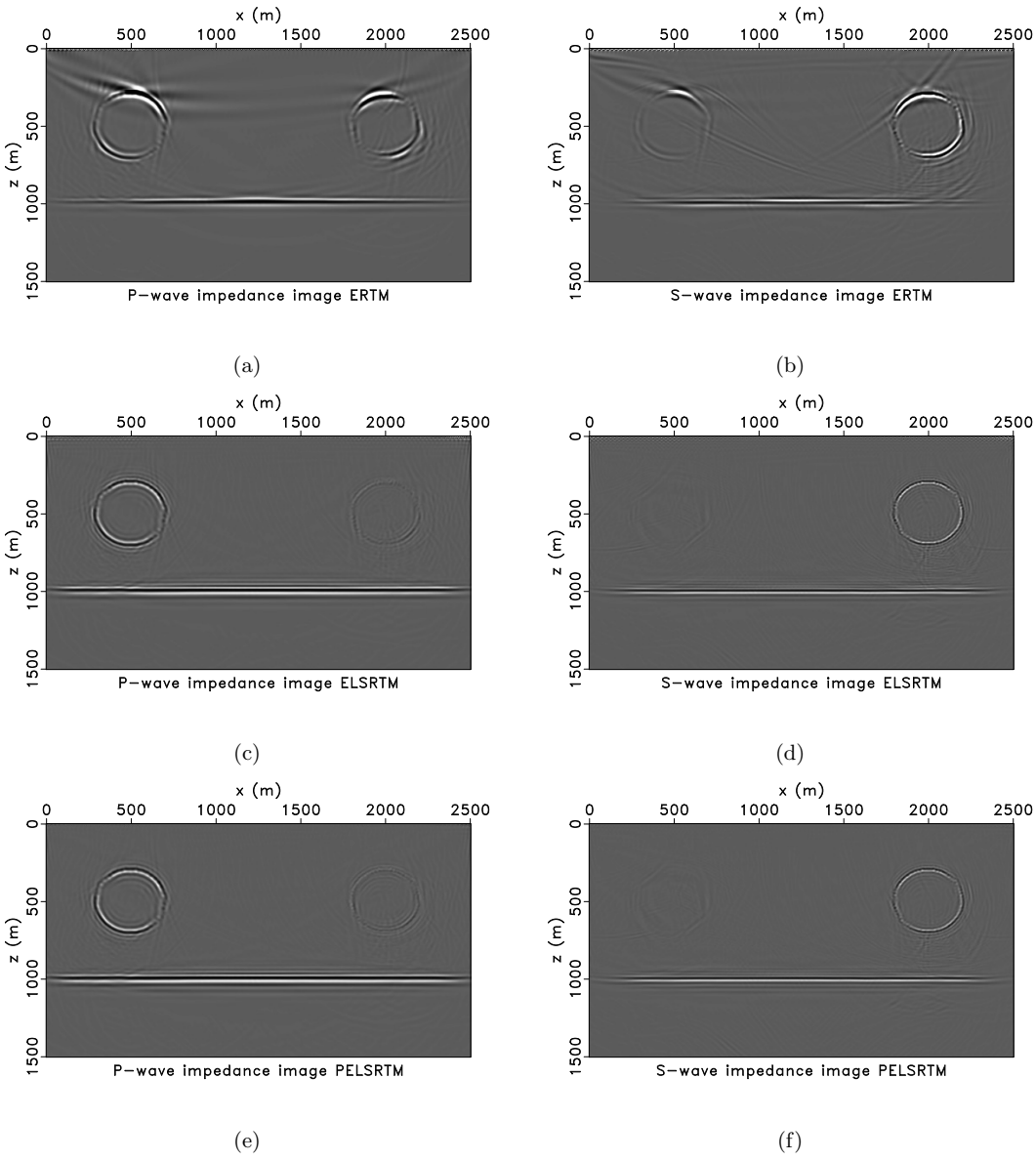


Figure 3.3: (a) P-wave impedance perturbation image estimated via elastic RTM. (b) S-wave impedance perturbation estimated via elastic RTM. (c) P-wave impedance perturbation image estimated via elastic LSRTM. (d) S-wave impedance perturbation image estimated via elastic LSRTM. (e) P-wave impedance perturbation image estimated via preconditioned elastic LSRTM. (f) S-wave impedance perturbation image estimated via preconditioned elastic LSRTM.

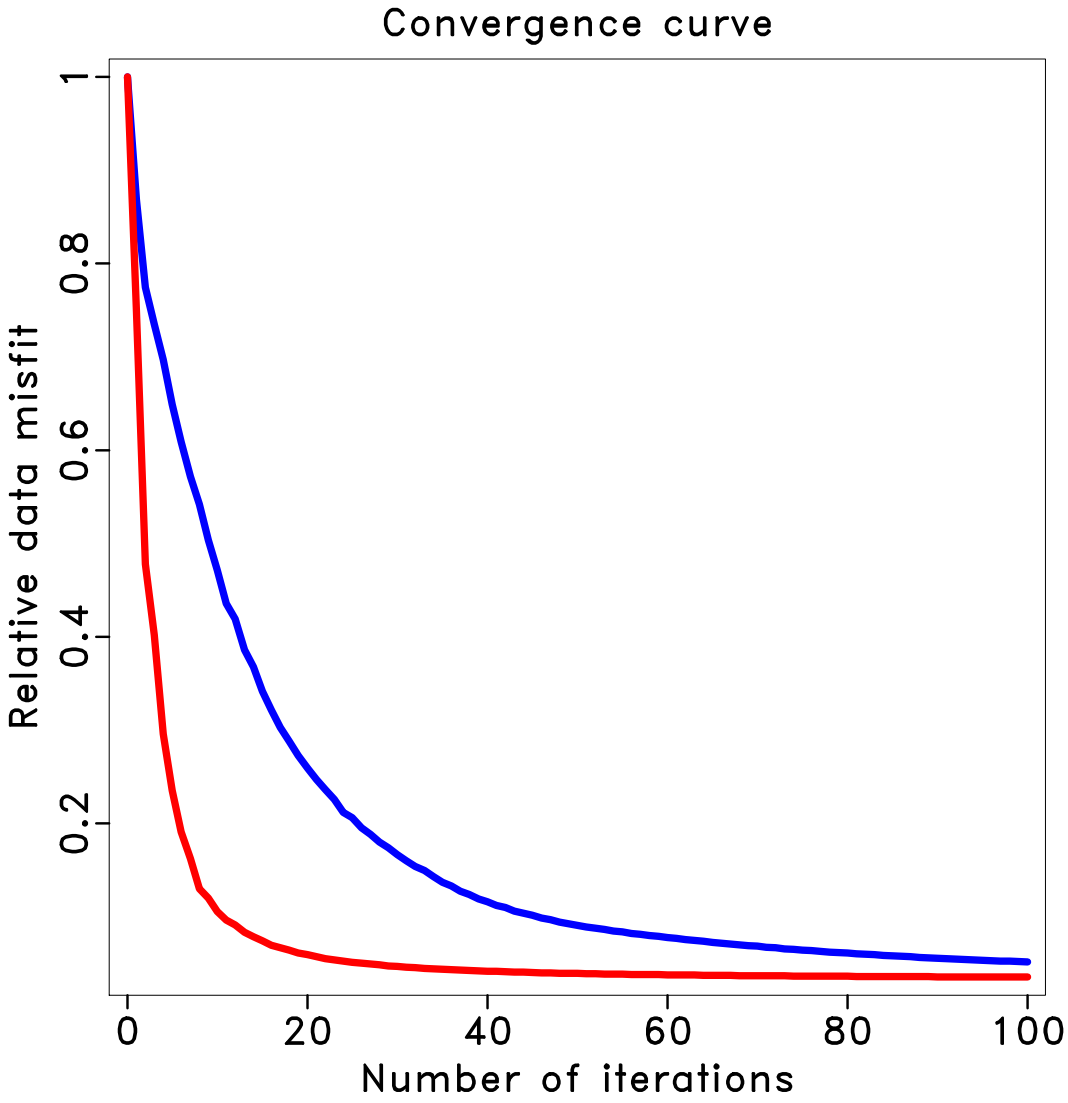


Figure 3.4: Comparison of relative data misfit convergence curves for elastic LSRTM (blue) and preconditioned elastic LSRTM (red) for Camembert model.

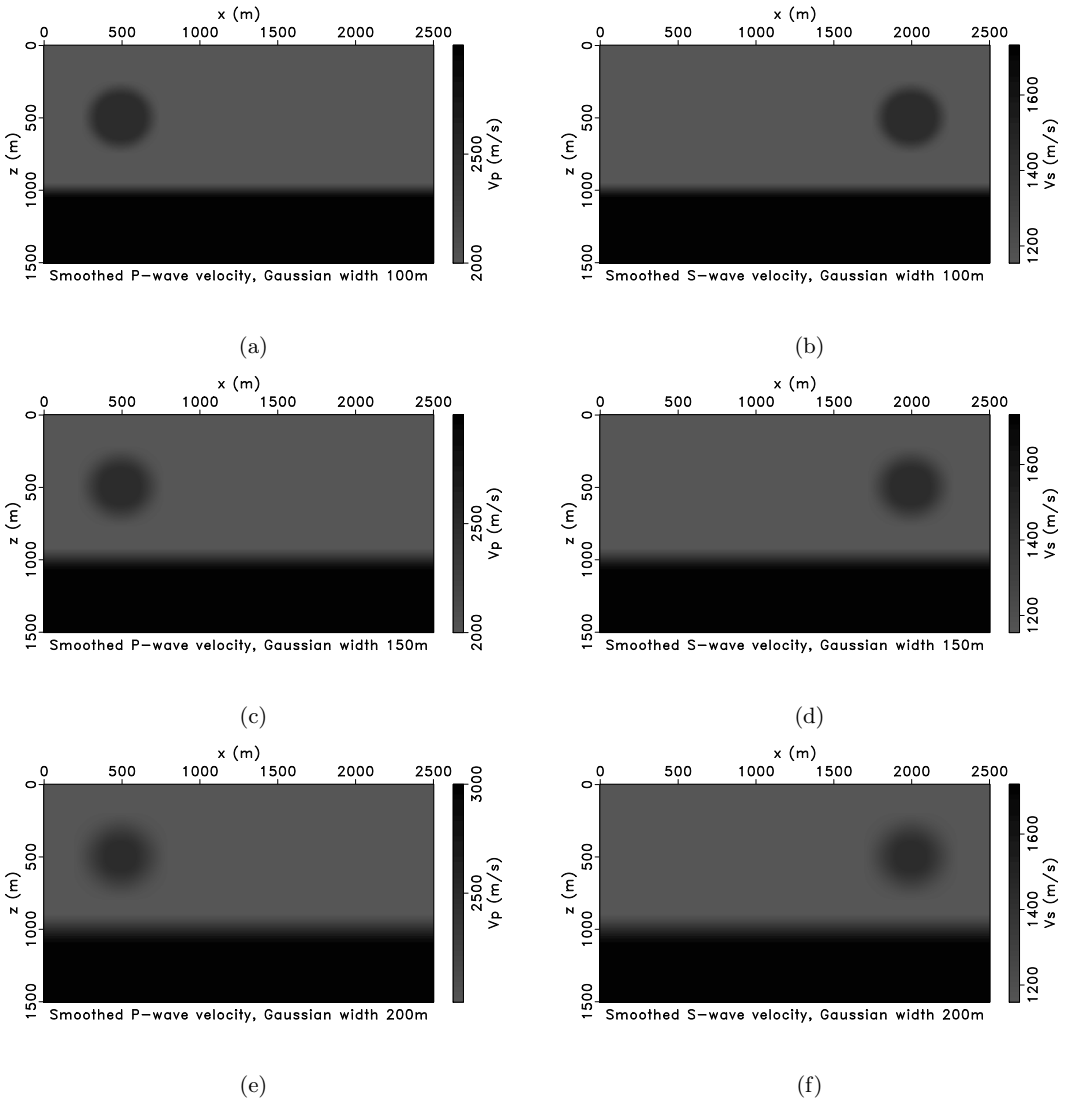


Figure 3.5: Analysis of the influence of the background models on elastic LSRTM. (a)-(b) Smoothed P- and S-wave velocity models using a 2D Gaussian function of width $W = 100$ m. (c)-(d) Smoothed P- and S-wave velocity models via 2D Gaussian function of width $W = 150$ m. (e)-(f) Smoothed P- and S-wave velocity models using a 2D Gaussian function of $W = 200$ m.

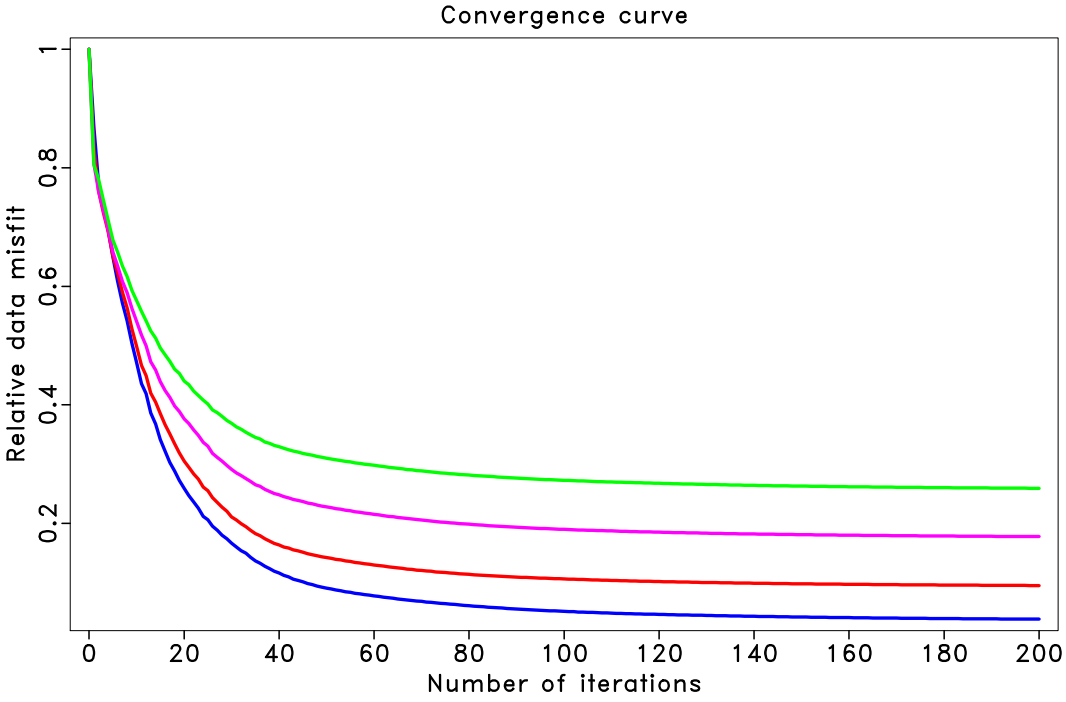


Figure 3.6: Comparison of relative data misfit convergence curves for elastic LSRTM for background models with different degree of smoothing: $W = 50$ m (Blue), $W = 100$ m (Red), $W = 150$ m (Purple) and $W = 200$ m (Green).

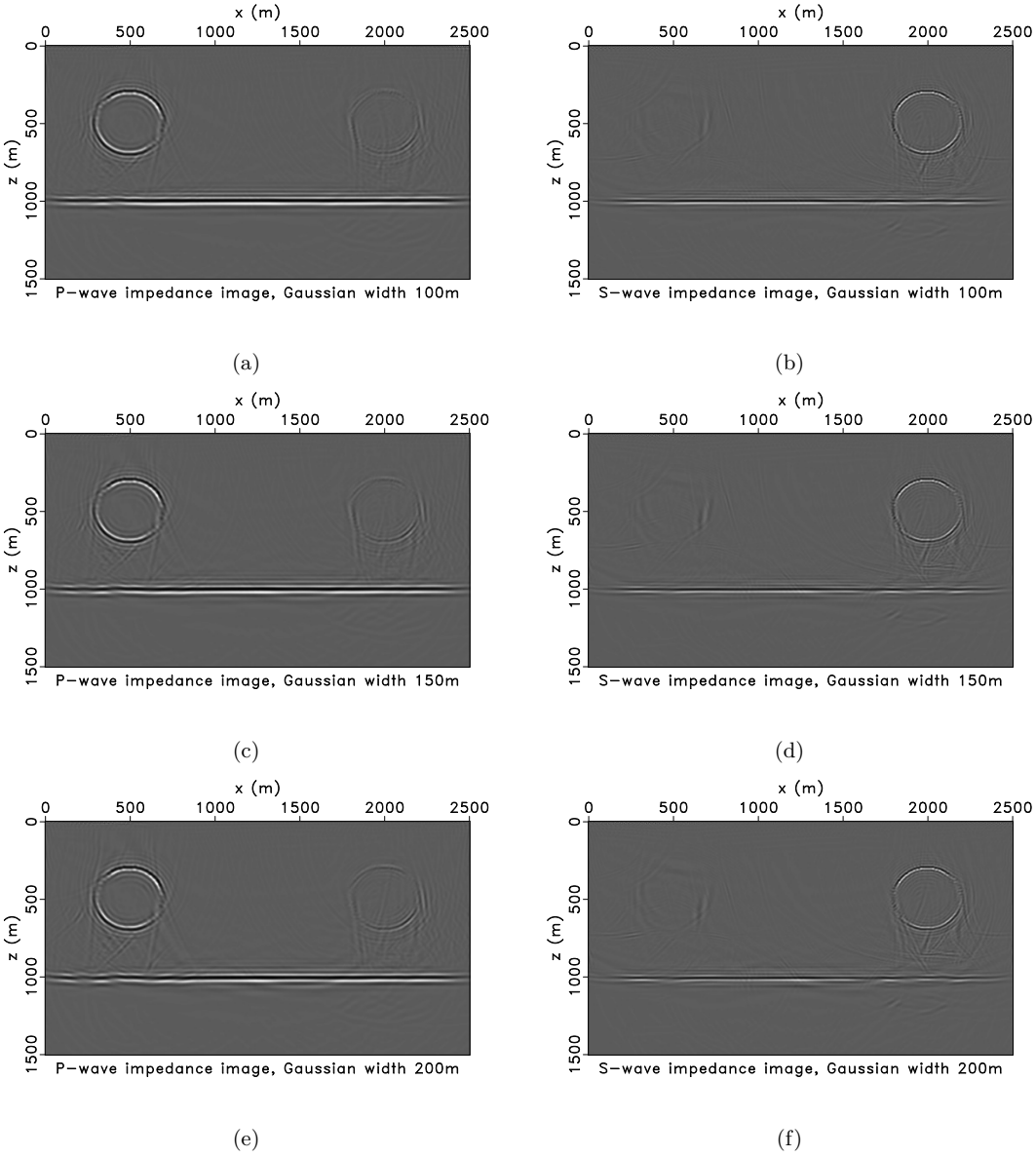


Figure 3.7: Comparison of elastic LSRTM results for background models with increasing degree of smoothing. (a)-(b) Inverted P- and S-wave impedance perturbation images using background models smoothed with a 2D Gaussian function of width $W = 100$ m. (c)-(d) Inverted P- and S-wave impedance perturbation images using background models smoothed with a 2D Gaussian function of width $W = 150$ m. (e)-(f) Inverted P- and S-wave impedance perturbation images using background models smoothed with a 2D Gaussian function of width $W = 200$ m.

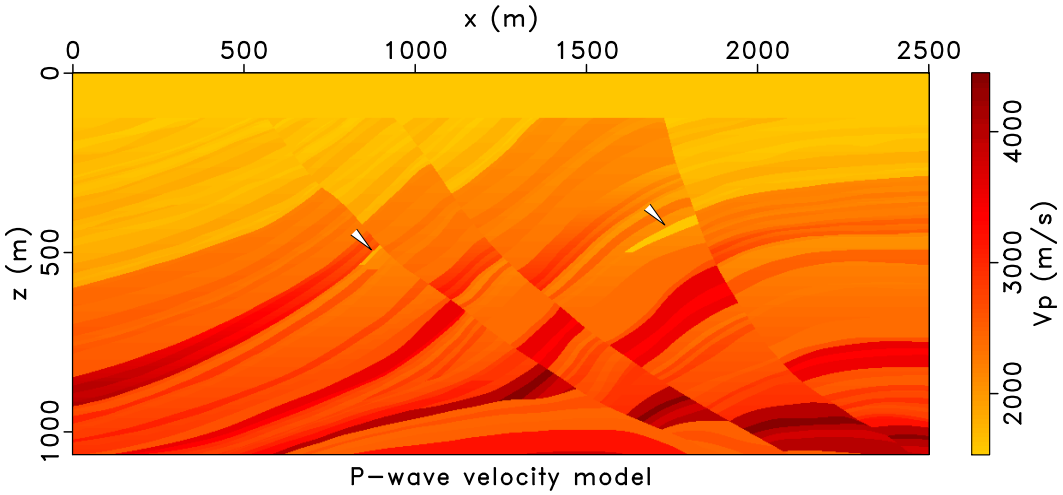
grid points to decrease the turnaround time of our tests. The water layer in the original Marmousi2 model was removed and replaced by a low-velocity layer to simulate a purely elastic model. Figure 3.8 shows the modified elastic Marmousi2 P- and S-wave velocity models. In the steep fault zone, there are two hydrocarbon reservoirs around depth 500 m that have decreased P-wave velocity and a small change in the S-wave velocity (indicated by white triangles). This uncorrelated P- and S-wave structure will cause cross-talk in elastic RTM images. We will show that the elastic LSRTM can attenuate the cross-talk. Density is assumed to be constant (2000 kg/m^3). The model has a dimension of 2500 m in horizontal distance and 1062.5 m in depth. A land acquisition geometry is simulated with 101 shots and 1001 receivers distributed on the surface of the earth. The shot interval is 25 m and receiver interval is 2.5 m. The central frequency of source function (Ricker wavelet) is 35 Hz. The "observed data" were simulated with our finite-difference code and the direct wave was removed from the observed data. The "observed data" are shown in Figure 3.9. Figure 3.10 shows the smoothed background velocity models for elastic RTM and elastic LSRTM. Smoothing was accomplished by convolving the true velocity models with a 35 m width 2D Gaussian function. Figure 3.11 shows the true P-wave and S-wave impedance perturbations with respect to the smoothed background models. From this figure, we can also observe that the P- and S-wave models are inconsistent in the two hydrocarbon reservoirs region at about 500 m depth in the steep fault zone (indicated by white triangles). Figure 3.12 shows the P- and S- wave impedance perturbation images obtained via elastic RTM. The elastic RTM algorithm has successfully imaged the geological structures. However, the amplitudes of the elastic images are unbalanced. Uncollapsed energy artifacts caused by not having a dense distribution of sources and limited aperture are also visible. Most important, the elastic RTM operator has generated cross-talk between P and S images in the two hydrocarbon reservoir areas because the P- and S-wave velocity structures are different. These problems are caused by the fact that the elastic RTM operator is an adjoint operator as opposed to an ideal inverse operator. The elastic least-squares reverse time migration (Figure 3.13) can solve these problems. Results for elastic LSRTM were computed after 98 iterations. The relative data misfit percentage reduces to about 40% (Figure 3.15). The geological structure of the elastic Marmousi2 model is complex. The data generated by finite-difference modelling contain internal multiples that are not honoured by the linearized Born modelling operator. We believe that the latter explains the inability of the algorithm to reduce the data misfit further. The elastic LSRTM corrected the unbalanced amplitudes and suppressed the low-frequency RTM artifacts and artifacts caused by limited aperture. The elastic LSRTM also generates high-resolution images. More important, the elastic LSRTM can successfully decouple elastic parameters and suppress multiparameter cross-talk in areas with hydrocarbon traps in both P and S impedance perturbation images. These benefits are the result of the embedded de-blurring process that is associated with the inversion of

the Hessian operator in elastic LSRTM.

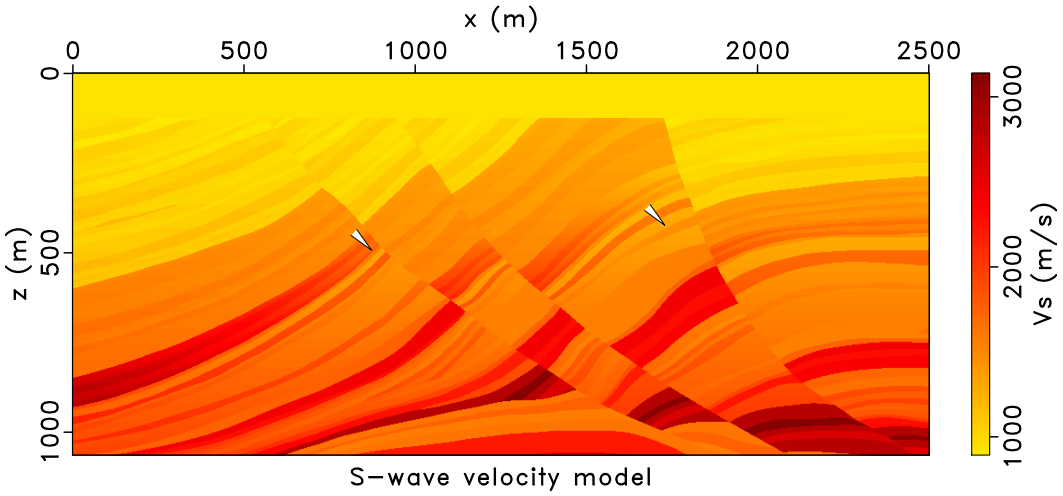
Figure 3.14 shows the pseudo-Hessian preconditioned elastic LSRTM after 58 iterations. The relative data misfit also reduces to about 40%. The preconditioned version of the elastic LSRTM yielded more amplitude balanced images than the unpreconditioned elastic LSRTM. To finalize our analysis, we also provide Figure 3.15 where we compared convergence curves for the unpreconditioned and preconditioned elastic LSRTM. Preconditioning with the pseudo-Hessian has lead to a visible improvement in convergence.

3.4 Conclusions

Elastic least-squares reverse time migration is formulated as a linearized elastic waveform inversion problem. The inversion is parameterized in terms of P-wave impedance perturbation and S-wave impedance perturbation. The formulations of elastic Born approximation operator and elastic reverse time migration operator are derived from the time-domain continuous adjoint-state method. The adjoint-state equation system is the same as the state equation system. The only difference is the replacement of an explosive source to an adjoint source. After developing the functional formulations for our forward and adjoint operators, we have discretized the elastic Born and RTM operators. The numerical discretized versions of the two operators pass the dot product test. This allows us to use the conjugate gradient least squares (CGLS) algorithm for solving the least-squares optimization problem. The Hessian is implicitly inverted via the adjoint-state method combined with conjugate gradient least squares algorithm. We investigated adopting the diagonal of the pseudo-Hessian operator to precondition the elastic LSRTM and thereby, to accelerate its convergence. The elastic LSRTM produces high-resolution images with fewer artifacts and more balanced amplitudes than elastic RTM. More important, elastic LSRTM can reduce cross-talk artifacts between P- and S-wave impedance perturbations that are present in elastic RTM images. In essence, the off-diagonal elements of the Hessian operator are attenuated by the inversion process. The pseudo-Hessian preconditioning operator adopted in our work not only accelerates the convergence of the elastic LSRTM but also improves the overall amplitude response of our images.

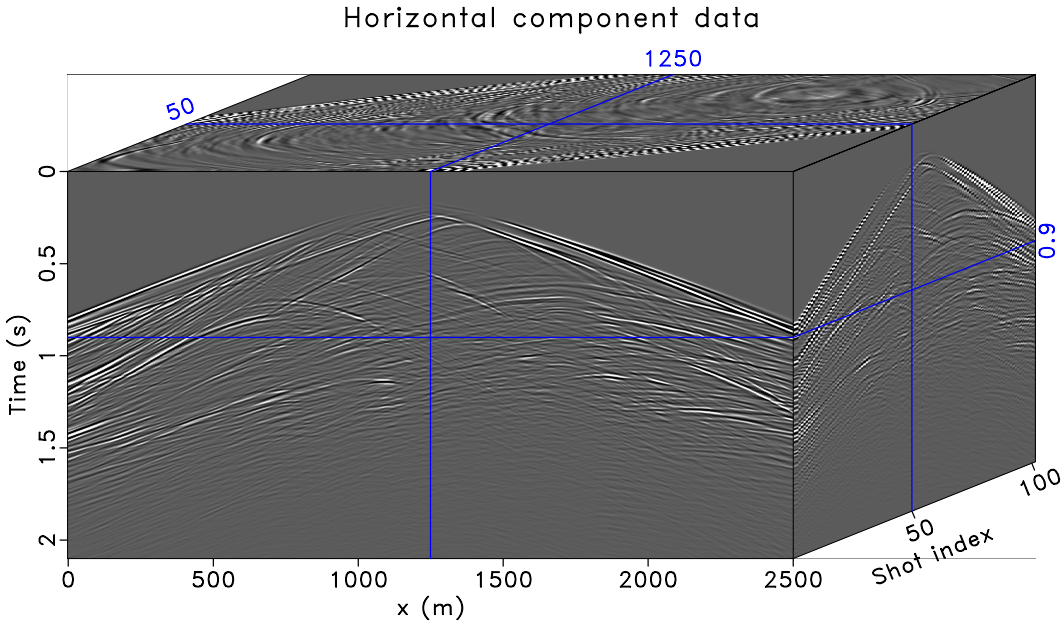


(a)

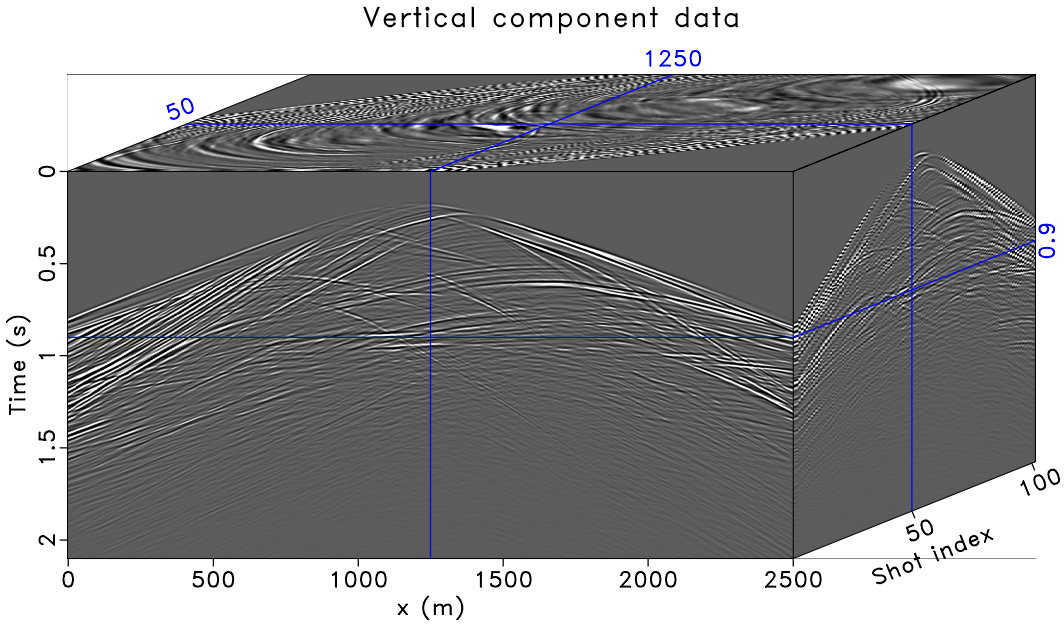


(b)

Figure 3.8: Elastic Marmousi2 model. (a) P-wave velocity model. (b) S-wave velocity model.

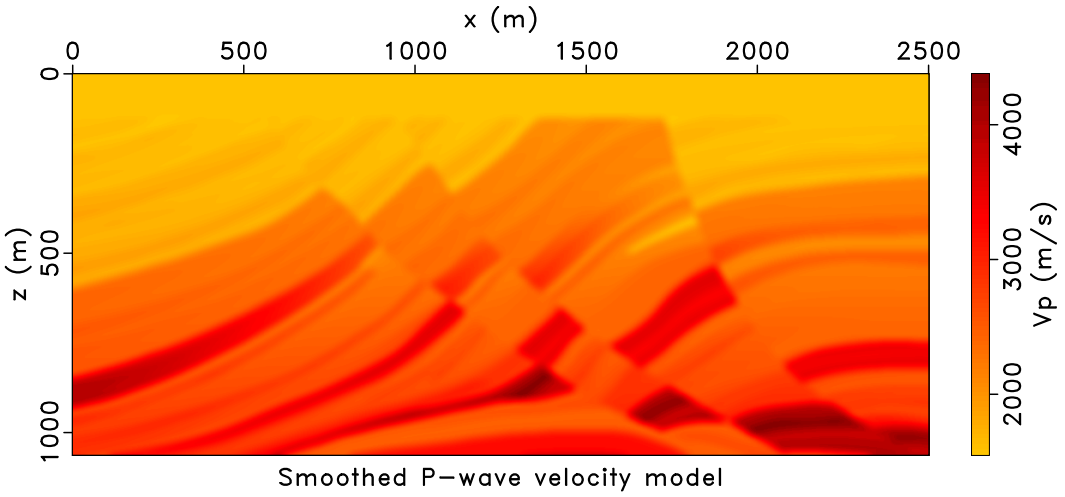


(a)

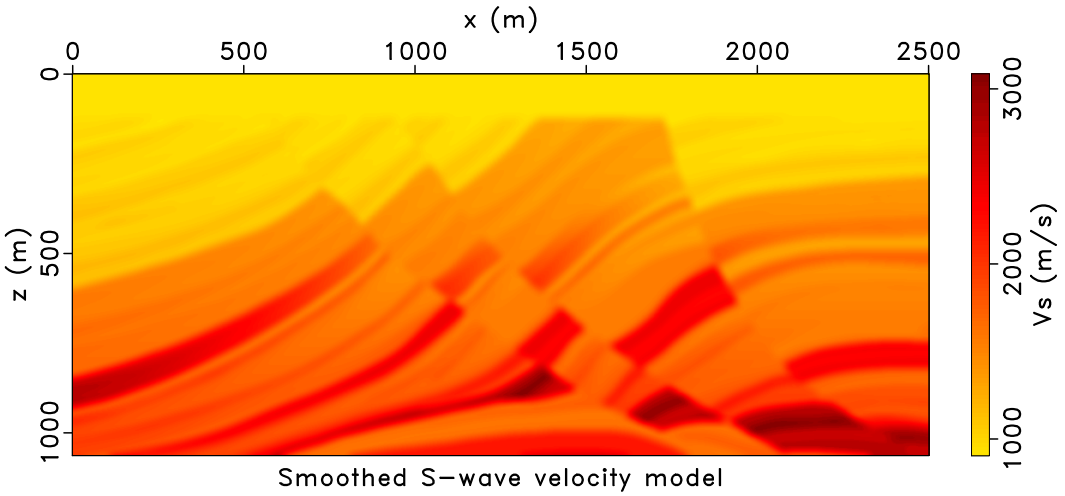


(b)

Figure 3.9: Prestack multicomponent data for elastic Marmousi2 model. (a) Horizontal particle velocity data. (b) Vertical particle velocity data.

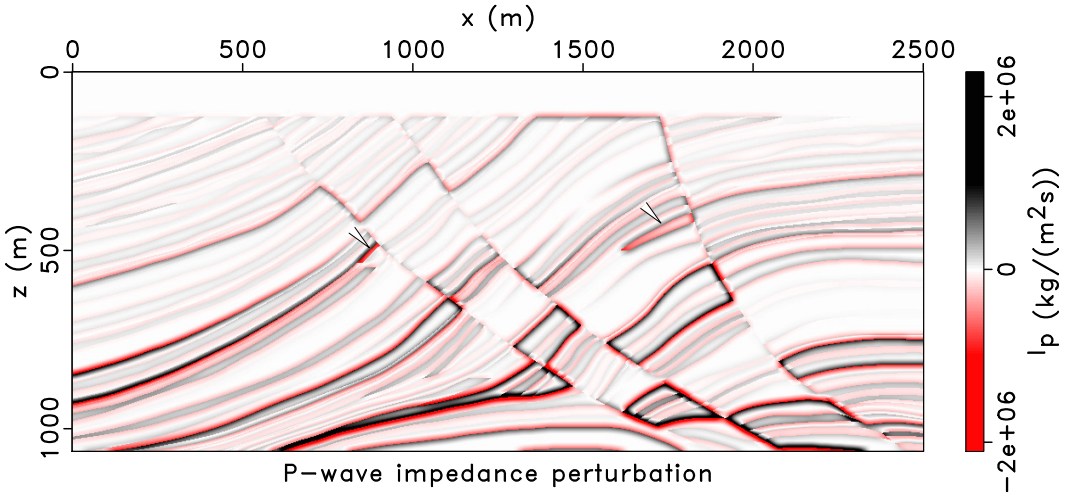


(a)

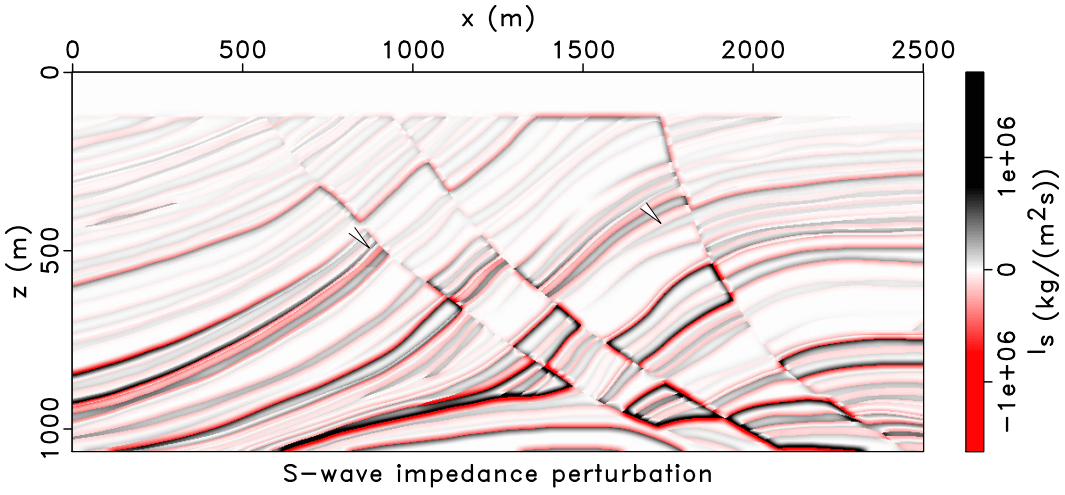


(b)

Figure 3.10: Elastic Marmousi2 model. (a) Smoothed P-wave velocity model. (b) Smoothed S-wave velocity model.

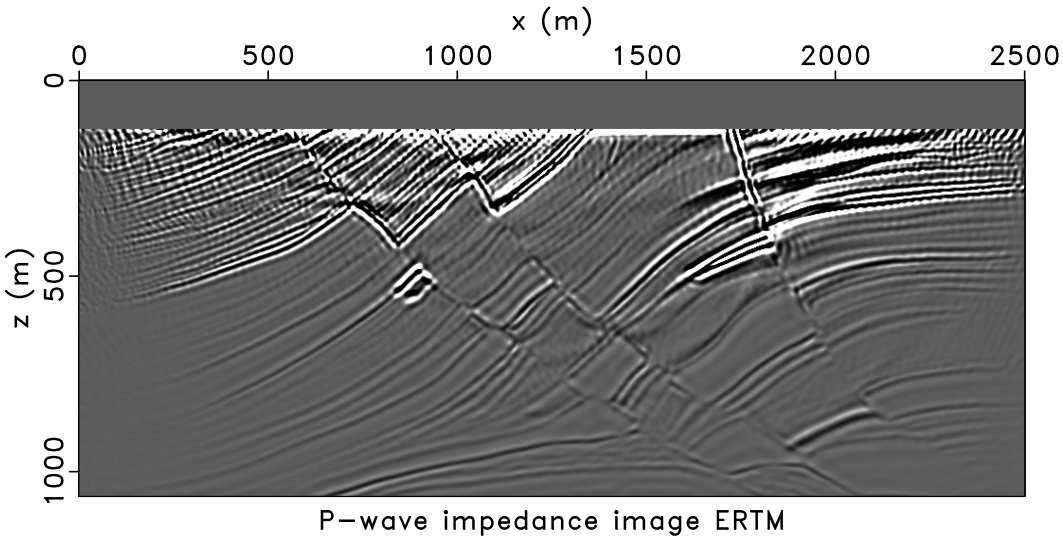


(a)

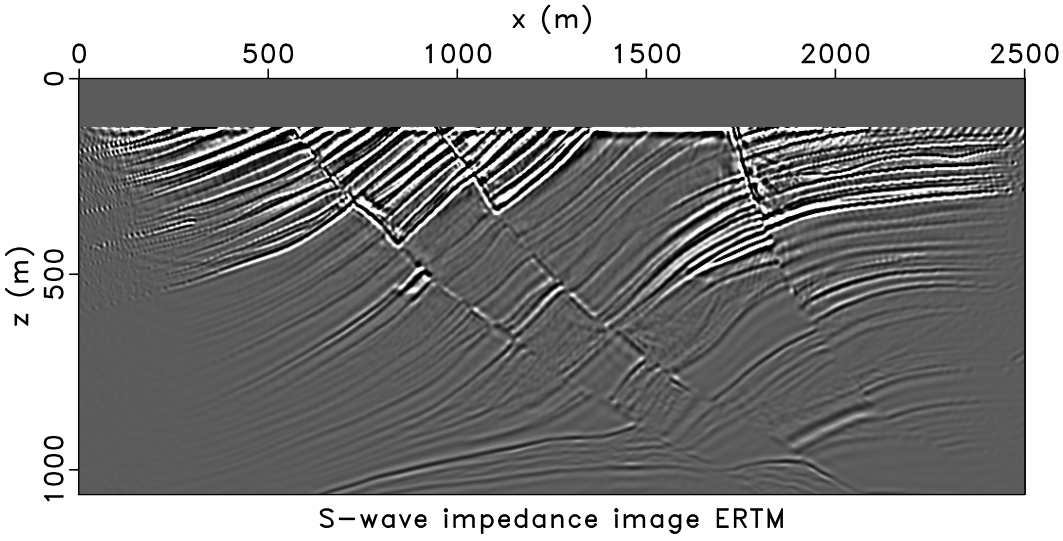


(b)

Figure 3.11: Elastic Marmousi2 model. (a) True P-wave impedance perturbation. (b) True S-wave impedance perturbation.

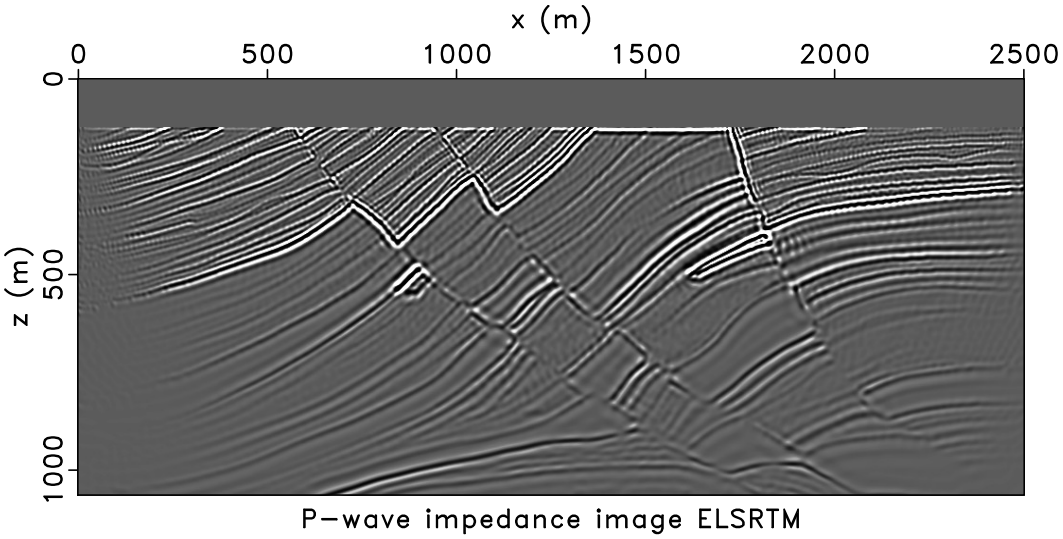


(a)

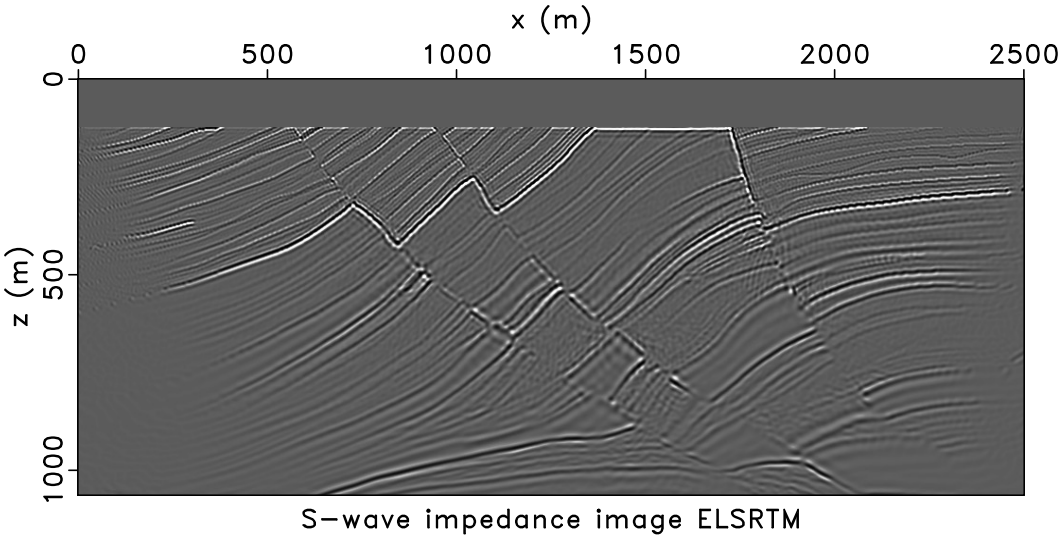


(b)

Figure 3.12: Images obtained via elastic RTM. (a) P-wave impedance perturbation image. (b) S-wave impedance perturbation image.

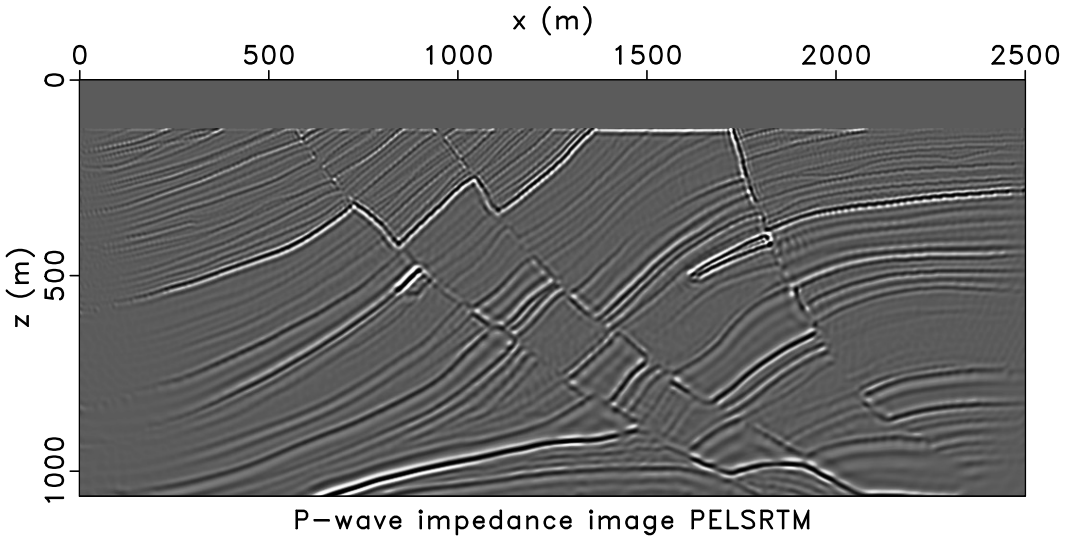


(a)

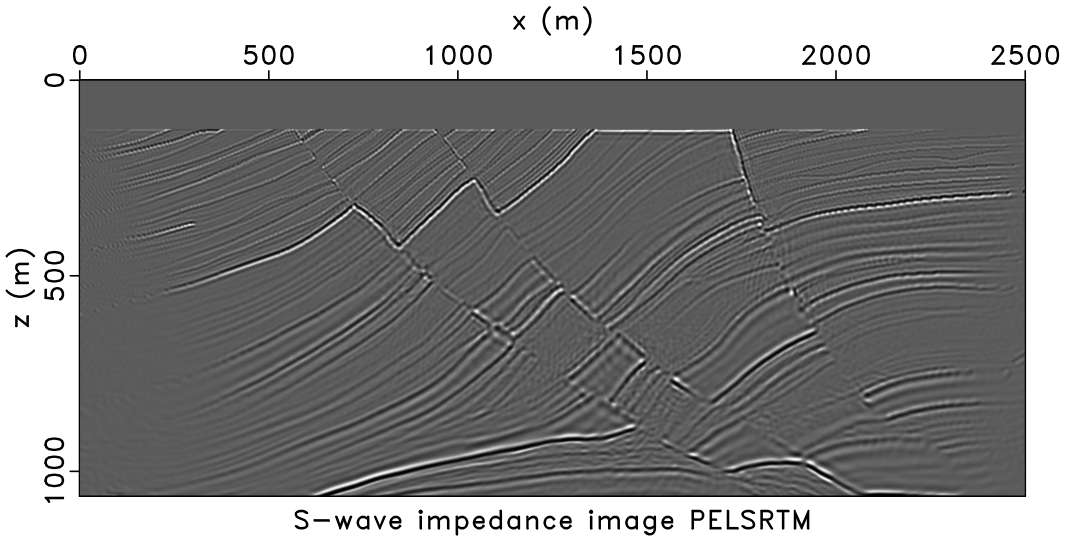


(b)

Figure 3.13: Images obtained via elastic LSRTM. (a) P-wave impedance perturbation image. (b) S-wave impedance perturbation image.



(a)



(b)

Figure 3.14: Images obtained via preconditioned elastic LSRTM. (a) P-wave impedance perturbation image. (b) S-wave impedance perturbation image.

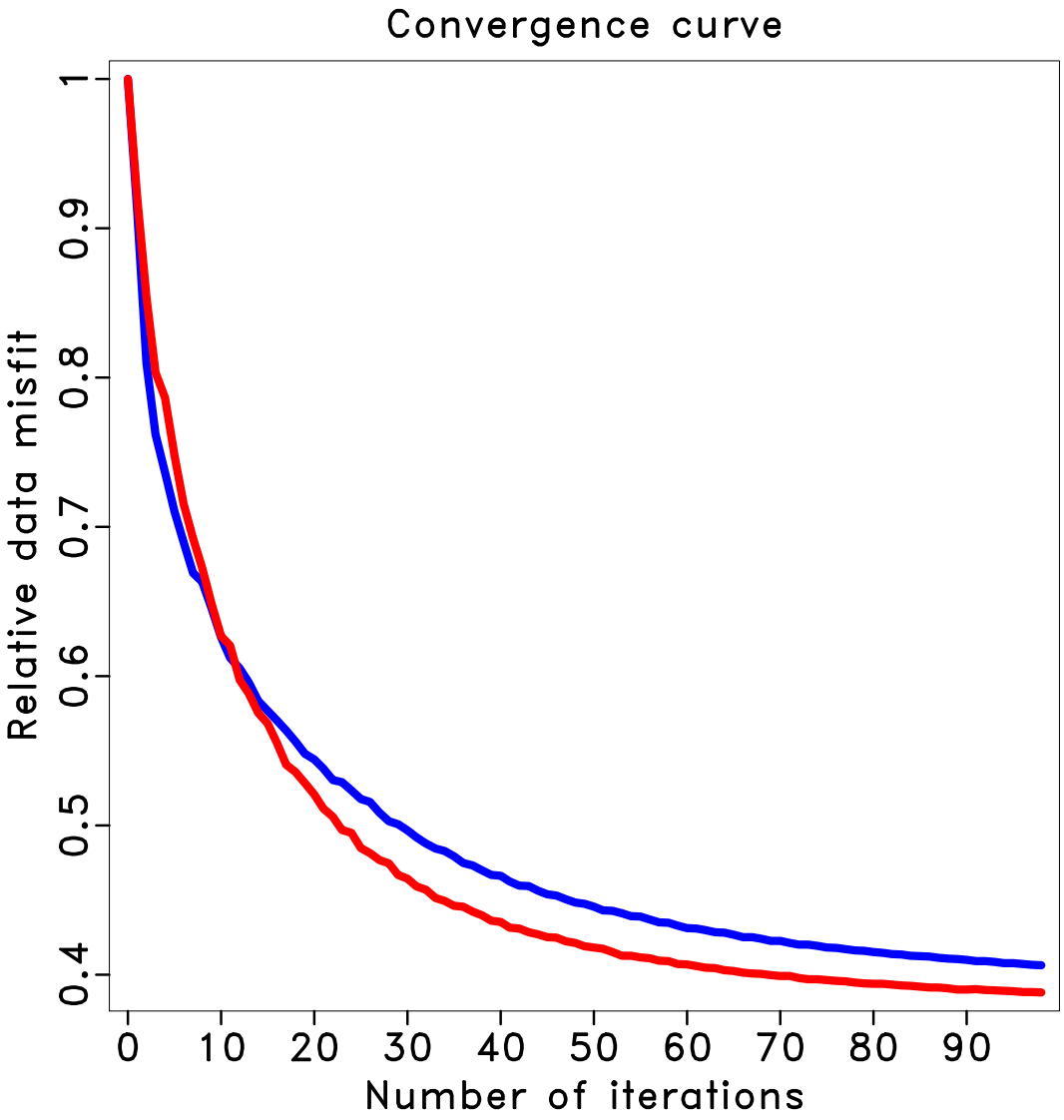


Figure 3.15: Comparison of relative data misfit convergence curves for elastic LSRTM (blue) and preconditioned elastic LSRTM (red) for the Marmousi2 model.

CHAPTER 4

The importance of including density in elastic least-squares reverse time migration: multiparameter crosstalk and convergence [1]

Time-domain elastic least-squares reverse time migration (LSRTM) can provide higher resolution images with fewer artifacts and a superior balance of amplitudes than elastic reverse time migration (RTM). More important, it can mitigate the crosstalk between P- and S-wave images. In previously proposed elastic LSRTM algorithms, density is either assumed to be constant or known. In other words, the density perturbation is not part of the least-squares inversion formulation. Neglecting density in elastic LSRTM may lead to crosstalk artifacts in the P- and S-wave images. In this article, we propose a time-domain three-parameter elastic LSRTM algorithm to simultaneously invert for density, P- and S-wave velocity perturbation images. We derive the elastic Born approximation and elastic RTM operators using the continuous adjoint-state method. We carefully discretize the two operators to assure that they pass the dot-product test. This allows us to use the conjugate gradient least-squares (CGLS) method to solve the least-squares migration problem. We evaluate the proposed algorithm on two synthetic examples. We show that our proposed three-parameter elastic LSRTM can reduce the multiparameter crosstalk among density, P- and S-wave velocity perturbation images and it can estimate the density perturbation image. Moreover, including density image in the elastic LSRTM inversion can improve the convergence of the least-squares inversion.

[1] A version of this chapter has been submitted to Geophysical Journal International.

4.1 Introduction

Seismic imaging techniques aim at portraying subsurface structure using seismic data. Reverse time migration (RTM) (McMechan, 1983; Baysal et al., 1983; Whitmore, 1983; Chang and McMechan, 1986; Etgen, 1986) is one of the most important seismic imaging methods. RTM utilizes the two-way acoustic wave equation for extrapolating wavefields into the interior of the earth. For this reason, RTM can handle steep and complex geological structures such as sedimentary areas with salt inclusions (Etgen et al., 2009). Due to the rapid development of computing power, RTM has become one of the most popular methods for industry applications of imaging techniques for resource exploration and exploitation. Acoustic RTM methods approximate the elastic solid earth by a fluid. However, an elastic formulation provides a better approximation to realistic media at exploration scales. Different elastic RTM methods have been developed by applying the excitation-time imaging condition (Sun and McMechan, 1986; Chang and McMechan, 1987; Sun and McMechan, 2001), the crosscorrelation imaging condition (Yan and Sava, 2008; Du et al., 2012), or utilizing elastic full-waveform inversion (FWI) gradients (Tarantola, 1986; Virieux and Operto, 2009; Luo et al., 2009; Zhu et al., 2009).

Migration is the adjoint of the forward Born modelling. The migrated images may suffer from relative low resolution, unbalanced amplitudes due to geometric spreading and acquisition footprint. Those issues can be mitigated by the least-squares migration (LSM). Different LSM techniques have been proposed in seismic imaging: least-squares Kirchhoff migration (Tarantola, 1984b; Lambare et al., 1992; Nemeth et al., 1999; Trad, 2017), least-squares one-way wave-equation migration (Kuehl and Sacchi, 2003; Rickett, 2003; Wang et al., 2005; Tang, 2009; Kaplan et al., 2010; Kazemi and Sacchi, 2015; Cheng et al., 2016) and least-squares reverse time migration (LSRTM) (Bourgeois et al., 1989; Ostmo et al., 2002; Dai et al., 2012; Dong et al., 2012; Zhang et al., 2015; Wong et al., 2015; Yao and Jakubowicz, 2016; Xue et al., 2016; Hou and Symes, 2016; Yang et al., 2016b; Chen et al., 2017). For inverting multicomponent data, elastic LSM methods have also been developed. For example, elastic least-squares Kirchhoff migration (Beydoun and Mendes, 1989; Jin et al., 1992), elastic least-squares one-way wave-equation migration (Stanton and Sacchi, 2015, 2017), and elastic LSRTM (Anikiev et al., 2013; Xu et al., 2016; Feng and Schuster, 2017; Duan et al., 2017; Chen and Sacchi, 2017a; Ren et al., 2017; Gu et al., 2017; Guo and McMechan, 2018).

Conventional elastic LSRTM algorithms do not include density image in the inversion. It inverts for P- and S-wave images. We call this type of elastic LSRTM a two-parameter elastic LSRTM. The density is either assumed to be constant or already known. However, this assumption is not valid in realistic earth media. Neglecting the density image in elastic

LSRTM may result in crosstalk artifacts in the P- and S-wave images. Chen and Sacchi (2017a) derive an elastic LSRTM algorithm including the density image component. However, they did not include density image inversion in their numerical examples. Sun et al. (2017) study a frequency-domain elastic LSRTM with density variation. Qu et al. (2018) present an elastic LSRTM with density variation based on P- and S-wave decoupled elastic velocity-stress wave equation. In this chpater, we propose a time-domain three-parameter elastic LSRTM algorithm to simultaneously invert for density perturbation, P- and S-wave velocity perturbation (Chen and Sacchi, 2018). The latter complements our previous work on elastic LSRTM (Chen and Sacchi, 2017a). Our three-parameter elastic LSRTM algorithm directly adopts the elastic wave equation without splitting the equation to P- and S-wave components as in Qu et al. (2018). We derive the elastic Born approximation and elastic RTM operators using the continuous adjoint-state method (Lions, 1971; Tarantola, 1988; Tromp et al., 2005; Plessix, 2006; Chen and Lee, 2015). We carefully discretize the two operators to ensure that they pass the dot-product test. The latter allows us to use the conjugate gradient least-squares (CGLS) method to solve the least-squares migration optimization problem. We show that the proposed three-parameter elastic LSRTM algorithm is able to mitigate the crosstalk among density, P- and S-wave velocity perturbations. Moreover, it improves the convergence and data fitting over the conventional two-parameter elastic LSRTM.

We have organized this article as follows. First, we describe the wave equation that we have adopted to simulate elastic wavefields. Then, we introduce the three-parameter elastic Born approximation and the elastic RTM operators. Subsequently, we present the proposed three-parameter elastic LSRTM algorithm. In the last section, we evaluate the performance of the proposed algorithm with numerical examples.

4.2 Theory

4.2.1 Elastic wave equation

The propagation of seismic wave in a heterogeneous, isotropic elastic earth media is described by the elastic wave equation (Virieux, 1986)

$$\begin{aligned}
&\rho\frac{\partial v_x}{\partial t} - \left(\frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xz}}{\partial z}\right) = 0,\\
&\rho\frac{\partial v_z}{\partial t} - \left(\frac{\partial \sigma_{xz}}{\partial x} + \frac{\partial \sigma_{zz}}{\partial z}\right) = 0,\\
&\frac{\partial \sigma_{xx}}{\partial t} - (\lambda+2\mu)\frac{\partial v_x}{\partial x} - \lambda\frac{\partial v_z}{\partial z} = f_{xx},\\
&\frac{\partial \sigma_{zz}}{\partial t} - (\lambda+2\mu)\frac{\partial v_z}{\partial z} - \lambda\frac{\partial v_x}{\partial x} = f_{zz},\\
&\frac{\partial \sigma_{xz}}{\partial t} - \mu\left(\frac{\partial v_x}{\partial z} + \frac{\partial v_z}{\partial x}\right) = 0,
\end{aligned} \tag{4.1}$$

where v_x and v_z are the horizontal and vertical particle velocity fields, σ_{xx}, σ_{xz} and σ_{zz} are the stress fields, ρ is the density, λ and μ are the Lamé parameters, and f_{xx} and f_{zz} are the explosive source terms. In the wave equation, we dropped the dependence on spatial and temporal coordinates $\mathbf{x}$ and t of our variables to make the notations concise but we understand that $v_x = v_x(\mathbf{x}, t)$, $\lambda = \lambda(\mathbf{x})$, etc. The elastic wave equation can be written in abstract functional form as follows

$$\mathbf{S}(\bar{\mathbf{m}})\mathbf{u} = \mathbf{f}, \tag{4.2}$$

where $\bar{\mathbf{m}} = (\rho, \lambda, \mu)^T$ denotes the model parameter vector, $\mathbf{S}$ is the wave equation operator, $\mathbf{u} = (v_x, v_z, \sigma_{xx}, \sigma_{zz}, \sigma_{xz})^T$ is the wavefield vector and $\mathbf{f} = (0, 0, f_{xx}, f_{zz}, 0)^T$ is the source vector. The seismic data are observed by receivers

$$\mathbf{d} = \mathbf{R}\mathbf{u}, \tag{4.3}$$

where $\mathbf{d} = (d_{v_x}, d_{v_z}, 0, 0, 0)^T$ is the seismic data vector and $\mathbf{R}$ is the sampling operator.

4.2.2 Three-parameter elastic Born approximation

Seismic imaging methods rely on the adjoint of the linearized forward problem. A perturbation around the known background model parameters

$$\rho \to \rho + \delta\rho, \tag{4.4a}$$
$$\lambda \to \lambda + \delta\lambda, \tag{4.4b}$$
$$\mu \to \mu + \delta\mu, \tag{4.4c}$$

leads to a perturbation of the wavefields

$$v_x \to v_x + \delta v_x, \tag{4.4d}$$
$$v_z \to v_z + \delta v_z, \tag{4.4e}$$
$$\sigma_{xx} \to \sigma_{xx} + \delta\sigma_{xx}, \tag{4.4f}$$
$$\sigma_{zz} \to \sigma_{zz} + \delta\sigma_{zz}, \tag{4.4g}$$
$$\sigma_{xz} \to \sigma_{xz} + \delta\sigma_{xz}. \tag{4.4h}$$

Inserting equation 4.4 into equation 4.1, subtracting equation 4.1, and dropping second and higher order terms leads to the Born approximation for the first-order velocity stress elastic wave equation system (Chen and Sacchi, 2017a)

$$\begin{aligned}
&\rho\frac{\partial\delta v_x}{\partial t} - \left(\frac{\partial\delta\sigma_{xx}}{\partial x} + \frac{\partial\delta\sigma_{xz}}{\partial z}\right) = -\delta\rho\dot{v}_x,\\
&\rho\frac{\partial\delta v_z}{\partial t} - \left(\frac{\partial\delta\sigma_{xz}}{\partial x} + \frac{\partial\delta\sigma_{zz}}{\partial z}\right) = -\delta\rho\dot{v}_z,\\
&\frac{\partial\delta\sigma_{xx}}{\partial t} - (\lambda+2\mu)\frac{\partial\delta v_x}{\partial x} - \lambda\frac{\partial\delta v_z}{\partial z} = (\delta\lambda+\delta\mu)\frac{\dot{\sigma}_{xx}+\dot{\sigma}_{zz}}{2(\lambda+\mu)} + \delta\mu\frac{\dot{\sigma}_{xx}-\dot{\sigma}_{zz}}{2\mu},\\
&\frac{\partial\delta\sigma_{zz}}{\partial t} - (\lambda+2\mu)\frac{\partial\delta v_z}{\partial z} - \lambda\frac{\partial\delta v_x}{\partial x} = (\delta\lambda+\delta\mu)\frac{\dot{\sigma}_{xx}+\dot{\sigma}_{zz}}{2(\lambda+\mu)} - \delta\mu\frac{\dot{\sigma}_{xx}-\dot{\sigma}_{zz}}{2\mu},\\
&\frac{\partial\delta\sigma_{xz}}{\partial t} - \mu\left(\frac{\partial\delta v_x}{\partial z} + \frac{\partial\delta v_z}{\partial x}\right) = \delta\mu\frac{\dot{\sigma}_{xz}}{\mu},
\end{aligned} \tag{4.5}$$

where $\mathbf{u} = (v_x, v_z, \sigma_{xx}, \sigma_{zz}, \sigma_{xz})^T$ is the incident wavefield or called source-side wavefield in the background model $\bar{\mathbf{m}} = (\rho, \lambda, \mu)^T$, $\delta\mathbf{u} = (\delta v_x, \delta v_z, \delta\sigma_{xx}, \delta\sigma_{zz}, \delta\sigma_{xz})^T$ is the scattered wavefield due to model perturbation $\delta\bar{\mathbf{m}} = (\delta\rho, \delta\lambda, \delta\mu)^T$, and over-dot means the time derivative. We parameterize our three-parameter elastic LSRTM in terms of density, P- and S-wave velocity perturbations. The different model parameter perturbations obey the

following expressions

$$\begin{pmatrix} \delta\rho \\ \delta\lambda \\ \delta\mu \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ V_p^2 - 2V_s^2 & 2\rho V_p & -4\rho V_s \\ V_s^2 & 0 & 2\rho V_s \end{pmatrix} \begin{pmatrix} \delta\varrho \\ \delta V_p \\ \delta V_s \end{pmatrix}, \tag{4.6}$$

where $V_p = \sqrt{(\lambda + 2\mu)/\rho}$ and $V_s = \sqrt{\mu/\rho}$ are background P- and S-wave velocities, δV_p and δV_s are P- and S-wave velocity perturbations.

We can also express the above expressions in abstract functional form. The model perturbation $\bar{\mathbf{m}} \rightarrow \bar{\mathbf{m}} + \delta\bar{\mathbf{m}}$ leads to a perturbation in wave equation 4.2

$$(\mathbf{S} + \delta\mathbf{S})(\mathbf{u} + \delta\mathbf{u}) = \mathbf{f}, \tag{4.7}$$

where $\delta\mathbf{S}$ is the wave equation operator perturbation, $\delta\mathbf{u}$ is the wavefield perturbation, $\mathbf{S}$ is the wave equation operator in background model and $\mathbf{u}$ is the wavefield in background model. Using wave equation $\mathbf{Su} = \mathbf{f}$ and neglecting second order term, equation 4.7 can be simplified as

$$\mathbf{S}\delta\mathbf{u} = -\delta\mathbf{S}\mathbf{u}. \tag{4.8}$$

The perturbed seismic data can be expressed as

$$\delta\mathbf{d} = \mathbf{R}\delta\mathbf{u} = -\mathbf{R}\mathbf{S}^{-1}\delta\mathbf{S}\mathbf{u} = -\mathbf{R}\mathbf{S}^{-1}\frac{\partial\mathbf{S}}{\partial\bar{\mathbf{m}}}\mathbf{u}\delta\bar{\mathbf{m}} \tag{4.9}$$

where $\mathbf{R}$ is the sampling operator, $\mathbf{S}^{-1}$ is the inverse of wave equation operator and the linear operator $\partial\mathbf{S}/\partial\bar{\mathbf{m}}$ denotes the diffraction pattern of density and Lamé parameters (Pageot et al., 2013; Operto et al., 2013). The vector $\delta\bar{\mathbf{m}} = (\delta\rho, \delta\lambda, \delta\mu)^T$ denotes density and Lamé parameters perturbations. The parameter perturbations are connected via

$$\delta\bar{\mathbf{m}} = \mathbf{T}\delta\mathbf{m}, \tag{4.10}$$

where $\mathbf{T}$ denotes the transformation matrix in equation 4.6, and $\delta\mathbf{m} = (\delta\varrho, \delta V_p, \delta V_s)^T$. The elastic Born approximation can be expressed in abstract form as

$$\delta\mathbf{d} = \mathbf{L}\delta\mathbf{m} = -\mathbf{R}\mathbf{S}^{-1}\frac{\partial\mathbf{S}}{\partial\bar{\mathbf{m}}}\mathbf{u}\mathbf{T}\delta\mathbf{m}, \tag{4.11}$$

where operator $\mathbf{L}$ indicates the elastic Born approximation operator.

4.2.3 Three-parameter elastic reverse time migration

Seismic migration estimates subsurface structural image using the seismic data recorded on the surface of the earth. Migration can be regarded as the adjoint of the Born approximation operator

$$\delta\mathbf{m}^* = \mathbf{L}^\dagger\delta\mathbf{d} = -\mathbf{T}^\dagger\left(\frac{\partial\mathbf{S}}{\partial\bar{\mathbf{m}}}\mathbf{u}\right)^\dagger\left(\mathbf{S}^{-1}\right)^\dagger\mathbf{R}^\dagger\delta\mathbf{d} = -\mathbf{T}^\dagger\left(\frac{\partial\mathbf{S}}{\partial\bar{\mathbf{m}}}\mathbf{u}\right)^\dagger\left(\mathbf{S}^\dagger\right)^{-1}\mathbf{R}^\dagger\delta\mathbf{d} \tag{4.12}$$

where † indicates the adjoint of an operator, $\mathbf{L}^\dagger$ is the elastic RTM operator and $\delta\mathbf{m}^*$ is the migrated elastic images ($\delta\mathbf{m}^* = (\delta\varrho^*, \delta V_p^*, \delta V_s^*)^T$). We introduce the adjoint-state variable $\mathbf{p} = \left(\mathbf{S}^\dagger\right)^{-1}\mathbf{R}^\dagger\delta\mathbf{d}$. The latter satisfies the adjoint-state equation corresponding to the state equation 4.2

$$\mathbf{S}^\dagger\mathbf{p} = \mathbf{R}^\dagger\delta\mathbf{d}, \tag{4.13}$$

where $\mathbf{S}^\dagger$ is the adjoint wave equation operator and $\mathbf{R}^\dagger\delta\mathbf{d}$ is the adjoint source. The migrated elastic images can be simplified as

$$\delta\mathbf{m}^* = -\mathbf{T}^\dagger\left(\frac{\partial\mathbf{S}}{\partial\bar{\mathbf{m}}}\mathbf{u}\right)^\dagger\mathbf{p} = -\mathbf{T}^\dagger\delta\bar{\mathbf{m}}^*, \tag{4.14}$$

where $\mathbf{T}^\dagger$ is the adjoint of the transformation matrix in equation 4.6 and $\delta\bar{\mathbf{m}}^* = (\delta\rho^*, \delta\lambda^*, \delta\mu^*)^T$. The adjoint-state equation corresponding to the first-order velocity-stress elastic wave equation 4.1 can be derived

$$\begin{aligned}
&-\rho\frac{\partial v_x}{\partial t} + \left(\frac{\partial\varsigma_{xx}}{\partial x} + \frac{\partial\varsigma_{xz}}{\partial z}\right) = \delta d_{v_x},\\
&-\rho\frac{\partial v_z}{\partial t} + \left(\frac{\partial\varsigma_{xz}}{\partial x} + \frac{\partial\varsigma_{zz}}{\partial z}\right) = \delta d_{v_z},\\
&-\frac{\partial\varsigma_{xx}}{\partial t} + (\lambda+2\mu)\frac{\partial v_x}{\partial x} + \lambda\frac{\partial v_z}{\partial z} = 0,\\
&-\frac{\partial\varsigma_{zz}}{\partial t} + (\lambda+2\mu)\frac{\partial v_z}{\partial z} + \lambda\frac{\partial v_x}{\partial x} = 0,\\
&-\frac{\partial\varsigma_{xz}}{\partial t} + \mu\left(\frac{\partial v_x}{\partial z} + \frac{\partial v_z}{\partial x}\right) = 0,
\end{aligned} \tag{4.15}$$

where $\mathbf{p} = (\upsilon_x, \upsilon_z, \varsigma_{xx}, \varsigma_{zz}, \varsigma_{xz})^T$ is the adjoint-state wavefield and $\delta\mathbf{d} = (\delta d_{v_x}, \delta d_{v_z}, 0, 0, 0)^T$ is the data residual. The migrated density and Lamé parameter images can be written as

$$\begin{aligned}
\delta\rho^* &= -\int (\dot{v}_x \upsilon_x + \dot{v}_z \upsilon_z) dt, \\
\delta\lambda^* &= \int \frac{(\dot{\sigma}_{xx} + \dot{\sigma}_{zz})(\varsigma_{xx} + \varsigma_{zz})}{4(\lambda+\mu)^2} dt, \\
\delta\mu^* &= \int \left[\frac{\dot{\sigma}_{xz}\varsigma_{xz}}{\mu^2} + \frac{(\dot{\sigma}_{xx} + \dot{\sigma}_{zz})(\varsigma_{xx} + \varsigma_{zz})}{4(\lambda+\mu)^2} + \frac{(\dot{\sigma}_{xx} - \dot{\sigma}_{zz})(\varsigma_{xx} - \varsigma_{zz})}{4\mu^2} \right] dt.
\end{aligned} \tag{4.16}$$

The density and Lamé parameter perturbations can be transformed to density and wave velocity perturbations

$$\begin{pmatrix} \delta\varrho^* \\ \delta V_p^* \\ \delta V_s^* \end{pmatrix} = \begin{pmatrix} 1 & V_p^2 - 2V_s^2 & V_s^2 \\ 0 & 2\rho V_p & 0 \\ 0 & -4\rho V_s & 2\rho V_s \end{pmatrix} \begin{pmatrix} \delta\rho^* \\ \delta\lambda^* \\ \delta\mu^* \end{pmatrix}. \tag{4.17}$$

4.2.4 Three-parameter elastic least-squares reverse time migration

To improve the resolution of images and reduce the cross-talk and artifacts in the images, we formulate the three-parameter elastic LSRTM as a least-squares inversion problem

$$J = \frac{1}{2} \sum_{i=1}^{N_s} \|\mathbf{L}_i \delta\mathbf{m} - \delta\mathbf{d}_i\|_2^2, \tag{4.18}$$

where $\mathbf{L}_i$ is the elastic Born approximation operator for the ith shot, $\delta\mathbf{d}_i$ is the ith shot gather, $\delta\mathbf{m} = (\delta\varrho, \delta V_p, \delta V_s)^T$ denotes elastic images, N_s indicates the number of shots and $\|\cdot\|_2$ indicates the ℓ_2 norm of vector. To accelerate the convergence of the inversion, we adopt the elastic pseudo-Hessian for preconditioning the system of equations (Shin et al., 2001a)

$$J = \frac{1}{2} \sum_{i=1}^{N_s} \|\mathbf{L}_i \mathbf{P} \delta\tilde{\mathbf{m}} - \delta\mathbf{d}_i\|_2^2, \tag{4.19}$$

where $\mathbf{P}$ denotes the inverse of the diagonal of pseudo-Hessian, $\delta\mathbf{m} = \mathbf{P}\delta\tilde{\mathbf{m}} = (\delta\varrho, \delta V_p, \delta V_s)^T$. It is important to mention that we carefully discretize the elastic Born and RTM operators ($\mathbf{L}$ and $\mathbf{L}^\dagger$) to assure they pass the dot-product test (Claerbout, 1992; Chen and Sacchi, 2017a). We adopt the preconditioned conjugate gradient least squares (PCGLS) algorithm (Bjorck, 1996) to solve equation 4.19. In this study, the damping regularization term was not included to the objective function 4.19. Instead, we adopt an iterative regularization method in which the iteration number plays the role of the regularization parameter (Hansen, 1998). The latter avoids the difficulty of picking regularization parameter.

4.3 Numerical examples

Our code was written in C language and parallelized with Message Passing Interface (MPI) over shots. Our forward modelling engine adopts a time domain staggered-grid finite-difference (FD) scheme (Virieux, 1986) to discretize the elastic wave equation and the unsplit Convolutional Perfectly Matched Layer (C-PML) boundary (Komatitsch and Martin, 2007) to absorb the artificial reflections from computational boundaries. In our code, the spatial FD order is selectable. The code automatically computes the FD coefficients from the user-specified FD order (Liu and Sen, 2009). Our elastic Born and RTM codes pass the dot-product test. The inversion is conducted with the PCGLS algorithm as mentioned above. We present two numerical examples to test the proposed algorithm. We compare the results of the proposed three-parameter elastic LSRTM and the results of two-parameter elastic LSRTM (Chen and Sacchi, 2017a). We emphasize that our synthetic multicomponent data are the solution of the time-domain elastic wave equation 4.1. We do not use the linearized Born modelling to generate data. The data contain full-wave modes except for direct waves.

4.3.1 Elastic inclusion model

Figure 4.1 a, c and e show the true P- and S-wave velocity and density models. This model may not be realistic. However, it is useful to demonstrate the multiparameter crosstalk in the elastic imaging. The model has 501×301 grid points with grid interval of 5 m. There are 101 shots located along the surface of the model with an interval of 25 m. There are 501 receivers located along the surface of the model with an interval of 5 m. We use a Ricker wavelet with central frequency 20 Hz as the source wavelet. The observed data (Figure 4.2) are simulated using our time domain elastic FD code. Figure 4.1 b, d and f are the migration P- and S-wave velocity and density models. The migration models are obtained by convolving the true models with a 2D Gaussian filter of 50 m width. Figure 4.3 a, c, and e show the true P- and S-wave velocity and density perturbations.

The elastic RTM generates images with strong multiparameter crosstalk (Figure 4.3 b, d, f). Results, in this case, cannot be interpreted properly. Moreover, there are high-amplitude low-frequency RTM artifacts in the images even after applying a Laplacian filtering. We adopted a number of 30 iterations for the two-parameter elastic LSRTM. The relative data misfit reduces to 10%. Our results show that the method significantly attenuated the crosstalk between the P- and S-wave velocity perturbation images (Figure 4.4 b and d). However, the density perturbation manifests as crosstalk in the estimated P- and S-wave velocity perturbation images. It is clear that the results will impede the proper interpretation of the P and S-wave images. The proposed three-parameter elastic LSRTM can suppress

the crosstalk and properly estimate the three images (Figure 4.4 a, c, e). Moreover, typical RTM artifacts have decreased, and the resolution of the images have significantly improved. The number of iterations of the three-parameter elastic LSRTM is 20, and the relative data misfit reduces to 6%. We also compare the convergence curves of the three-parameter elastic LSRTM and the two-parameter elastic LSRTM (Figure 4.5). Including density perturbation in the inversion leads to an improvement in the convergence of the iterative inversion.

4.3.2 Elastic Marmousi2 model

In this section, we evaluate the proposed algorithm on a more complex model: the elastic Marmousi2 model (Martin et al., 2006). We reduce the size of the original model to 1001×426 grids with grid interval 2.5 m. We also replace the water layer by a low-velocity layer in the original model. Figure 4.6 a, c and e show the true P- and S-wave velocity and density models. There are uncorrelated structures in the three models (indicated by the white triangles) representing potential hydrocarbon reservoirs. There are 101 shots located along the surface of the model with interval of 25 m. There are 1001 receivers located along the surface of the model with interval of 2.5 m. The source wavelet is a Ricker wavelet with central frequency 35 Hz. The observed data are simulated using our time domain elastic FD code (Figure 4.7). Figure 4.6 b, d and f are the smoothed background P- and S-wave velocity and density models. Background model smoothing is obtained by convolving the true models with a 2D Gaussian filter of 35 m width. Figure 4.8 a, c, and e show the true P- and S-wave velocity and density perturbations.

The images generated by elastic RTM contain strong low-frequency RTM artifacts (Figure 4.8 b, d, f). Moreover, the amplitudes for the shallow and deep parts of the images are unbalanced. A particular issue is that the elastic RTM generates crosstalk among the three elastic images in areas where the models are uncorrelated. The two-parameter elastic LSRTM (Chen and Sacchi, 2017a) can largely resolve those problems (Figure 4.9 b and d). However, it does not provide an estimation of the density perturbation image. These results were obtained after 100 iterations of the two-parameter elastic LSRTM. And the relative data misfit reduces to 40%. The proposed three-parameter elastic LSRTM is able to properly estimate the P- and S-wave velocity and density perturbation images (Figure 4.9 a, c and e). The images have more balanced amplitudes, fewer low-frequency RTM artifacts and reduced multiparameter crosstalk. These results were obtained after 100 iterations of the three-parameter elastic LSRTM. And the relative data misfit reduces to 23%. To appreciate details more clearly, we compare the elastic images in the three windows around the three white triangles in the model (Figure 4.6 a, c and e) as Figure 4.10, Figure 4.11 and Figure 4.12. The three-parameter elastic LSRTM generated images with the highest resolution and fewest artifacts and crosstalk. We compare the convergence curves of the three-parameter

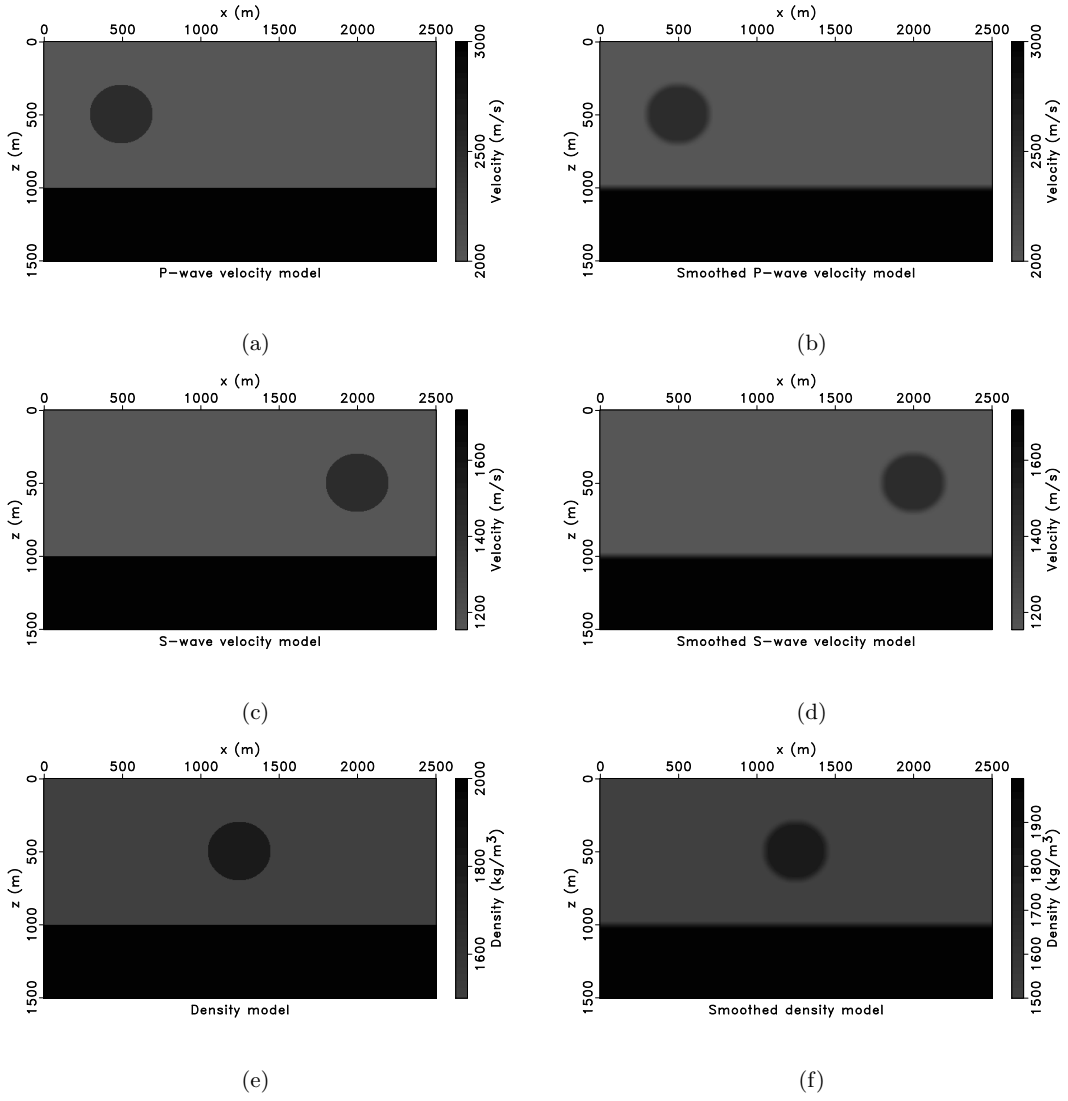
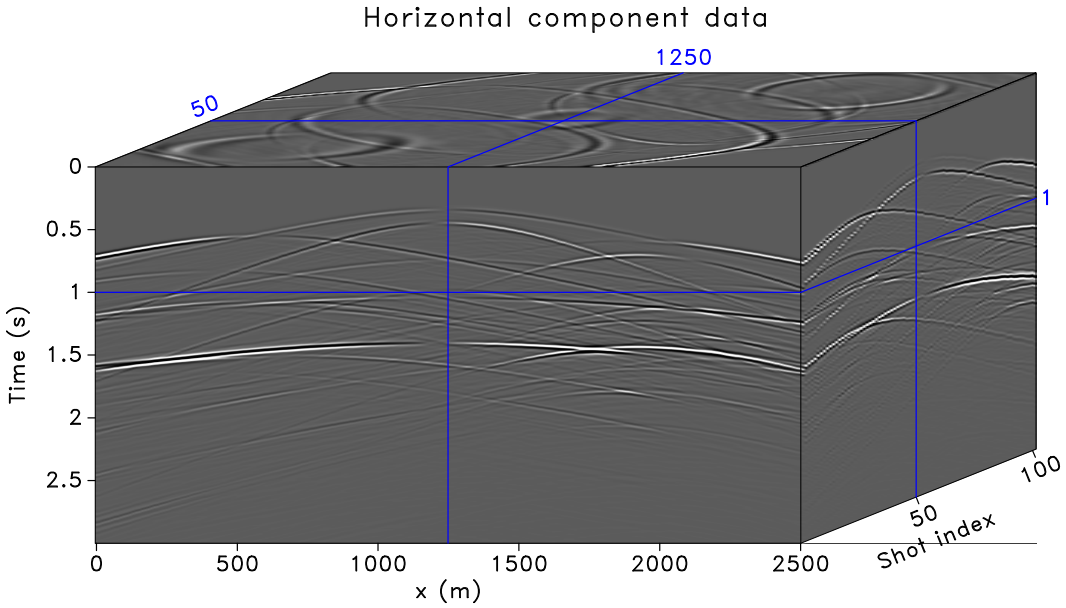
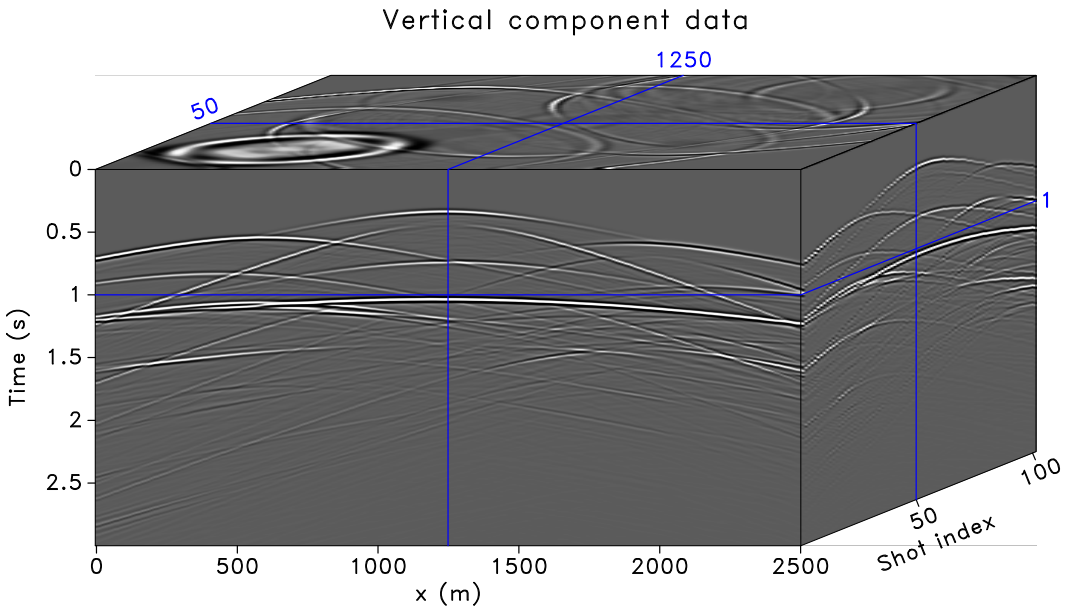


Figure 4.1: Elastic inclusion model. (a) P-wave velocity model. (b) Smoothed P-wave velocity model. (c) S-wave velocity model. (d) Smoothed S-wave velocity model. (e) Density model. (f) Smoothed density model.



(a)



(b)

Figure 4.2: Multicomponent data of the elastic inclusion model. (a) Horizontal particle velocity data. (b) Vertical particle velocity data.

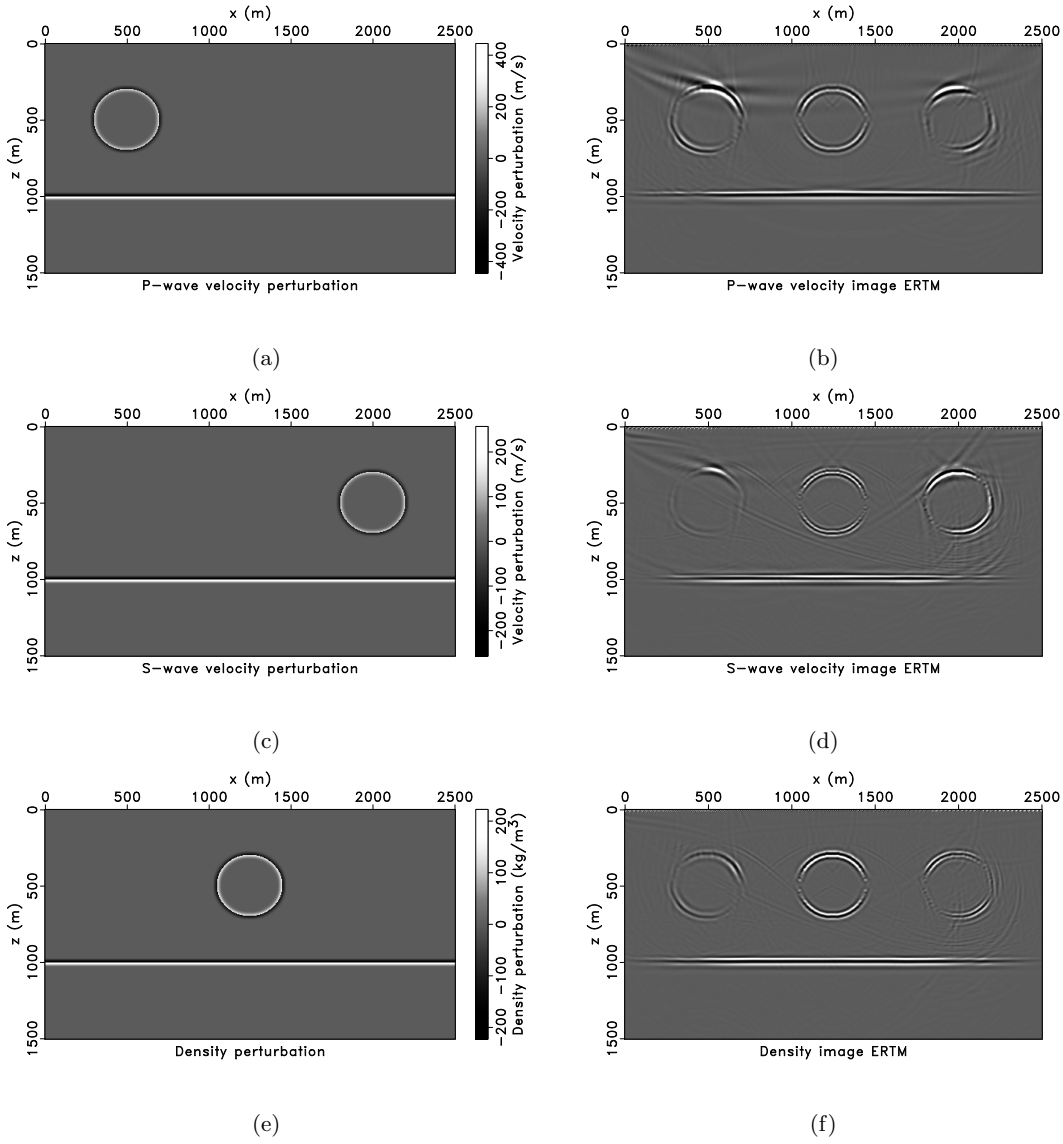


Figure 4.3: (a) True P-wave velocity perturbation. (b) P-wave velocity perturbation image estimated via elastic RTM. (c) True S-wave velocity perturbation. (d) S-wave velocity perturbation image estimated via elastic RTM. (e) True density perturbation. (f) Density perturbation image estimated via elastic RTM.

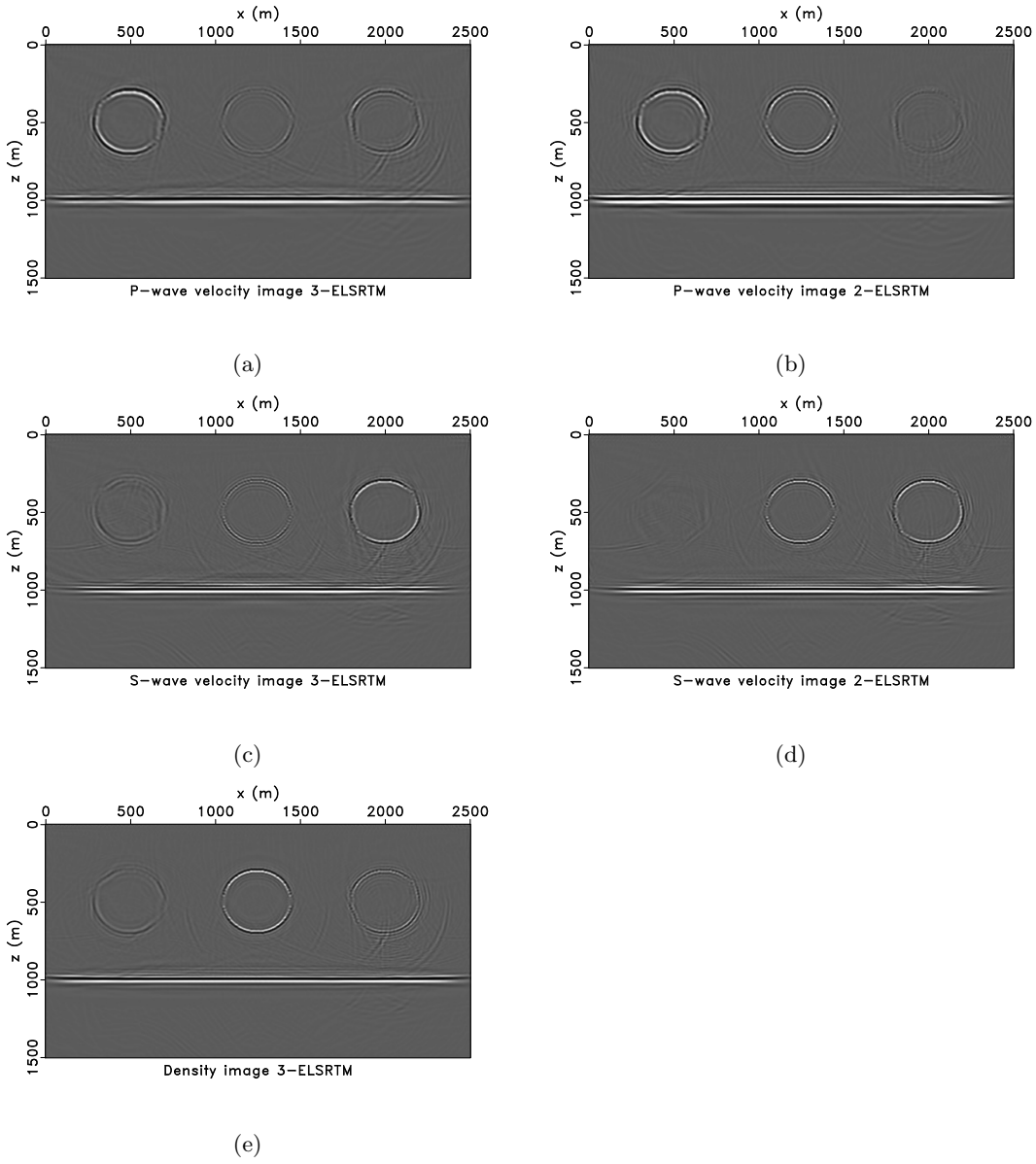


Figure 4.4: P-wave velocity perturbation image estimated via three-parameter elastic LSRTM (a) and two-parameter elastic LSRTM (b). S-wave velocity perturbation image estimated via three-parameter elastic LSRTM (c) and two-parameter elastic LSRTM (d). Density perturbation image estimated via three-parameter elastic LSRTM (e).

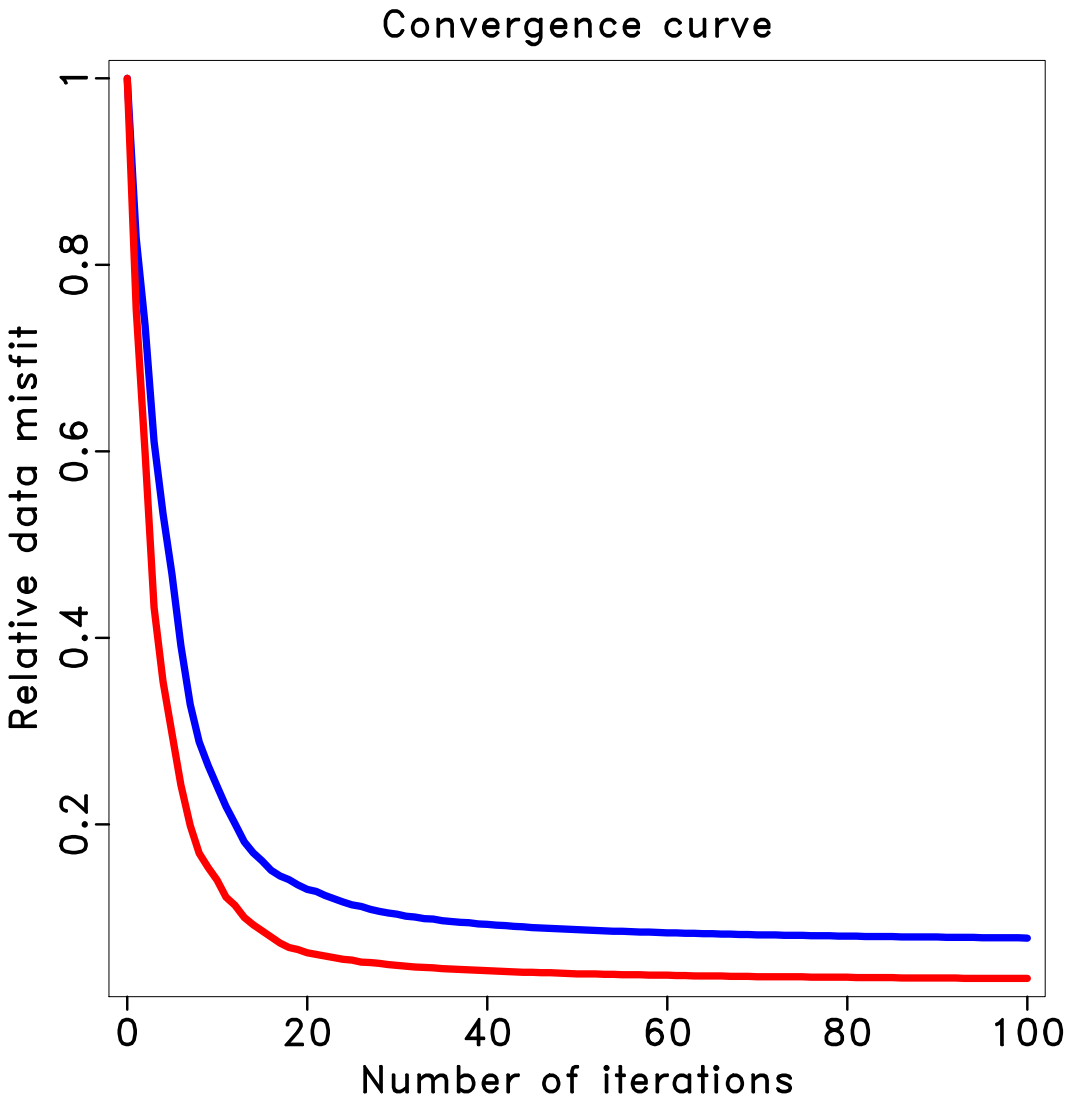


Figure 4.5: Comparison of the convergence curves of three-parameter elastic LSRTM (red) and two-parameter elastic LSRTM (blue) for elastic inclusion model.

elastic LSRTM and two-parameter elastic LSRTM in Figure 4.13. The three-parameter elastic LSRTM converges faster than the two-parameter elastic LSRTM.

4.4 Discussion

The newly developed three-parameter elastic LSRTM can mitigate the multiparameter crosstalk among density, P- and S-wave images. The method does not completely remove the crosstalk. From the radiation pattern analysis (Figure 1.3), the P-P wave responses caused by density perturbation and P-wave velocity perturbation are closely coupled at small scattering angle. The P-S converted wave response due to density and S-wave velocity perturbation are coupled at moderate scattering angle. It indicates that even the short wavelength of the density model may be resolved only partially. It may not be possible to completely decouple the three isotropic elastic images. It is also important to point out that our synthetic data are obtained via the finite-difference solution of the elastic wave equation. The data contain all modes including multiples. In other words, the data were not generated using the linearized Born modeling operator. This also adds complications to linearized least-squares waveform inversion because multiples are not honoured by the linearized Born forward modeling operator. A recommendation is to apply free-surface and internal multiple attenuation methods prior to least-squares migration (Verschuur et al., 1992; Weglein, 1999). However, we point out that the proper elimination of internal multiples is an important challenge (Berkhout and Verschuur, 2005).

4.5 Conclusions

The conventional two-parameter elastic LSRTM algorithm does not consider density image in the inversion. Neglecting density image in the inversion may generate crosstalk artifacts in P- and S-wave images. We propose a time-domain three-parameter elastic LSRTM method. It simultaneously inverts for density, P- and S-wave velocity perturbation images. We derive the elastic Born approximation and elastic RTM operators using the time-domain continuous adjoint-state method. We carefully discretize the two operators to assure that they pass the dot-product test. The latter allows us to use the CGLS method to solve the least-squares migration quadratic optimization problem on the fly. We observe that the proposed three-parameter elastic LSRTM can decouple the three isotropic elastic parameters and suppress the crosstalk. Moreover, it provides faster convergence and an improved data fitting than the two-parameter elastic LSRTM.

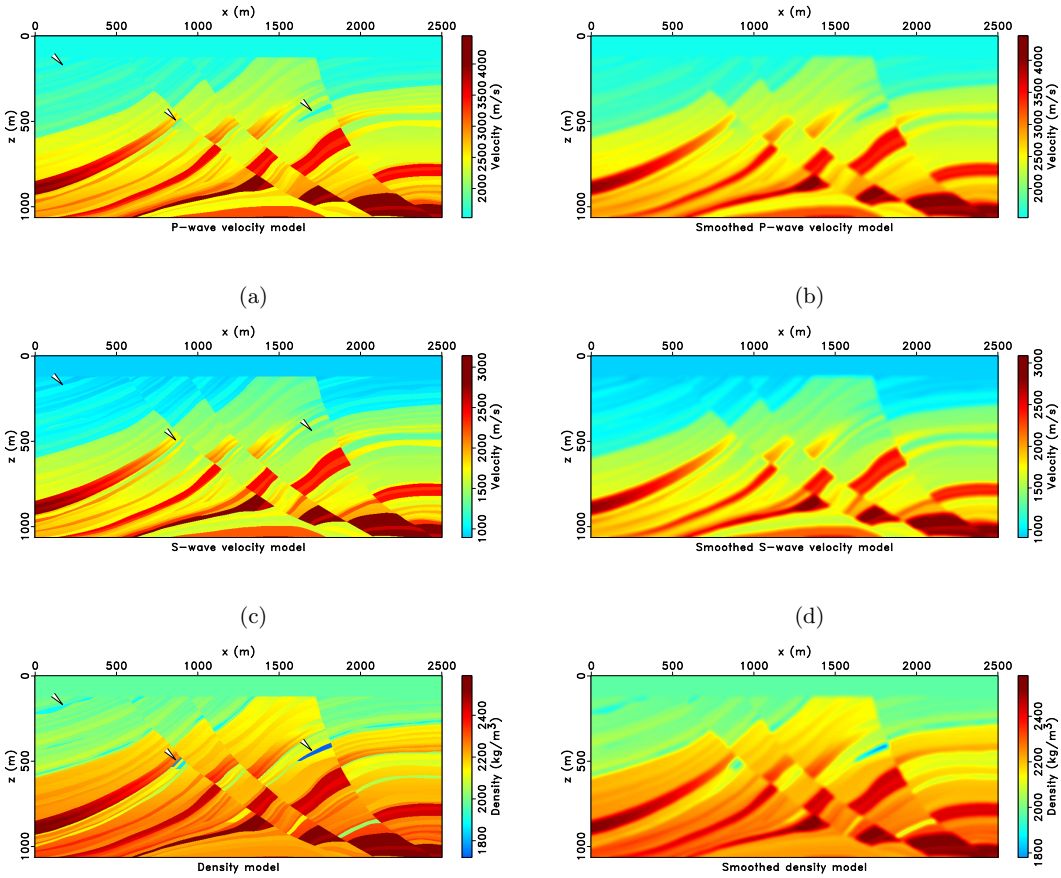
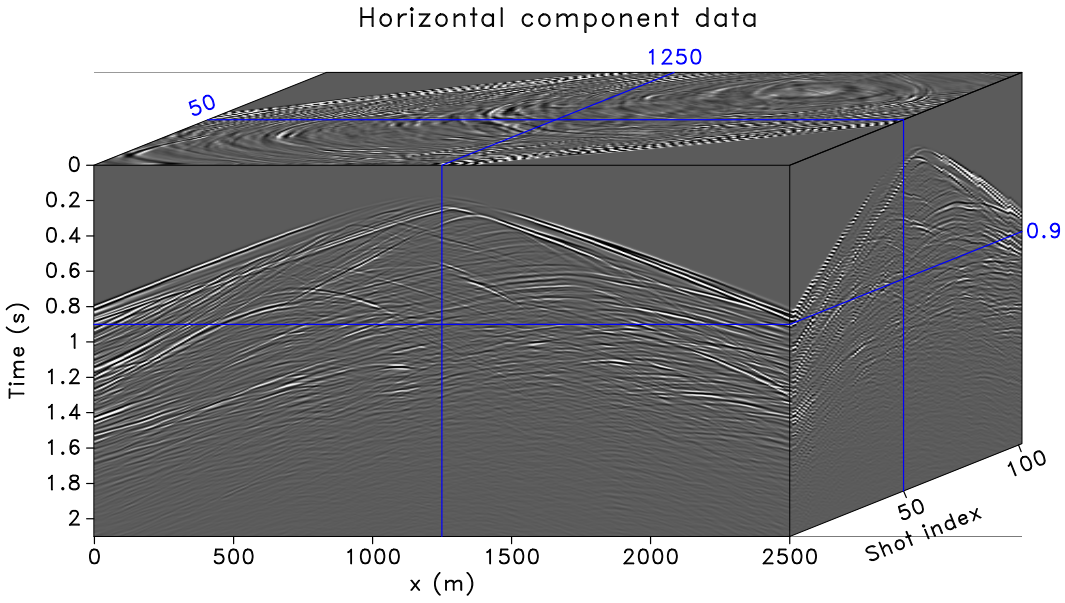
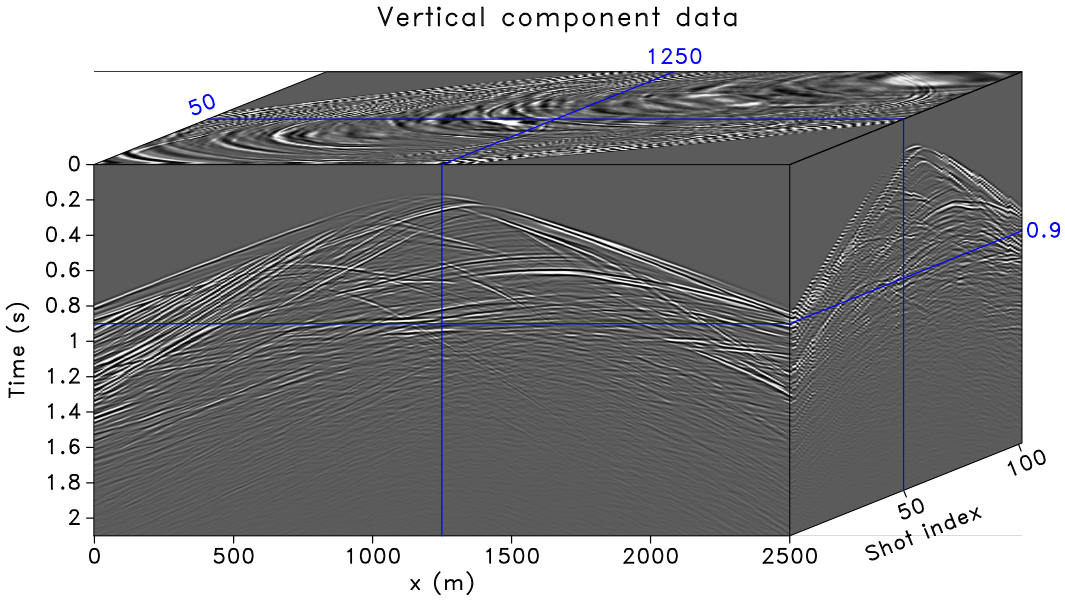


Figure 4.6: Elastic Marmousi2 model. (a) P-wave velocity model. (b) Smoothed P-wave velocity model. (c) S-wave velocity model. (d) Smoothed S-wave velocity model. (e) Density model. (f) Smoothed density model.



(a)



(b)

Figure 4.7: Multicomponent data of the elastic Marmousi2 model. (a) Horizontal particle velocity data. (b) Vertical particle velocity data.

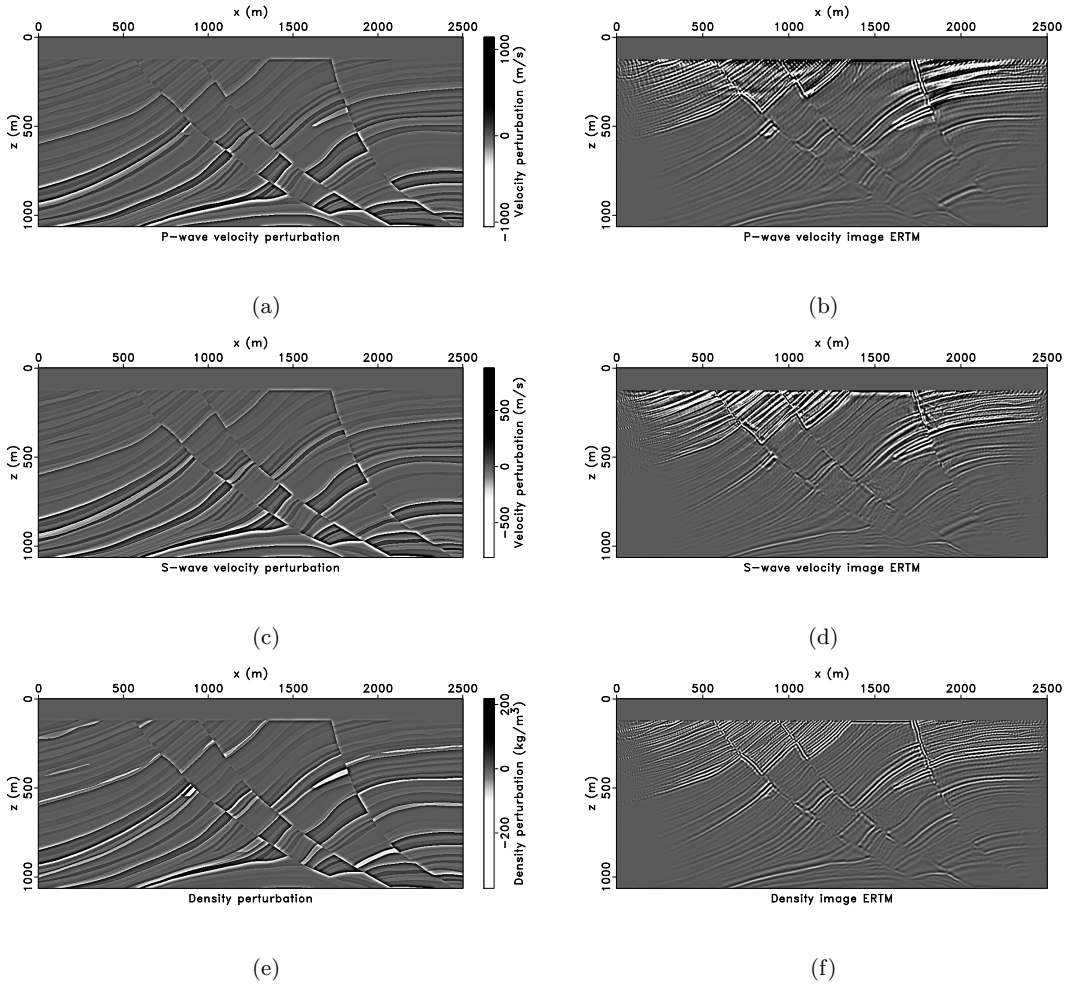


Figure 4.8: (a) True P-wave velocity perturbation. (b) P-wave velocity perturbation image estimated via elastic RTM. (c) True S-wave velocity perturbation. (d) S-wave velocity perturbation image estimated via elastic RTM. (e) True density perturbation. (f) Density perturbation image estimated via elastic RTM.

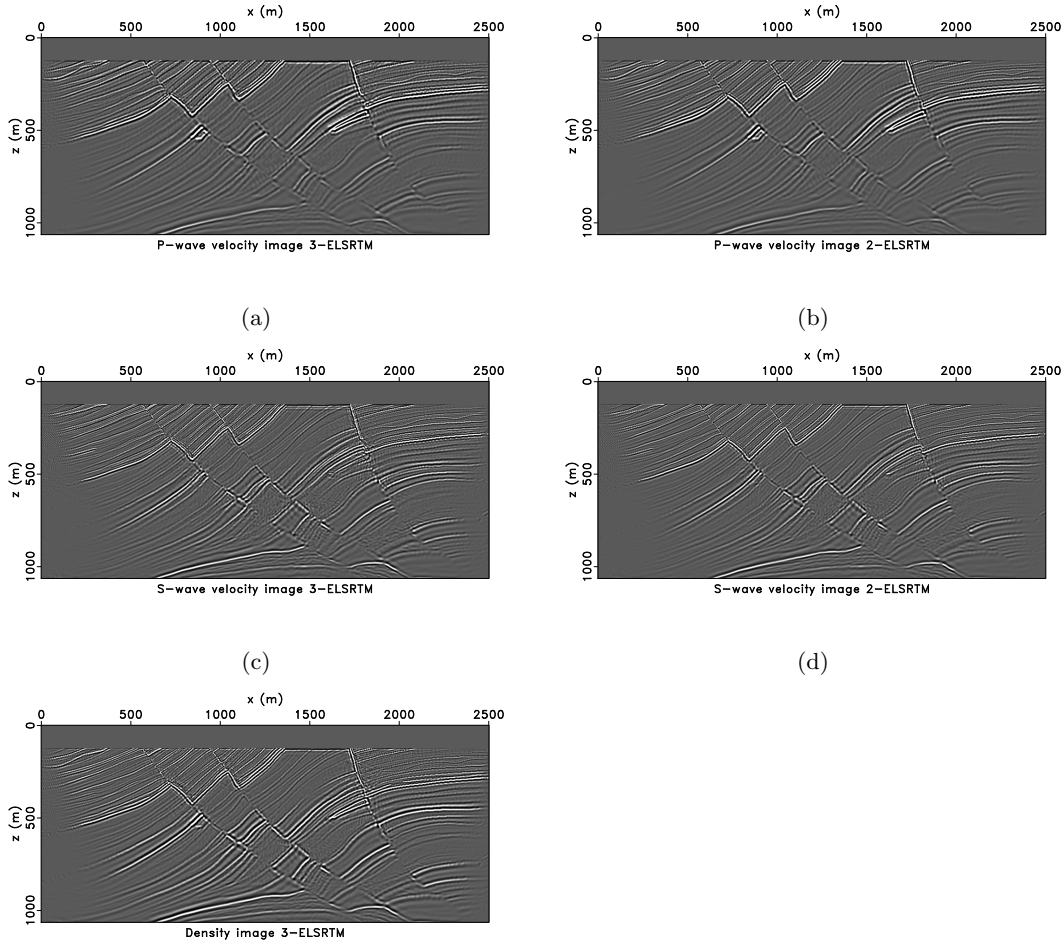


Figure 4.9: P-wave velocity perturbation image estimated via three-parameter elastic LSRTM (a) and two-parameter elastic LSRTM (b). S-wave velocity perturbation image estimated via three-parameter elastic LSRTM (c) and two-parameter elastic LSRTM (d). Density perturbation image estimated via three-parameter elastic LSRTM (e).

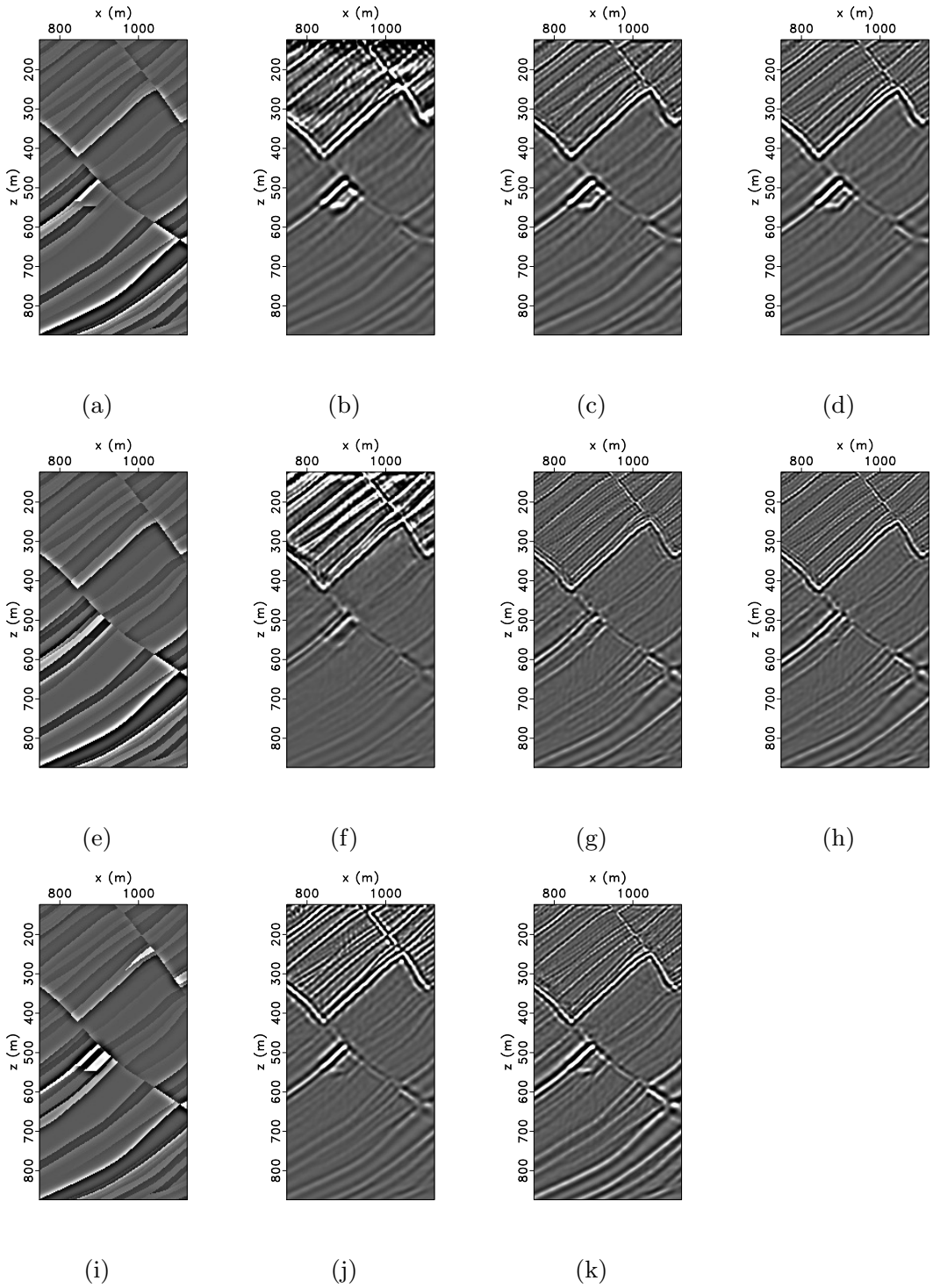


Figure 4.10: Comparison of images in the range 750 m $< x <$ 1125 m and 125 m $< z <$ 875 m. P-wave velocity perturbation: (a) true model, (b) elastic RTM, (c) three-parameter elastic LSRTM, (d) two-parameter elastic LSRTM. S-wave velocity perturbation: (e) true model, (f) elastic RTM, (g) three-parameter elastic LSRTM, (h) two-parameter elastic LSRTM. Density perturbation: (i) true model, (j) elastic RTM, (k) three-parameter elastic LSRTM.

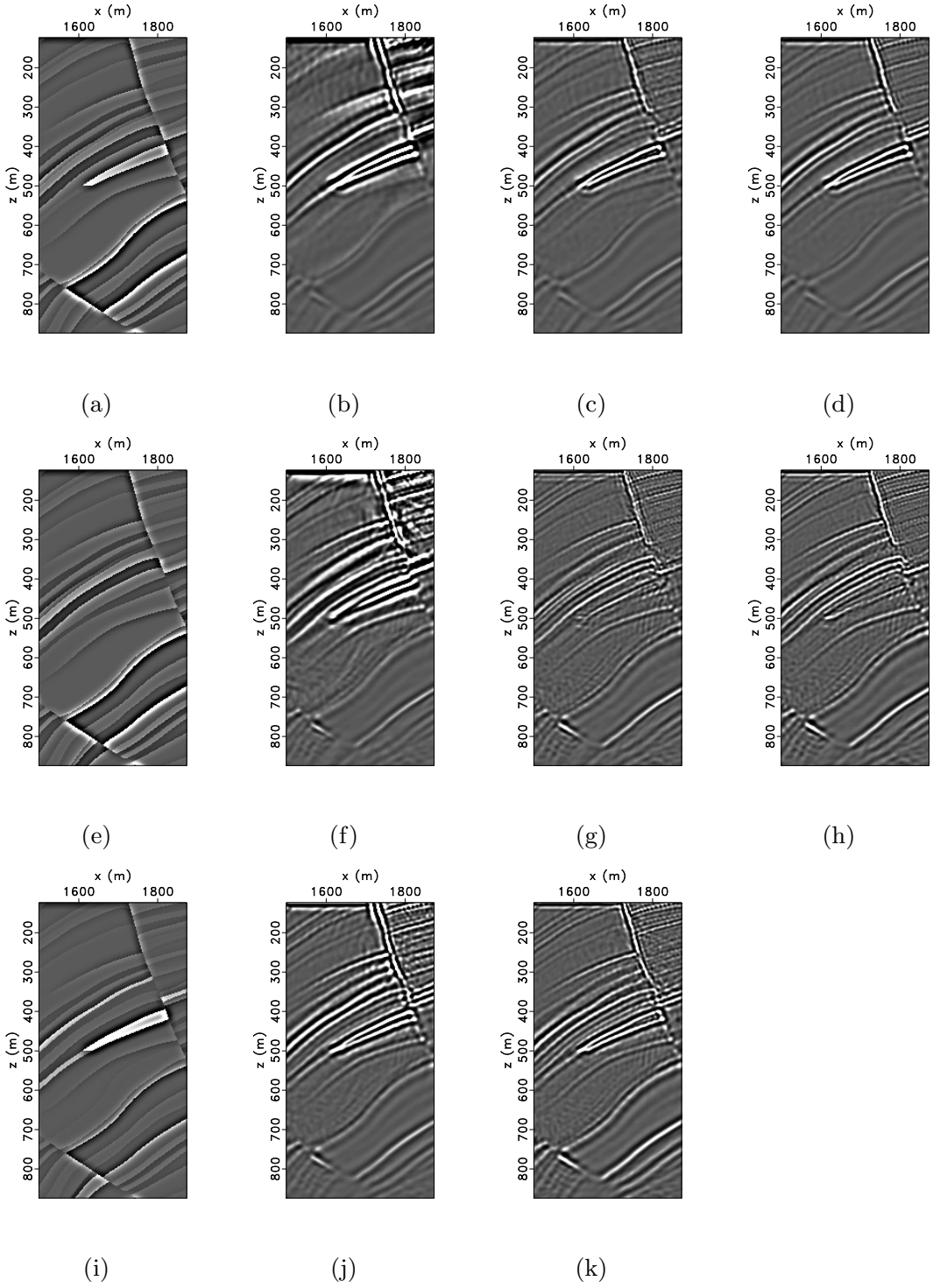


Figure 4.11: Comparison of images in the range 1500 m $< x <$ 1875 m and 125 m $< z <$ 875 m. P-wave velocity perturbation: (a) true model, (b) elastic RTM, (c) three-parameter elastic LSRTM, (d) two-parameter elastic LSRTM. S-wave velocity perturbation: (e) true model, (f) elastic RTM, (g) three-parameter elastic LSRTM, (h) two-parameter elastic LSRTM. Density perturbation: (i) true model, (j) elastic RTM, (k) three-parameter elastic LSRTM.

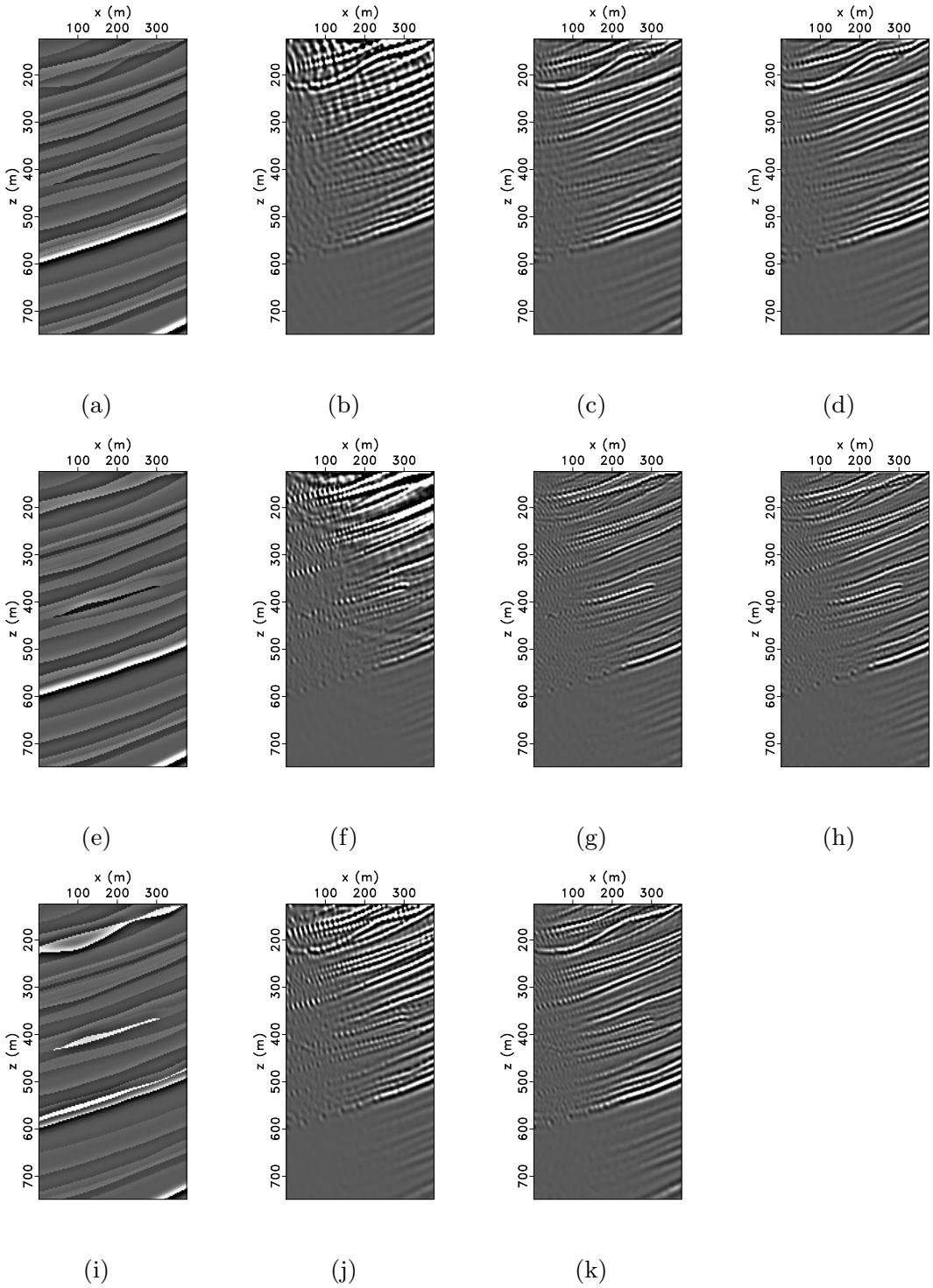


Figure 4.12: Comparison of images in the range 0 m $< x <$ 375 m and 125 m $< z <$ 750 m. P-wave velocity perturbation: (a) true model, (b) elastic RTM, (c) three-parameter elastic LSRTM, (d) two-parameter elastic LSRTM. S-wave velocity perturbation: (e) true model, (f) elastic RTM, (g) three-parameter elastic LSRTM, (h) two-parameter elastic LSRTM. Density perturbation: (i) true model, (j) elastic RTM, (k) three-parameter elastic LSRTM.

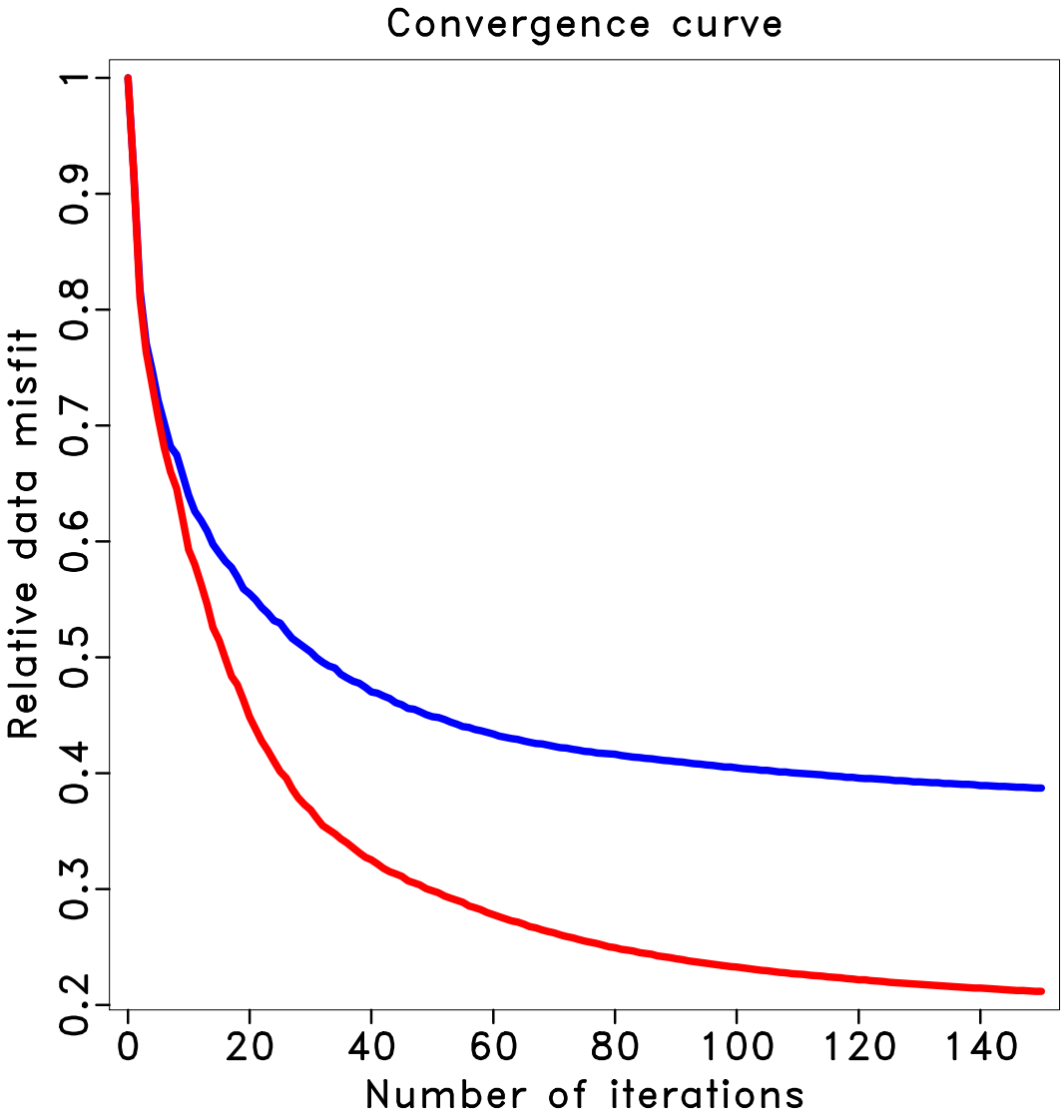


Figure 4.13: Comparison of the convergence curves of three-parameter elastic LSRTM (red) and two-parameter elastic LSRTM (blue) for elastic Marmousi2 model.

CHAPTER 5

Time-domain matrix-free elastic Gauss-Newton full-waveform inversion via adjoint-state method [1]

We propose a time-domain matrix-free elastic Gauss-Newton full-waveform inversion (FWI) algorithm. The elastic Gauss-Newton FWI is considered as an iterative elastic LSRTM problem. The proposed algorithm consists of two nested iteration loops: the outer Gauss-Newton nonlinear iterations and the inner conjugate gradient least-squares (CGLS) linear iterations. The Gauss-Newton search direction in each outer FWI iteration is computed using the CGLS method. This step is equivalent to applying elastic LSRTM on data residuals. The Fréchet derivative operator is the elastic Born modelling operator and the adjoint of the Fréchet derivative operator is the elastic RTM operator. The CGLS algorithm can be safely used for solving the Gauss-Newton search direction because our discretized numerical versions of elastic Born and RTM operators passed the dot-product test. Our proposed algorithm is matrix-free because it only requires the forward and adjoint Fréchet derivative operators applied "on the fly" to vectors. The latter is achieved via the adjoint-state method. We use the proposed algorithm to simultaneously invert for P- and S-wave velocities. The proposed elastic Gauss-Newton FWI generates slightly better-inverted models than the nonlinear conjugate gradient (NLCG) method. The elastic Gauss-Newton FWI converges faster than the elastic NLCG FWI.

5.1 Introduction

Full-waveform inversion (FWI) (Bamberger et al., 1982; Lailly, 1983; Tarantola, 1984a; Pratt, 1999; Virieux and Operto, 2009) aims at estimating subsurface model parameters

[1] A version of this chapter is in preparation for submission for publication.

using full-wave mode data recorded by seismic receivers. It is a nonlinear and ill-posed inverse problem. Tarantola (1984a) formulates time-domain acoustic FWI as a nonlinear least-squares inverse problem (Tarantola and Valette, 1982). The inverse problem was proposed to be solved using a gradient-based iterative method. Tarantola (1984a) adopts the adjoint-state method (Lions, 1971) to compute the gradient of the objective function to avoid the expensive explicit computation of the Fréchet derivative (Bamberger et al., 1982; Lailly, 1983). Gauthier et al. (1986) conduct the first numerical study of 2D acoustic FWI in the time domain with synthetic seismic reflection and transmission data. Gauthier et al. (1986) point out the importance of transmission data in the recovering of the low spatial frequency content of the model. Bunks et al. (1995) propose a multiscale method to mitigate the cycle-skipping problem in time-domain FWI. Subsequently, a more efficient FWI formulation was proposed in the frequency domain (Song et al., 1995; Pratt et al., 1996, 1998; Pratt, 1999; Pratt and Shipp, 1999; Ravaut et al., 2004).

The above-mentioned work is mainly focused on monoparameter acoustic inversion. Tarantola (1986) extends the FWI theory to isotropic elastic earth media. By analyzing diffraction patterns, Tarantola (1986) proposes a hierarchical inversion strategy for time-domain elastic FWI. For the long wavelengths of the model, the inversion is parameterized in terms of P- and S-wave velocities. For the short wavelengths of the model, he suggests first optimizing for the P-wave impedance, then optimizing for the S-wave impedance, and finally optimizing for the density. Mora (1987b, 1988) numerically study the 2D time-domain elastic FWI of synthetic reflection and transmission multicomponent data. The three isotropic elastic parameters are simultaneously inverted for. For reflection data inversion, Mora (1987b) points out that the coupling between P-wave velocity and density are not well resolved when using the velocity-density parameterization. For reflection and transmission data inversion, Mora (1988) observes that the low-wavenumber content of density cannot be resolved because density has little effect on seismic wave traveltimes. Crase et al. (1990) propose a robust time-domain elastic FWI based on the ℓ_1 norm minimization criterion to invert for short wavelengths of both P- and S-wave impedances. The latter is applied to a real marine streamer dataset (Crase et al., 1990) and to land seismic reflection data (Crase et al., 1992). The 2D time-domain elastic FWI has also been applied to real multicomponent ocean-bottom cable (OBC) seismic data using a hierarchical inversion scheme (Singh et al., 2008; Sears et al., 2008, 2010). The inversion scheme entails first inverting for P-wave velocity using the vertical component, second inverting for S-wave velocity using the vertical component, and finally inverting for S-wave velocity using horizontal component. The density is updated using Gardner's empirical relationship for P-wave velocity. Kohn et al. (2012) investigate the influence of model parameterization in 2D time-domain elastic FWI using synthetic examples. The three isotropic elastic parameters are simultaneously optimized with a hard constraint applied on the density during inversion. Time-domain elastic FWI

has also been extended to three dimensional (Guasch et al., 2012; Vigh et al., 2014; Borisov and Singh, 2015; Albertin et al., 2016; Borisov et al., 2018; Oh et al., 2018). Even on modern high-performance computers, 3D time-domain elastic FWI is still computational challenging. In these studies, the P- and S-wave velocities are simultaneously inverted for. The density is either kept constant (Vigh et al., 2014; Albertin et al., 2016; Borisov et al., 2018; Oh et al., 2018) or updated using empirical relationship during inversion (Guasch et al., 2012; Borisov and Singh, 2015). The development of elastic FWI has also seen significant progress in the frequency domain. Gelis et al. (2007) implemented a 2D frequency-domain elastic FWI using Born and Rytov approximations on synthetic data. Choi et al. (2008a,b) proposed a 2D frequency-domain elastic FWI for density, P- and S-wave velocities. Brossier et al. (2009b) investigate a 2D frequency-domain elastic FWI for synthetic land seismic surface and body wave data. Brossier et al. (2009a, 2010) apply the ℓ_1 norm data misfit function in the 2D frequency-domain elastic FWI. Romdhane et al. (2011) apply 2D frequency-domain elastic FWI for characterizing 2D shallow structure in the presence of complex topography. Pageot et al. (2013) apply 2D frequency-domain elastic FWI on synthetic teleseismic data for lithospheric imaging. The density model is assumed to be constant in those studies (Gelis et al., 2007; Brossier et al., 2009b,a; Romdhane et al., 2011; Pageot et al., 2013). Jeong et al. (2012) propose a hierarchical inversion scheme for 2D frequency-domain elastic FWI. In the first stage, the Lamé parameters are optimized with fixed density. In the second stage, the density and Lamé parameters are simultaneously optimized.

The convergence rate of gradient-based optimization methods is linear. Newton's methods converge quadratically, and the convergence rate is grid independent (Santosa and Symes, 1988; Burstedde and Ghattas, 2009). The first discussion of Gauss-Newton methods for time-domain FWI problem can be traced back to Tarantola (1984a) where they call it the "total inversion" method (Tarantola and Valette, 1982). Santosa and Symes (1988) proposed a truncated Newton method and apply it to 1D time-domain acoustic FWI problem. The Hessian is not explicitly computed and stored. In each Newton's iteration, the Hessian is implicitly inverted using the conjugate gradient algorithm. The latter requires only computation of the action of the Hessian operator on a vector with the help of the second-order adjoint-state method. Santosa and Symes (1988) also discussed the Gauss-Newton method where the second order derivative is discarded. Akcelik et al. (2002) implemented a 3D time-domain acoustic Gauss-Newton-Krylov FWI. The latter is an extension of Santosa and Symes (1988)'s Gauss-Newton method to three dimensions. The Gauss-Newton Hessian is implicitly inverted using the symmetric conjugate gradient algorithm. Epanomeritakis et al. (2008) extended Gauss-Newton-Krylov FWI to the 3D time-domain elastic case (Akcelik et al., 2002). However, Epanomeritakis et al. (2008) only invert for a single parameter (shear modulus) in their numerical examples. Burstedde and Ghattas (2009) investigate a 1D time-domain acoustic Newton FWI using the Newton-Krylov method proposed in (San-

tosa and Symes, 1988). The conjugate gradient is terminated early to avoid directions of negative curvature (Burstedde and Ghattas, 2009). Anagaw and Sacchi (2012) and Metivier et al. (2013, 2014) apply the truncated Newton method to 2D frequency-domain acoustic FWI.

Pratt et al. (1998) formulated the Gauss-Newton and Newton method in the frequency domain. The Gauss-Newton Hessian or the full Hessian matrix is explicitly computed and inverted. Time-domain elastic Gauss-Newton FWI has also been investigated in Sheen et al. (2006). In Sheen et al. (2006), the Fréchet derivative and Hessian matrices are computed explicitly using the reciprocity of the Green's function. Pan et al. (2016) discussed the Gauss-Newton and full-Newton method in elastic HTI media. However, it is also based on explicitly computing and inverting the Hessian matrix. Obviously, those methods are too expensive and unfeasible for application to large-scale problems.

In this chapter, I propose a time-domain matrix-free elastic Gauss-Newton FWI algorithm (Chen and Sacchi, 2017c). The proposed algorithm consists of two loops of iterations: the outer Gauss-Newton nonlinear iterations (Nocedal and Wright, 2006) and the inner conjugate gradient least-squares (CGLS) linear iterations (Hestenes and Stiefel, 1952; Paige and Saunders, 1982). The outer nonlinear iteration uses a parabolic fitting line search (Vigh et al., 2009) to estimate the step size. The Gauss-Newton search direction in each outer FWI iteration is computed using the matrix-free CGLS algorithm. I recognize that this step is actually equivalent to applying an elastic LSRTM to data residuals (Chen and Sacchi, 2017a), with the Fréchet derivative operator as the elastic Born modelling operator and the adjoint of Fréchet derivative operator as the elastic RTM operator. The CGLS algorithm can be safely used for solving the Gauss-Newton search direction because our discretized numerical versions of elastic Born and RTM operators passed the dot-product test (Mora, 1987a; Claerbout, 1992). In the inner CGLS linear iterations, the step size is analytically calculated without of the need of a line search. The inner CGLS linear iterations are preconditioned using the elastic pseudo-Hessian operator (Shin et al., 2001a; Chen and Sacchi, 2017a). The proposed algorithm is matrix-free because it only requires the forward Fréchet derivative and adjoint Fréchet derivative operators applied "on the fly" to vectors. The operators are applied on vectors efficiently via the adjoint-state method (Lions, 1971). The proposed algorithm is adopted to simultaneously invert for P- and S-wave velocities. One has preferred to avoid inverting for density because the seismic traveltimes are not sensitive to density. Only the seismic wave amplitude is sensitive to density. From an analysis of diffraction pattern (Forgues and Lambare, 1997; Tarantola, 1986), the wavefield scattered by a density perturbation is mainly detectable at small aperture angles. The latter means that only short-to-intermediate wavelengths of the density can be reconstructed. If no proper constraint or a priori information is imposed on density, the inverted density by isotropic elastic FWI will deviate from true density. This agrees with observations made in

previous studies (Choi et al., 2008a,b; Jeong et al., 2012; Kohn et al., 2012). The proposed elastic Gauss-Newton FWI generates slightly better-inverted models than the nonlinear conjugate gradient method.

This chapter is organized as follows: First, we describe the system of equations that we have adopted to forward model elastic wavefields. Then, we discuss the general time-domain elastic Gauss-Newton FWI algorithm. Subsequently, we propose to formulate the step of solving Gauss-Newton search direction as an elastic LSRTM problem. We discuss the elastic Born and RTM operators and the numerical adjointness of those operators. Furthermore, we propose to solve the Gauss-Newton search direction using the CGLS algorithm and also propose to precondition the CGLS algorithm by elastic pseudo-Hessian. Then, we summarize the time-domain matrix-free elastic Gauss-Newton FWI algorithm. In the last section, we provide numerical examples that permit one to evaluate the performance of the proposed algorithm.

5.2 Theory

5.2.1 The forward problem

In this study, we consider a 2D heterogeneous, isotropic elastic earth media. The propagation of 2D P-SV waves in this media can be expressed by the elastodynamic equations (Virieux, 1986)

$$\begin{aligned}
&\rho\frac{\partial v_x}{\partial t} - \left(\frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xz}}{\partial z}\right) = 0,\\
&\rho\frac{\partial v_z}{\partial t} - \left(\frac{\partial \sigma_{xz}}{\partial x} + \frac{\partial \sigma_{zz}}{\partial z}\right) = 0,\\
&\frac{\partial \sigma_{xx}}{\partial t} - (\lambda+2\mu)\frac{\partial v_x}{\partial x} - \lambda\frac{\partial v_z}{\partial z} = f_{xx},\\
&\frac{\partial \sigma_{zz}}{\partial t} - (\lambda+2\mu)\frac{\partial v_z}{\partial z} - \lambda\frac{\partial v_x}{\partial x} = f_{zz},\\
&\frac{\partial \sigma_{xz}}{\partial t} - \mu\left(\frac{\partial v_x}{\partial z} + \frac{\partial v_z}{\partial x}\right) = 0,
\end{aligned} \tag{5.1}$$

where v_x and v_z are the horizontal and vertical particle velocity fields, σ_{xx}, σ_{xz} and σ_{zz} are the stress fields, f_{xx} and f_{zz} are the explosive source terms, ρ is density, λ and μ the Lamé coefficients. These coefficients describe the spatially variable property of the earth. The latter is related to the seismic P- and S- wave velocities via $\lambda + 2\mu = \rho V_p^2$ and $\mu = \rho V_s^2$. To make the notation concise, we dropped the dependency on spatial and temporal coordinates $\mathbf{x}$ and t of our variables. Using a more general operator notation, the elastic wave equation

may be compactly written as

$$\mathbf{S}(\mathbf{m})\mathbf{u} = \mathbf{f}, \tag{5.2}$$

where $\mathbf{u} = (v_x, v_z, \sigma_{xx}, \sigma_{zz}, \sigma_{xz})^T$ is the wavefield vector, $\mathbf{f} = (0, 0, f_{xx}, f_{zz}, 0)^T$ is the source vector, $\mathbf{S}$ is the wave equation operator, and $\mathbf{m}$ denotes the model parameter vector. The wavefield $\mathbf{u} = \mathbf{u}(\mathbf{m})$ can be considered as a nonlinear implicit function of the model parameters $\mathbf{m}$. Seismic data are recorded by receivers deployed on the surface of the earth

$$\mathbf{d} = \mathbf{R}\mathbf{u}, \tag{5.3}$$

where $\mathbf{R}$ is the sampling operator (often also called the restriction operator).

5.2.2 Time-domain elastic Gauss-Newton full-waveform inversion

Elastic FWI estimates the subsurface isotropic elastic parameters from the observed seismic data. It is usually formulated as a nonlinear least-squares inverse problem that minimizes the functional (Tarantola, 1984a, 1986)

$$J(\mathbf{m}) = \frac{1}{2}\sum_{i=1}^{N_s} \|\mathbf{d}_i(\mathbf{m}) - \mathbf{d}_i^{obs}\|_2^2, \tag{5.4}$$

where $\mathbf{d}_i(\mathbf{m})$ represents the forward modelled seismic data for ith shot, $\mathbf{d}_i^{obs}$ is the observed seismic data for ith shot, and $\|\cdot\|_2$ denotes the ℓ_2 norm of a vector. Because of the large size of the model space, global optimization methods are prohibitively expensive for solving FWI problem. Instead, the FWI problem is usually solved via iterative local optimization algorithms. Introducing a model perturbation $\mathbf{m} \rightarrow \mathbf{m}+\delta\mathbf{m}$ and a second-order Taylor series expansion, the objective function in the vicinity of $\mathbf{m}$ is given by the following expression (Virieux and Operto, 2009)

$$J(\mathbf{m}+\delta\mathbf{m}) = J(\mathbf{m}) + \delta\mathbf{m}^T\frac{\partial J(\mathbf{m})}{\partial \mathbf{m}} + \frac{1}{2}\delta\mathbf{m}^T\frac{\partial^2 J(\mathbf{m})}{\partial \mathbf{m}^2}\delta\mathbf{m} + \mathcal{O}(\|\delta\mathbf{m}\|^3), \tag{5.5}$$

where $\partial J(\mathbf{m})/\partial\mathbf{m}$ is the gradient and $\partial^2 J(\mathbf{m})/\partial\mathbf{m}^2$ is the Hessian of the objective function. In the vicinity of $\mathbf{m}$, the objective function is linearized and an optimal model update $\delta\mathbf{m}$ should satisfy $\partial J(\mathbf{m}+\delta\mathbf{m})/\partial\delta\mathbf{m}^T = 0$. This leads to the following result

$$\begin{aligned}
\delta\mathbf{m} &= -\left[\frac{\partial^2 J(\mathbf{m})}{\partial \mathbf{m}^2}\right]^{-1}\frac{\partial J(\mathbf{m})}{\partial \mathbf{m}} \\
&= -\left\{\sum_{i=1}^{N_s}\left[\left(\frac{\partial \mathbf{d}_i}{\partial \mathbf{m}}\right)^\dagger\left(\frac{\partial \mathbf{d}_i}{\partial \mathbf{m}}\right) + \left(\frac{\partial^2 \mathbf{d}_i}{\partial \mathbf{m}^2}\right)^\dagger(\mathbf{d}_i - \mathbf{d}_i^{obs})\right]\right\}^{-1}\left[\sum_{i=1}^{N_s}\left(\frac{\partial \mathbf{d}_i}{\partial \mathbf{m}}\right)^\dagger(\mathbf{d}_i - \mathbf{d}_i^{obs})\right],
\end{aligned} \tag{5.6}$$

where † denotes the adjoint of an operator, $\partial \mathbf{d}_i / \partial \mathbf{m}$ is the Fréchet derivative operator. The latter is also called the elastic Born approximation operator. In equation 5.6, the term in the braces is called the full Hessian of the FWI problem (Pratt et al., 1998; Fichtner and Trampert, 2011; Metivier et al., 2013). The second term in the braces corresponds to second-order multiple scattering. This term is small when the objective function is close to a minimum. Dropping the second-order term leads to the Gauss-Newton update (Pratt et al., 1998)

$$\delta \mathbf{m} = - \left[\sum_{i=1}^{N_s} \left(\frac{\partial \mathbf{d}_i}{\partial \mathbf{m}} \right)^{\dagger} \left(\frac{\partial \mathbf{d}_i}{\partial \mathbf{m}} \right) \right]^{-1} \left[\sum_{i=1}^{N_s} \left(\frac{\partial \mathbf{d}_i}{\partial \mathbf{m}} \right)^{\dagger} \left(\mathbf{d}_i - \mathbf{d}_i^{obs} \right) \right], \tag{5.7}$$

where the term in the first bracket is the called the approximated Hessian of FWI

$$\mathbf{H}_a = \sum_{i=1}^{N_s} \left(\frac{\partial \mathbf{d}_i}{\partial \mathbf{m}} \right)^{\dagger} \left(\frac{\partial \mathbf{d}_i}{\partial \mathbf{m}} \right), \tag{5.8}$$

and the term in the second bracket is the gradient of the FWI problem

$$\mathbf{g} = \sum_{i=1}^{N_s} \left(\frac{\partial \mathbf{d}_i}{\partial \mathbf{m}} \right)^{\dagger} \left(\mathbf{d}_i - \mathbf{d}_i^{obs} \right). \tag{5.9}$$

We can observe that the approximated Hessian is the cascade of forward and adjoint Fréchet derivative operators. Moreover, the gradient is obtained by applying the adjoint Fréchet derivative operator to data residuals. The gradient of the elastic FWI problem is equivalent to a prestack elastic RTM algorithm applied to data residuals (Tarantola, 1986). Using equations 5.8 and 5.9, the Gauss-Newton update 5.7 can be simplified as follows

$$\delta \mathbf{m} = -{\mathbf{H}_a}^{-1} \mathbf{g}. \tag{5.10}$$

The elastic Gauss-Newton FWI iteratively minimizes the nonlinear least-squares objective function by updating the model in the Gauss-Newton direction $\delta \mathbf{m}$. The updated model at $(k+1)$th iteration can be written as

$$\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} + \eta^{(k)} \delta \mathbf{m}^{(k)} = \mathbf{m}^{(k)} - \eta^{(k)} (\mathbf{H}_a^{(k)})^{-1} \mathbf{g}^{(k)}, \tag{5.11}$$

where $\mathbf{m}^{(k)}$ is the model at the kth iteration, $\eta^{(k)}$ is the step size at the kth iteration, and $\delta \mathbf{m}^{(k)}$ is the Gauss-Newton search direction at the kth iteration, $\mathbf{H}_a^{(k)}$ is the approximated Hessian at the kth iteration and $\mathbf{g}^{(k)}$ is the gradient at the kth iteration. In this study, we estimate the step size via parabolic fitting. The general time-domain elastic Gauss-Newton FWI algorithm is summarized by Algorithm 3

Algorithm 3 Time-domain elastic Gauss-Newton FWI

Initialize: $\mathbf{m}^{(0)}$
for $k = 0, 1, \cdots$ **while** it does not converge **do**
1. Forward modelling: $\mathbf{d}_i^{(k)} = \mathbf{d}_i(\mathbf{m}^{(k)})$, $i = 1, \cdots, N_s$
2. Compute data residuals: $\delta\mathbf{d}_i^{(k)} = \mathbf{d}_i^{obs} - \mathbf{d}_i^{(k)}$, $i = 1, \cdots, N_s$
3. Compute Gauss-Newton search direction: $\delta\mathbf{m}^{(k)} = -(\mathbf{H}_a^{(k)})^{-1}\mathbf{g}^{(k)}$
4. Compute step size $\eta^{(k)}$ via parabolic fitting line search.
5. $\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} + \eta^{(k)}\delta\mathbf{m}^{(k)}$

end

5.2.3 Solving for the Gauss-Newton step via the CGLS algorithm combined with the adjoint-state method

Formulation of the Gauss-Newton step as an elastic LSRTM problem

In this section, we will drop the FWI iteration index (k) of the approximated Hessian, gradient and model update to make the notation more concise but we understand that these quantities depend on the FWI iteration index. In the definition of the Gauss-Newton update (equation 5.11), the approximated Hessian $\mathbf{H}_a$ needs to be inverted. However, explicitly computing and inverting the Hessian operator are prohibitively expensive for realistic scale problem. Instead, we solve for the Gauss-Newton step update via the matrix-free conjugate gradient least-squares algorithm combing with adjoint-state method. Gauss-Newton update 5.7 can be computed via solving normal equations

$$\sum_{i=1}^{N_s}\left(\frac{\partial\mathbf{d}_i}{\partial\mathbf{m}}\right)^{\dagger}\left(\frac{\partial\mathbf{d}_i}{\partial\mathbf{m}}\right)\delta\mathbf{m} = -\sum_{i=1}^{N_s}\left(\frac{\partial\mathbf{d}_i}{\partial\mathbf{m}}\right)^{\dagger}\left(\mathbf{d}_i - \mathbf{d}_i^{obs}\right), \tag{5.12}$$

or compactly written as,

$$\mathbf{H}_a\delta\mathbf{m} = -\mathbf{g}. \tag{5.13}$$

If we use $\mathbf{L}_i$ to denote the Fréchet derivative operator $(\partial\mathbf{d}_i/\partial\mathbf{m})$ and $\delta\mathbf{d}_i$ to denote the data residual $(\mathbf{d}_i^{obs} - \mathbf{d}_i)$ for ith shot, the normal equations 5.12 become

$$\sum_{i=1}^{N_s}\mathbf{L}_i^{\dagger}\mathbf{L}_i\delta\mathbf{m} = \sum_{i=1}^{N_s}\mathbf{L}_i^{\dagger}\delta\mathbf{d}_i. \tag{5.14}$$

Solving above normal equations is equivalent to solving a least-squares inversion problem that minimizes

$$J_{gn}(\delta\mathbf{m}) = \frac{1}{2}\sum_{i=1}^{N_s}\|\mathbf{L}_i\delta\mathbf{m} - \delta\mathbf{d}_i\|_2^2. \tag{5.15}$$

We recognize that this is actually an elastic LSRTM formulation (Chen and Sacchi, 2017a). The Fréchet derivative operator $\mathbf{L}_i$ is the elastic Born modelling operator and the adjoint of Fréchet derivative operator $\mathbf{L}_i^\dagger$ is the elastic RTM operator. In the following sections, we will briefly show the expressions for these two operators. More details on the derivations of those operators can be found in Chapter 3.

Fréchet derivative operator L: elastic Born modelling operator

As discussed in previous section, the Fréchet derivative of the objective function of elastic FWI is indeed the elastic Born approximation. The elastic Born approximation describes that the incident wavefield hits a scatterer and generates the scattered wavefield. It maps from model perturbation (scatterers) to data perturbation (scattered energy).

$$\delta\mathbf{d} = \mathbf{L}\delta\mathbf{m}. \tag{5.16}$$

The source-side incident wavefield in the current background model is computed using the elastic wave equation

$$\begin{aligned}
&\rho\frac{\partial v_x}{\partial t} - \left(\frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xz}}{\partial z}\right) = 0,\\
&\rho\frac{\partial v_z}{\partial t} - \left(\frac{\partial \sigma_{xz}}{\partial x} + \frac{\partial \sigma_{zz}}{\partial z}\right) = 0,\\
&\frac{\partial \sigma_{xx}}{\partial t} - (\lambda+2\mu)\frac{\partial v_x}{\partial x} - \lambda\frac{\partial v_z}{\partial z} = f_{xx},\\
&\frac{\partial \sigma_{zz}}{\partial t} - (\lambda+2\mu)\frac{\partial v_z}{\partial z} - \lambda\frac{\partial v_x}{\partial x} = f_{zz},\\
&\frac{\partial \sigma_{xz}}{\partial t} - \mu\left(\frac{\partial v_x}{\partial z} + \frac{\partial v_z}{\partial x}\right) = 0.
\end{aligned} \tag{5.17}$$

We assume that the model parameters are perturbed around a background model such that $\rho \to \rho + \delta\rho$, $\lambda \to \lambda + \delta\lambda$ and $\mu \to \mu + \delta\mu$. The wavefields will be perturbed in the sense that $v_x \to v_x + \delta v_x$, $v_z \to v_z + \delta v_z$, $\sigma_{xx} \to \sigma_{xx} + \delta\sigma_{xx}$, $\sigma_{zz} \to \sigma_{zz} + \delta\sigma_{zz}$ and $\sigma_{xz} \to \sigma_{xz} + \delta\sigma_{xz}$. We substitute the perturbed model parameters and wavefields into equation 5.17. After canceling second order terms involving small perturbations, we end up with the following

partial differential equations (Chen and Sacchi, 2017a)

$$
\begin{aligned}
&\rho\frac{\partial\delta v_x}{\partial t}-\left(\frac{\partial\delta\sigma_{xx}}{\partial x}+\frac{\partial\delta\sigma_{xz}}{\partial z}\right)=-\delta\rho\dot{v}_x,\\
&\rho\frac{\partial\delta v_z}{\partial t}-\left(\frac{\partial\delta\sigma_{xz}}{\partial x}+\frac{\partial\delta\sigma_{zz}}{\partial z}\right)=-\delta\rho\dot{v}_z,\\
&\frac{\partial\delta\sigma_{xx}}{\partial t}-(\lambda+2\mu)\frac{\partial\delta v_x}{\partial x}-\lambda\frac{\partial\delta v_z}{\partial z}=(\delta\lambda+\delta\mu)\frac{\dot{\sigma}_{xx}+\dot{\sigma}_{zz}}{2(\lambda+\mu)}+\delta\mu\frac{\dot{\sigma}_{xx}-\dot{\sigma}_{zz}}{2\mu},\\
&\frac{\partial\delta\sigma_{zz}}{\partial t}-(\lambda+2\mu)\frac{\partial\delta v_z}{\partial z}-\lambda\frac{\partial\delta v_x}{\partial x}=(\delta\lambda+\delta\mu)\frac{\dot{\sigma}_{xx}+\dot{\sigma}_{zz}}{2(\lambda+\mu)}-\delta\mu\frac{\dot{\sigma}_{xx}-\dot{\sigma}_{zz}}{2\mu},\\
&\frac{\partial\delta\sigma_{xz}}{\partial t}-\mu\left(\frac{\partial\delta v_x}{\partial z}+\frac{\partial\delta v_z}{\partial x}\right)=\delta\mu\frac{\dot{\sigma}_{xz}}{\mu}.
\end{aligned}
\tag{5.18}
$$

where the vector $\delta\mathbf{u}=(\delta v_x,\delta v_z,\,\delta\sigma_{xx},\delta\sigma_{zz},\delta\sigma_{xz})^T$ is the scattered wavefield due to model perturbations $\delta\rho$, $\delta\lambda$ and $\delta\mu$, and $\mathbf{u}=(v_x,v_z,\sigma_{xx},\sigma_{zz},\sigma_{xz})^T$ is the incident wavefield field in the background model ρ,λ,μ, and the over-dot means the time derivative. The right side of equation 5.18 is the so-called "secondary source". Equation 5.18 is the first-order Born approximation equation in an heterogeneous, isotropic elastic media. The scattered data are obtained by sampling the scattered wavefield at the receiver positions

$$\delta\mathbf{d}=\mathbf{R}\delta\mathbf{u}. \tag{5.19}$$

For elastic FWI problem, parameterization in wave compressional and shear wave speeds is better than a parameterization in Lamé parameters (Tarantola, 1986; Kohn et al., 2012). By adopting the chain rule, the model perturbations obey the following relationships

$$
\begin{pmatrix}\delta\rho\\ \delta\lambda\\ \delta\mu\end{pmatrix}=\begin{pmatrix}1 & 0 & 0\\ V_p^2-2V_s^2 & 2\rho V_p & -4\rho V_s\\ V_s^2 & 0 & 2\rho V_s\end{pmatrix}\begin{pmatrix}\delta\varrho\\ \delta V_p\\ \delta V_s\end{pmatrix}, \tag{5.20}
$$

where $\mathbf{m}=(\rho,V_p,V_s)^T$ represents the background model and $\delta\mathbf{m}=(\delta\varrho,\delta V_p,\delta V_s)^T$ the model perturbation. Equation 5.17, 5.18, 5.19 and 5.20 define the elastic Born modelling (equation 5.16).

Adjoint of Fréchet derivative operator $\mathbf{L}^\dagger$: The elastic RTM operator

The adjoint of the Fréchet derivative of the elastic FWI objective function is the elastic RTM operator. The latter is the adjoint of the elastic Born modelling operator that maps data perturbation to model perturbation.

$$\delta\mathbf{m}^*=\mathbf{L}^\dagger\delta\mathbf{d}, \tag{5.21}$$

where $\delta\mathbf{m}^*$ is the adjoint model perturbation ($\delta\mathbf{m}^* = (\delta\varrho^*, \delta V_p^*, \delta V_s^*)^T$). The adjoint-state equation for state equation 5.1 can be derived using the adjoint-state method (Chen and Sacchi, 2017a)

$$\begin{aligned}
&-\rho\frac{\partial \upsilon_x}{\partial t} + \left(\frac{\partial \varsigma_{xx}}{\partial x} + \frac{\partial \varsigma_{xz}}{\partial z}\right) = \delta d_{v_x},\\
&-\rho\frac{\partial \upsilon_z}{\partial t} + \left(\frac{\partial \varsigma_{xz}}{\partial x} + \frac{\partial \varsigma_{zz}}{\partial z}\right) = \delta d_{v_z},\\
&-\frac{\partial \varsigma_{xx}}{\partial t} + (\lambda+2\mu)\frac{\partial \upsilon_x}{\partial x} + \lambda\frac{\partial \upsilon_z}{\partial z} = 0,\\
&-\frac{\partial \varsigma_{zz}}{\partial t} + (\lambda+2\mu)\frac{\partial \upsilon_z}{\partial z} + \lambda\frac{\partial \upsilon_x}{\partial x} = 0,\\
&-\frac{\partial \varsigma_{xz}}{\partial t} + \mu\left(\frac{\partial \upsilon_x}{\partial z} + \frac{\partial \upsilon_z}{\partial x}\right) = 0,
\end{aligned} \tag{5.22}$$

where the vector $\mathbf{p} = (\upsilon_x, \upsilon_z, \varsigma_{xx}, \varsigma_{zz}, \varsigma_{xz})^T$ is the adjoint-state wavefield, the vector $\delta\mathbf{d} = (\delta d_{v_x}, \delta d_{v_z})^T$ represents data residuals. We assume that the observed data are two-component vector particle velocity data. The adjoint-state wave equation represents the data residual injected as an adjoint source and propagated in time-reversal mode. The adjoint-state wavefield is the back-propagated wavefield or sometime called the receiver-side wavefield in the migration jargon. The adjoint model perturbations ($\delta\mathbf{m}^* = (\delta\rho^*, \delta\lambda^*, \delta\mu^*)^T$) can be expressed as (Chen and Sacchi, 2017a)

$$\begin{aligned}
\delta\rho^* &= -\int (\dot{v}_x\upsilon_x + \dot{v}_z\upsilon_z)dt,\\
\delta\lambda^* &= \int \frac{(\dot{\sigma}_{xx}+\dot{\sigma}_{zz})(\varsigma_{xx}+\varsigma_{zz})}{4(\lambda+\mu)^2}dt,\\
\delta\mu^* &= \int \left[\frac{\dot{\sigma}_{xz}\varsigma_{xz}}{\mu^2} + \frac{(\dot{\sigma}_{xx}+\dot{\sigma}_{zz})(\varsigma_{xx}+\varsigma_{zz})}{4(\lambda+\mu)^2} + \frac{(\dot{\sigma}_{xx}-\dot{\sigma}_{zz})(\varsigma_{xx}-\varsigma_{zz})}{4\mu^2}\right]dt.
\end{aligned} \tag{5.23}$$

In elastic FWI, wave speeds are easier to resolve than Lamé parameters (Tarantola, 1986; Forgues and Lambare, 1997). We use the following parameter transformation

$$\begin{pmatrix} \delta\varrho^* \\ \delta V_p^* \\ \delta V_s^* \end{pmatrix} = \begin{pmatrix} 1 & V_p^2 - 2V_s^2 & V_s^2 \\ 0 & 2\rho V_p & 0 \\ 0 & -4\rho V_s & 2\rho V_s \end{pmatrix} \begin{pmatrix} \delta\rho^* \\ \delta\lambda^* \\ \delta\mu^* \end{pmatrix}, \tag{5.24}$$

to transform the Lamé parameters perturbations to wave speeds perturbations. Equation 5.17, 5.22, 5.23 and 5.24 define the elastic RTM operator (equation 5.21).

Solving for the Gauss-Newton step using preconditioned CGLS algorithm

It is important to precondition the linear system of equations to accelerate the convergence of the CGLS algorithm of the Gauss-Newton step. We adopted the elastic pseudo-Hessian for preconditioning the problem (Shin et al., 2001a; Chen and Sacchi, 2017a). The preconditioned version of elastic LSRTM minimizes

$$J_{gn}(\delta\tilde{\mathbf{m}}) = \frac{1}{2}\sum_{i=1}^{N_s} \|\mathbf{L}_i\mathbf{P}\delta\tilde{\mathbf{m}} - \delta\mathbf{d}_i\|_2^2, \tag{5.25}$$

where $\mathbf{P}$ represents the pseudo-Hessian preconditioning. The preconditioned conjugate gradient least squares (PCGLS) algorithm (Bjorck, 1996) is summarized in Algorithm 4. The

Algorithm 4 Preconditioned CGLS algorithm

Initialize
- $\delta\mathbf{m}^{(0)} = \mathbf{0}$
- $\mathbf{r}_i^{(0)} = \delta\mathbf{d}_i,\ i = 1, \cdots, N_s$
- $\mathbf{s}^{(0)} = \mathbf{P}^\dagger\left(\sum_{i=1}^{N_s}\mathbf{L}_i^\dagger\mathbf{r}_i^{(0)}\right)$ `//compute the preconditioned gradient`
- $\mathbf{p}^{(0)} = \mathbf{s}^{(0)}$
- $\gamma^{(0)} = \|\mathbf{s}^{(0)}\|_2^2$

for $l = 0, 1, \cdots$ **while** not converge **do**
- $\mathbf{t}^{(l)} = \mathbf{P}\mathbf{p}^{(l)}$,
- $\mathbf{q}_i^{(l)} = \mathbf{L}_i\mathbf{t}^{(l)},\ i = 1, \cdots, N_s$ `//Born forward modeling`
- $\delta^{(l)} = \sum_{i=1}^{N_s}\|\mathbf{q}_i^{(l)}\|_2^2$
- $\alpha^{(l)} = \gamma^{(l)}/\delta^{(l)}$ `//calculate the step size`
- $\delta\mathbf{m}^{(l+1)} = \delta\mathbf{m}^{(l)} + \alpha^{(l)}\mathbf{t}^{(l)}$ `//update the model`
- $\mathbf{r}_i^{(l+1)} = \mathbf{r}_i^{(l)} - \alpha^{(l)}\mathbf{q}_i^{(l)},\ i = 1, \cdots, N_s$ `//compute data residuals`
- $\mathbf{s}^{(l+1)} = \mathbf{P}^\dagger\left(\sum_{i=1}^{N_s}\mathbf{L}_i^\dagger\mathbf{r}_i^{(l+1)}\right)$ `//compute the preconditioned gradient`
- $\gamma^{(l+1)} = \|\mathbf{s}^{(l+1)}\|_2^2$
- $\beta^{(l)} = \gamma^{(l+1)}/\gamma^{(l)}$
- $\mathbf{p}^{(l+1)} = \mathbf{s}^{(l+1)} + \beta^{(l)}\mathbf{p}^{(l)}$ `//compute the conjugate direction`

end

output of Algorithm 4 is the inverted multiparameter model perturbation $\delta\mathbf{m} = \mathbf{P}\delta\tilde{\mathbf{m}} = (\delta\varrho, \delta V_p, \delta V_s)^T$. The latter is the Gauss-Newton search direction for elastic FWI.

5.2.4 Time-domain matrix-free elastic Gauss-Newton FWI

As discussed in the previous section, we propose a time-domain elastic Gauss-Newton FWI using CGLS algorithm to solve the internal Gauss-Newton linear system of equations. We stress that one does not need to formulate or invert the Hessian matrix explicitly. Instead, the Hessian is iteratively inverted by the CGLS iterations. The CGLS only requires two

operators $\mathbf{L}$ and $\mathbf{L}^\dagger$ that are applied "on the fly" to vectors. The operators $\mathbf{L}$ and $\mathbf{L}^\dagger$ are efficiently applied to vectors via the adjoint-state method. We name this algorithm, the time-domain matrix-free elastic Gauss-Newton FWI. The algorithm is summarized in Algorithm 5.

Algorithm 5 Time-domain matrix-free elastic Gauss-Newton FWI

Initialize: $\mathbf{m}^{(0)}$

for $k = 0, 1, \cdots$ **while** it does not converge **do**

1. Forward modelling: $\mathbf{d}_i^{(k)} = \mathbf{d}_i(\mathbf{m}^{(k)})$, $i = 1, \cdots, N_s$
2. Compute data residual: $\delta\mathbf{d}_i^{(k)} = \mathbf{d}_i^{obs} - \mathbf{d}_i^{(k)}$, $i = 1, \cdots, N_s$
3. Compute Gauss-Newton search direction $\delta\mathbf{m}^{(k)}$ by solving

$$\min_{\delta\tilde{\mathbf{m}}^{(k)}} \quad \frac{1}{2}\sum_{i=1}^{N_s} \|\mathbf{L}_i^{(k)}\mathbf{P}^{(k)}\delta\tilde{\mathbf{m}}^{(k)} - \delta\mathbf{d}_i^{(k)}\|_2^2$$

Apply the preconditioned CGLS algorithm 4.

4. Compute step size $\eta^{(k)}$ via parabolic fitting line search.
5. $\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} + \eta^{(k)}\delta\mathbf{m}^{(k)}$

end

5.3 Numerical Examples

Our software code was developed using the C language and parallelized via the Message Passing Interface (MPI). The parallelization is implemented in shots. Our forward modelling code adopts a time domain staggered-grid finite-difference (FD) scheme (Virieux, 1986) to discretize the elastic wave equation. The algorithm also uses the unsplit Convolutional Perfectly Matched Layer (C-PML) boundary condition (Komatitsch and Martin, 2007) to absorb the artificial reflections arising from the computational boundaries. In our code, the spatial FD order is selectable. The code automatically computes the FD coefficients from the user-specified FD order (Liu and Sen, 2009). Our elastic Born code ($\mathbf{L}$) and RTM code ($\mathbf{L}^\dagger$) pass the dot-product test (Mora, 1987a; Claerbout, 1992; Chen and Sacchi, 2017a). When computing the FWI gradient, we do not save the source-side wavefield. Instead, we adopt source-wavefield reconstruction method (Gauthier et al., 1986; Dussaud et al., 2008) that allows us to only save wavefield on boundaries and the final time frame. The inversion code follows the Algorithm 5. Note that, our elastic Gauss-Newton FWI utilize the multiscale method (Kolb et al., 1986; Bunks et al., 1995). The time-domain data are filtered to form sets of data with different frequency bands. Our FWI code inverts the low-frequency band data first and then inverts high-frequency data. The inverted models from an early scale are used as the initial model for inversion of higher scales. That is to say, there is another extra loop over frequency bands outside the loops provided in Algorithm 5. There are in total three loops of iterations for our elastic Gauss-Newton FWI algorithm.

The proposed time-domain matrix-free elastic FWI method was tested on the elastic inclusion model and the elastic Marmousi2 model (Martin et al., 2006). The observed data are simulated with a time-domain elastic staggered-grid finite-difference method. The C-PML boundary condition was applied on four boundaries of the model. In other words, we do not consider surface waves. The observed data are two-component particle velocity fields. In this study, we simultaneously invert for P- and S- wave velocities and assume that the density is known.

5.3.1 Elastic inclusion model

We design an elastic inclusion model to analyze the resolution of the proposed elastic Gauss-Newton FWI. Figure 5.1 shows the true P- and S-wave velocity models. The square shape velocity anomalies are embedded in a two-layer models. The P- and S-wave velocity models are uncorrelated. Density is fixed at 2000 kg/m^3. The model has a dimension of 2.0 km in horizontal axis and 1.5 km in depth with 201×151 grid points. There are in total 51 shots and 201 receivers deployed along the surface that simulates a fixed-spread survey

geometry. The shot interval is 40 m and receiver interval is 10 m. A 10 Hz central frequency Ricker wavelet is used to simulate an explosive source. The multicomponent observed data are simulated using our elastic staggered-grid finite-difference code. The forward modelling is second-order accurate in time and eighth-order accurate in space ($\mathcal{O}(\Delta t^2, \Delta x^8)$). The observed multicomponent data cubes are shown in Figure 5.2. We compare the results of conventional elastic FWI based on the nonlinear conjugate gradient method (hereafter, we call it elastic NLCG FWI for short) and the proposed elastic Gauss-Newton FWI. In the elastic NLCG FWI, we used the elastic pseudo-Hessian to precondition the gradient. The line search is also based on parabolic fitting. Both the elastic NLCG FWI and the elastic Gauss-Newton FWI utilize the multiscale method (Bunks et al., 1995). The four frequency bands for the multiscale inversion are: 0-5 Hz, 0-10 Hz, 0-15 Hz and 0-30 Hz. Figure 5.3 shows the starting P- and S-wave velocity models for elastic NLCG FWI and elastic Gauss-Newton FWI. The initial models contain two layers without the velocity anomalies.

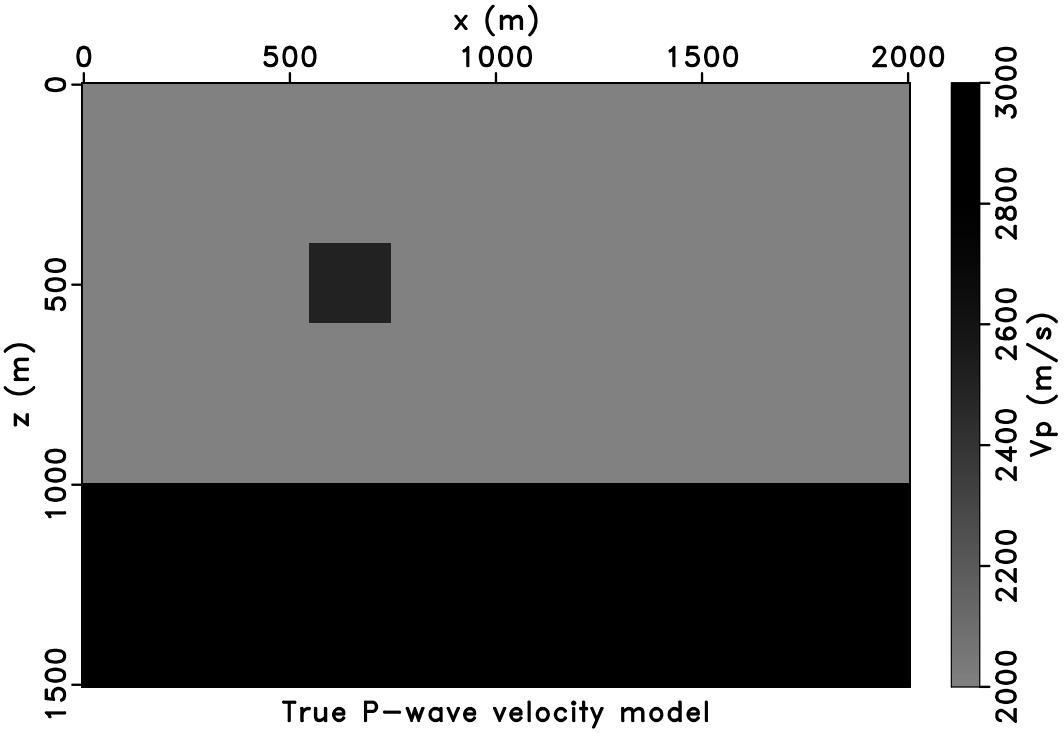
The results of elastic NLCG FWI are shown in Figure 5.4 a and Figure 5.5 a. These results were computed after 50 nonlinear conjugate gradient iterations. The elastic Gauss-Newton FWI has iterated 10 times for the outer FWI loop and 20 times for inner CGLS loop (Figure 5.4 b and Figure 5.5 b). Both elastic NLCG FWI and the proposed elastic Gauss-Newton obtain satisfactory results. The elastic Gauss-Newton FWI provides slightly better results than the elastic NLCG FWI. The results of elastic Gauss-Newton FWI contain fewer artifacts and crosstalk contamination. To show the details more clearly, we display the profiles of the inverted P-wave velocity models in Figure 5.6 and the profiles of the inverted S-wave velocity models in Figure 5.7. We can observe that the elastic Gauss-Newton FWI has fully recovered the amplitude of the velocity anomalies and removed the crosstalk between the P- and S-wave velocity models. We also examine the data fitting of the two elastic FWI algorithms in Figure 5.8 and Figure 5.9. Figure 5.10 compares the converge curves of the elastic NLCG FWI and the elastic Gauss-Newton FWI for four different frequency bands. The relative data misfit for frequency band fb is defined as

$$misfit = \frac{\sum_{i=1}^{N_s} \|\mathbf{d}_i^{fb} - \mathbf{d}_i^{obs^{fb}}\|_2^2}{\sum_{i=1}^{N_s} \|\mathbf{d}_i^{obs^{fb}}\|_2^2}. \tag{5.26}$$

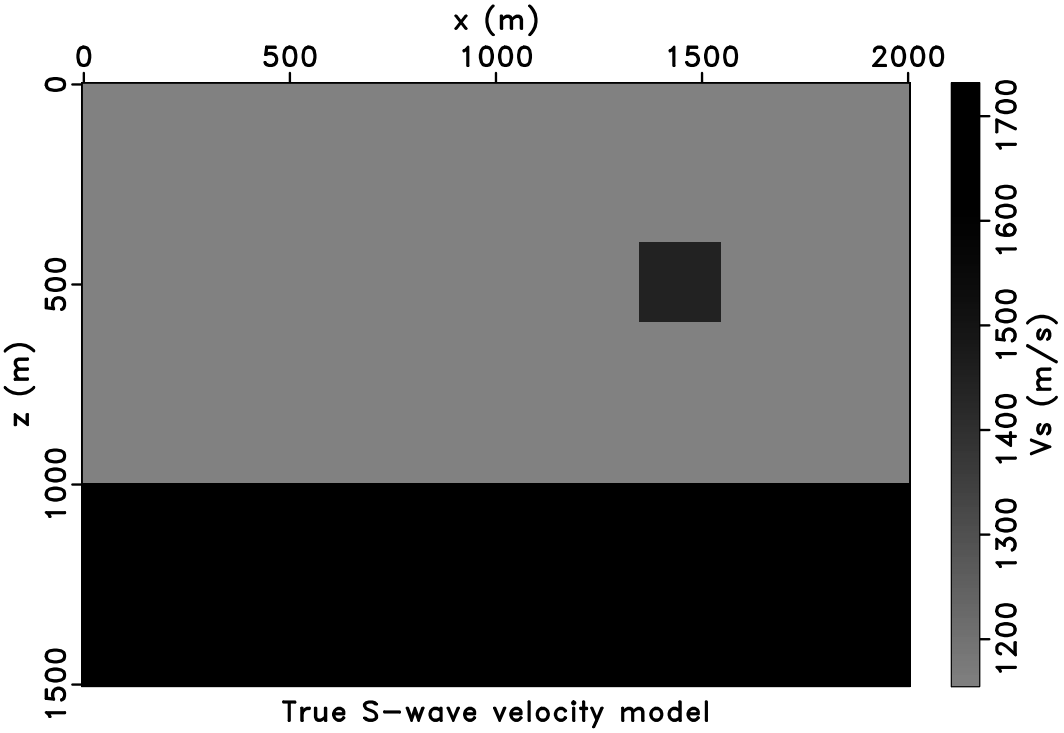
The proposed elastic Gauss-Newton converges much faster than the elastic NLCG FWI.

5.3.2 Elastic Marmousi2 model

Figure 5.11 shows the true P- and S-wave velocity models of the elastic Marmousi2 model. The P- and S-wave velocity models are uncorrelated in hydrocarbon reservoir areas (indicated by the white triangles). Those areas are identified by the low P-wave velocities,

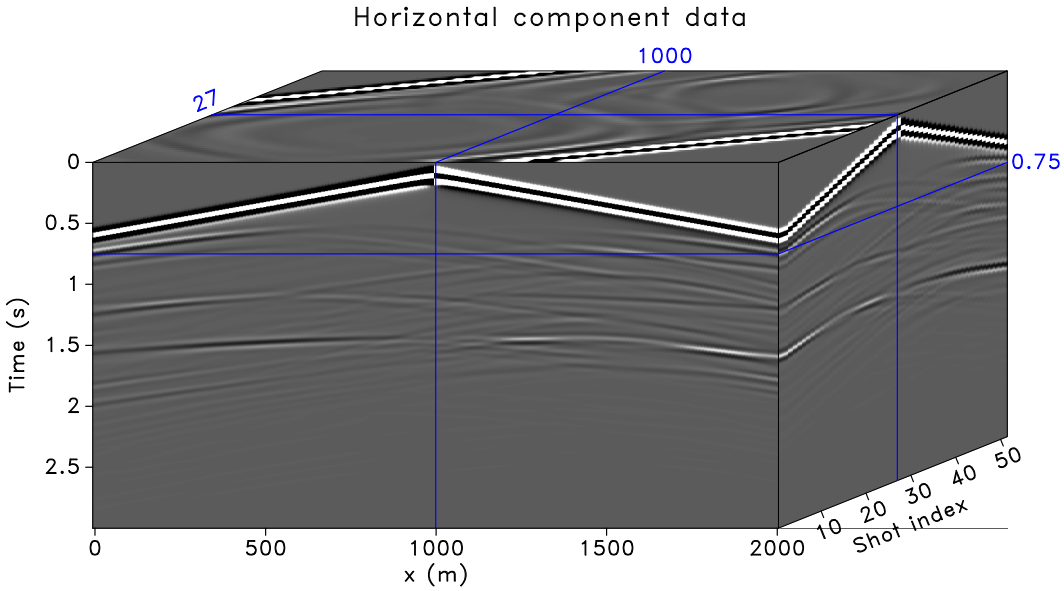


(a)

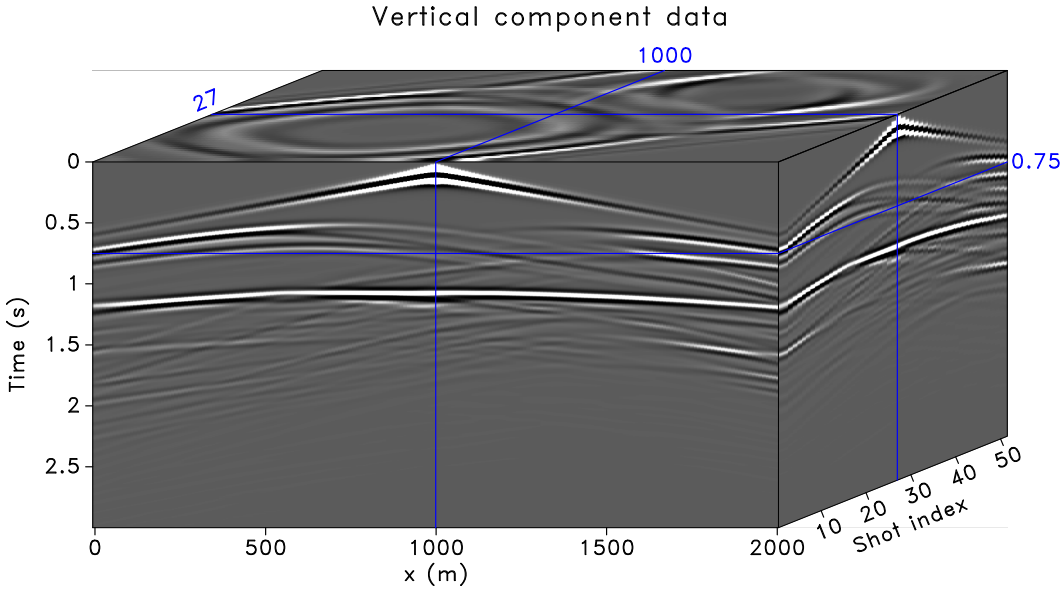


(b)

Figure 5.1: Elastic inclusion model. (a) True P-wave velocity model. (b) True S-wave velocity model.

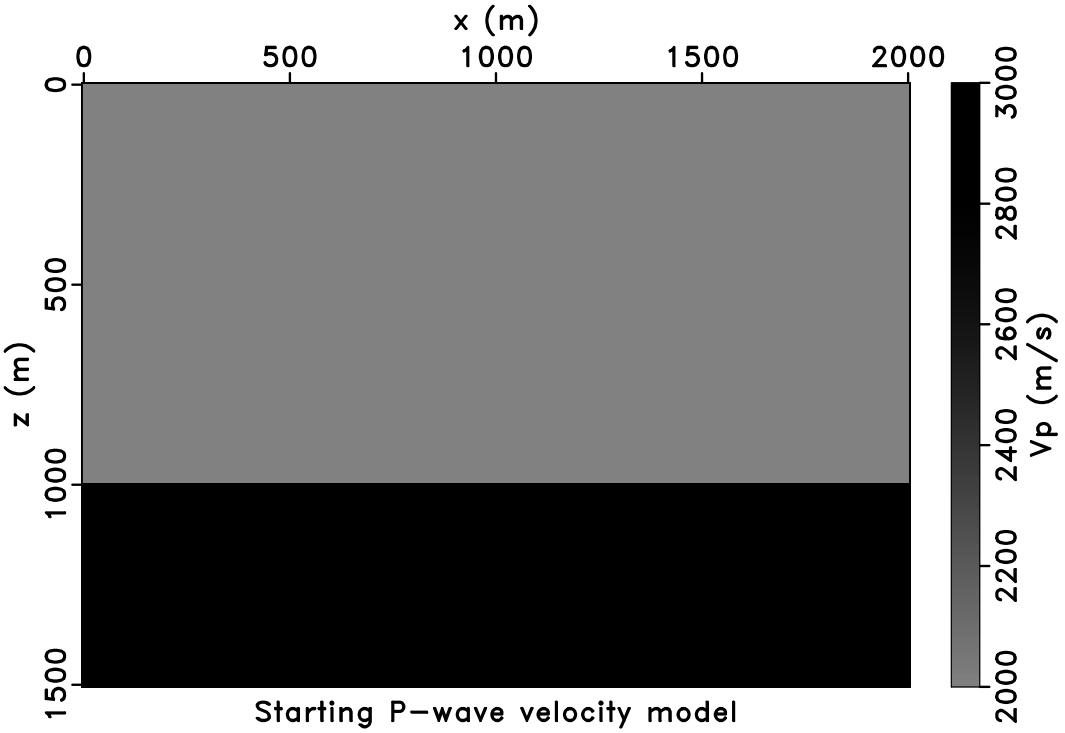


(a)

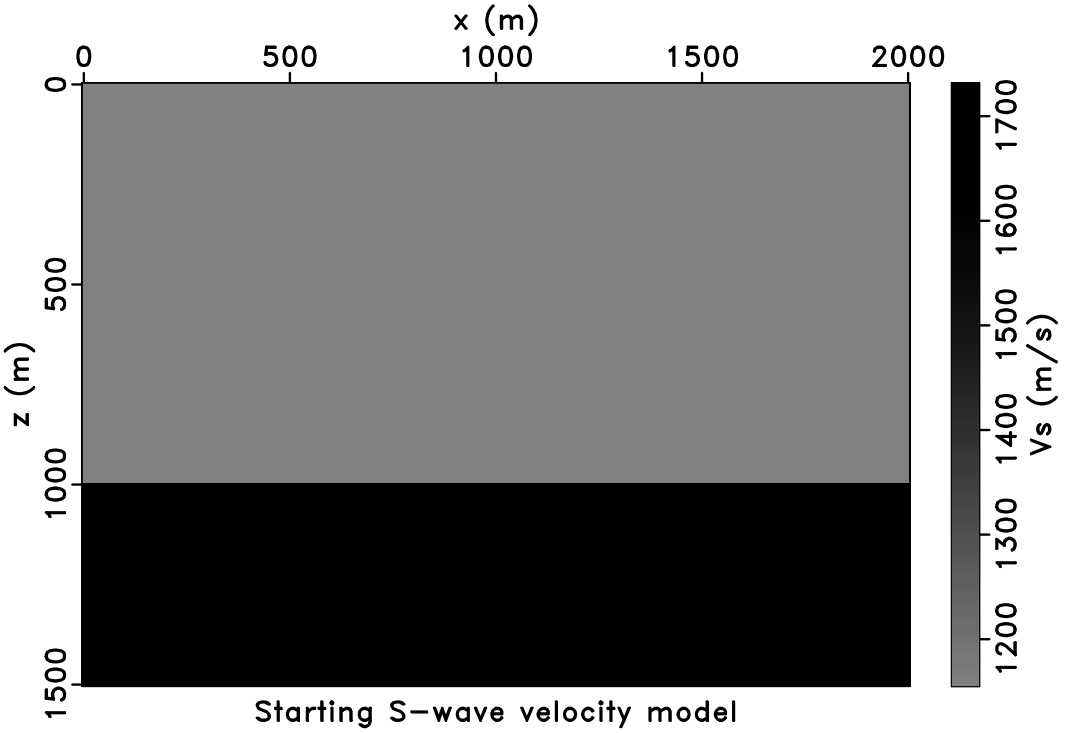


(b)

Figure 5.2: Prestack multicomponent data. (a) Horizontal particle velocity data. (b) Vertical partical velocity data.

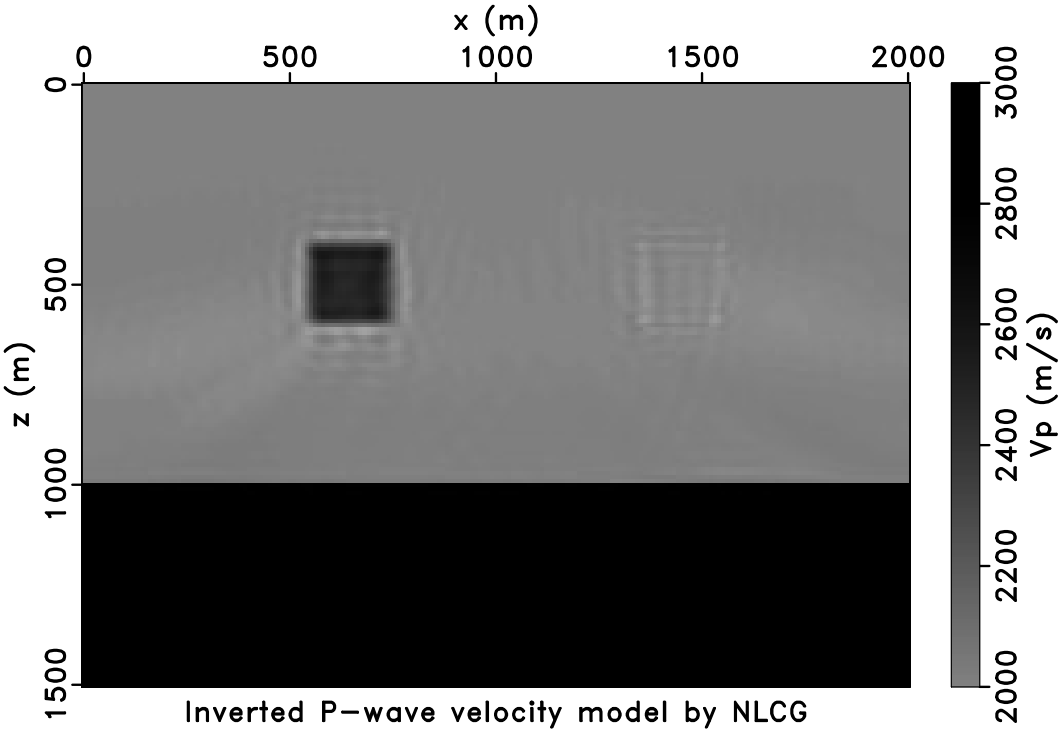


(a)

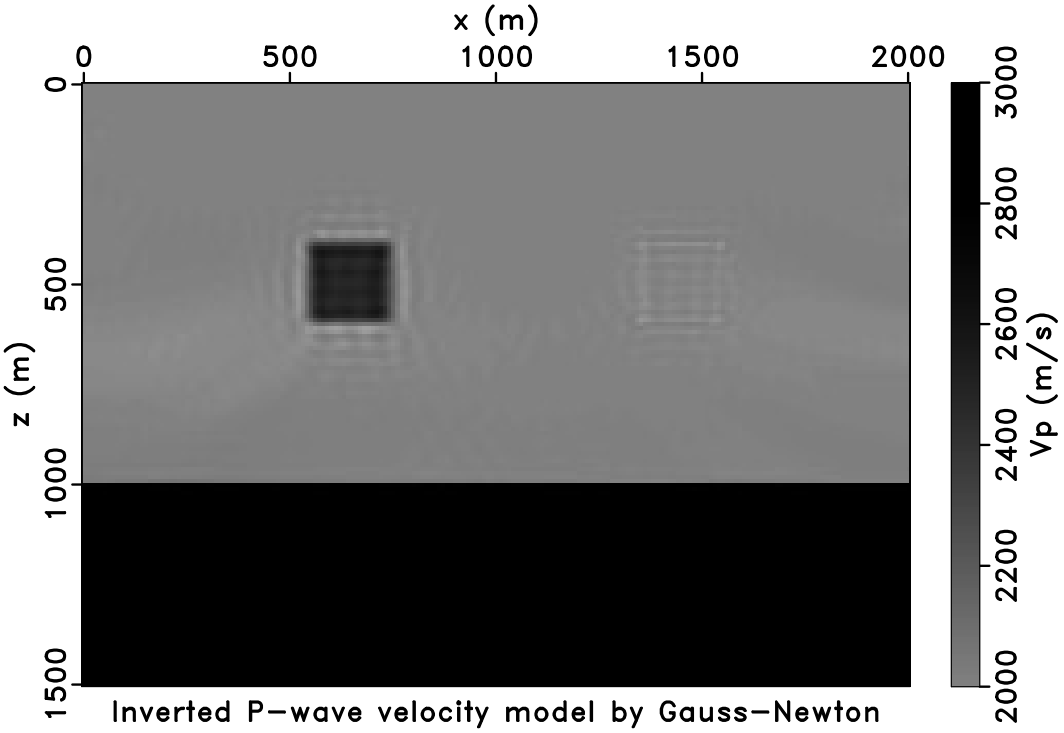


(b)

Figure 5.3: (a) Starting P-wave velocity model. (b) Starting S-wave velocity model.

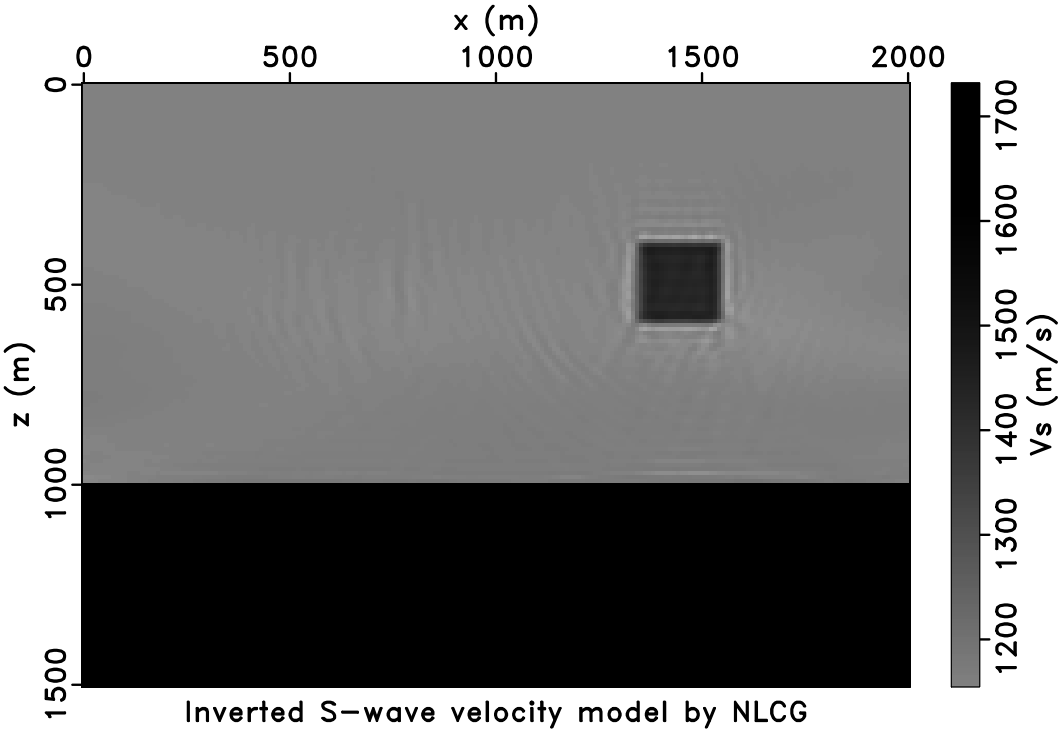


(a)

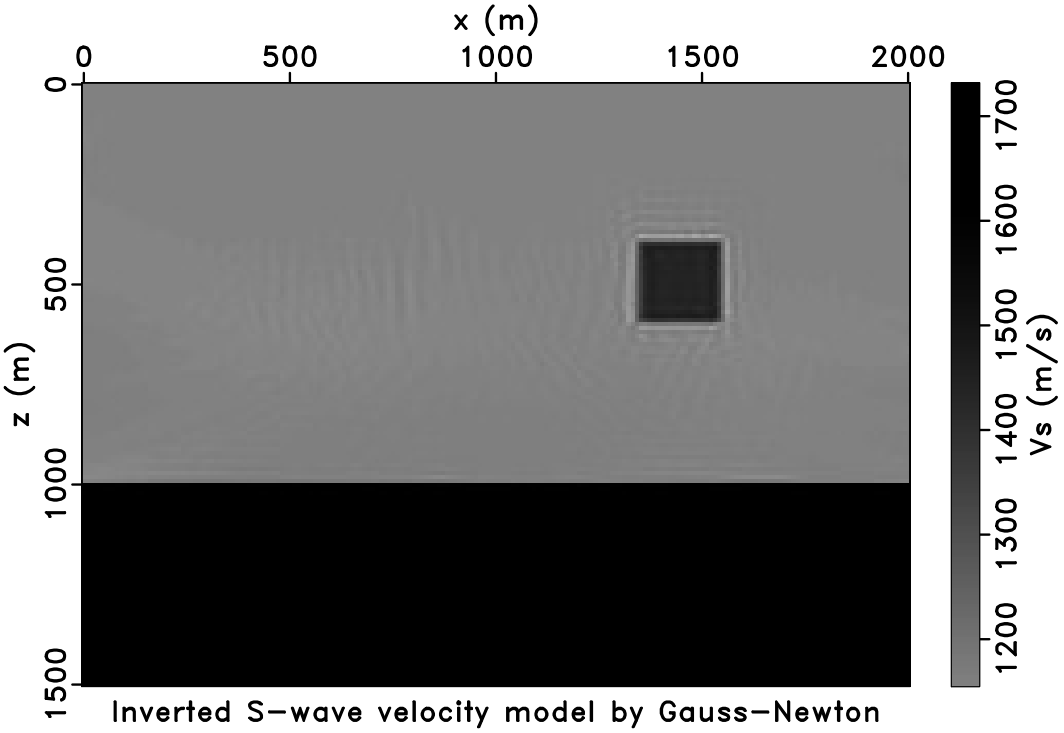


(b)

Figure 5.4: (a) Inverted P-wave velocity model by elastic NLCG FWI. (b) Inverted P-wave velocity model by elastic Gauss-Newton FWI.

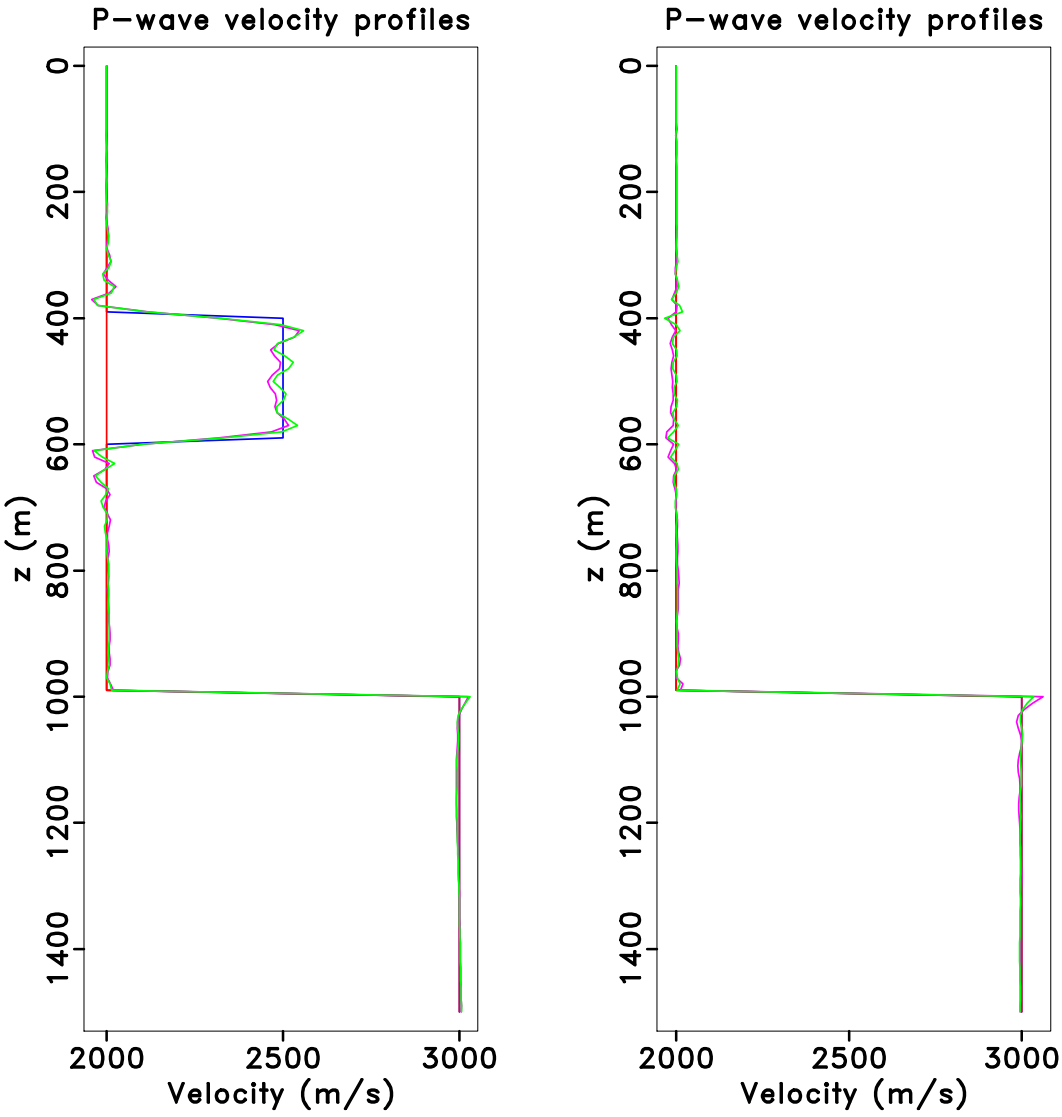


(a)



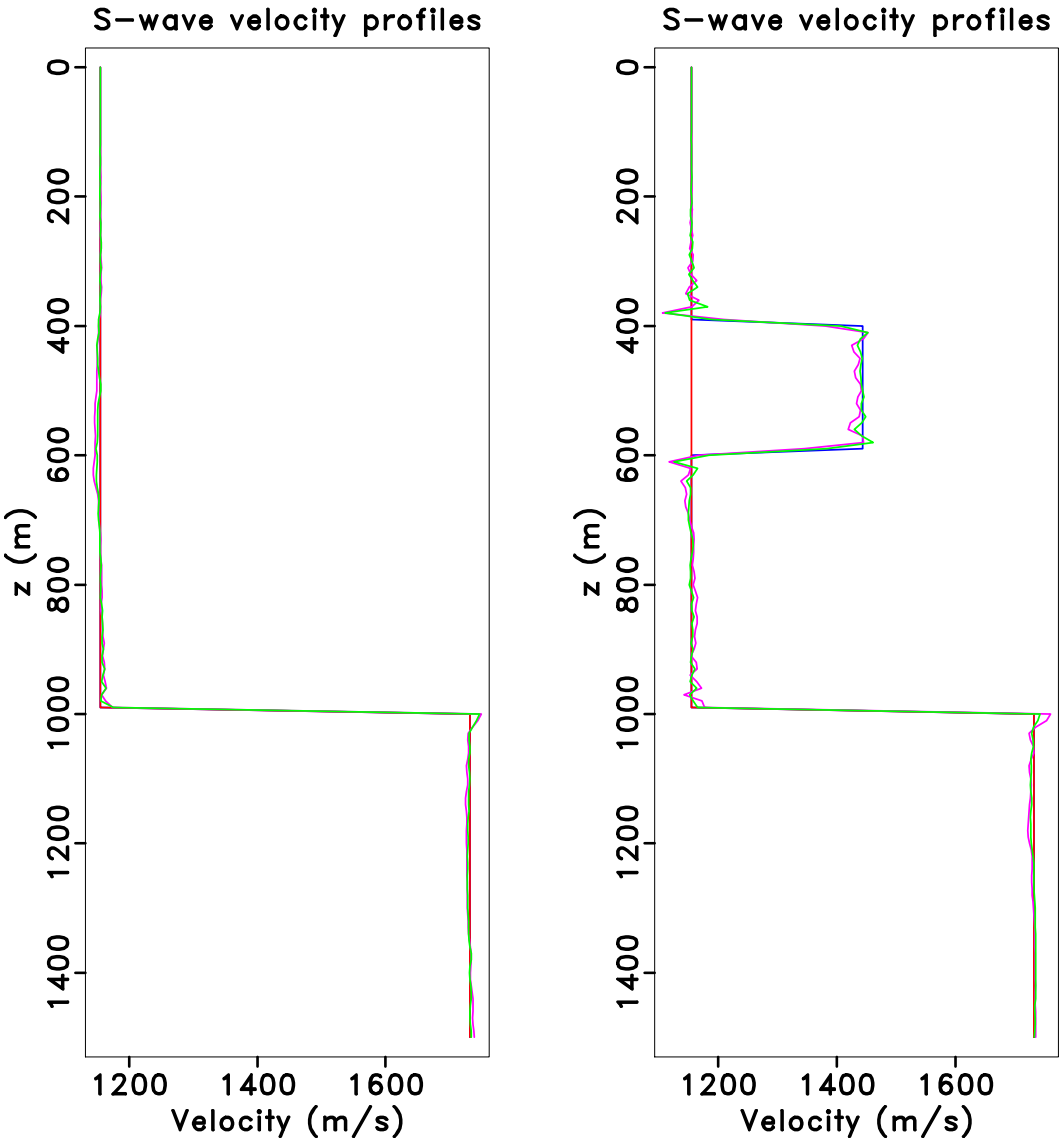
(b)

Figure 5.5: (a) Inverted S-wave velocity model by elastic NLCG FWI. (b) Inverted S-wave velocity model by elastic Gauss-Newton FWI.



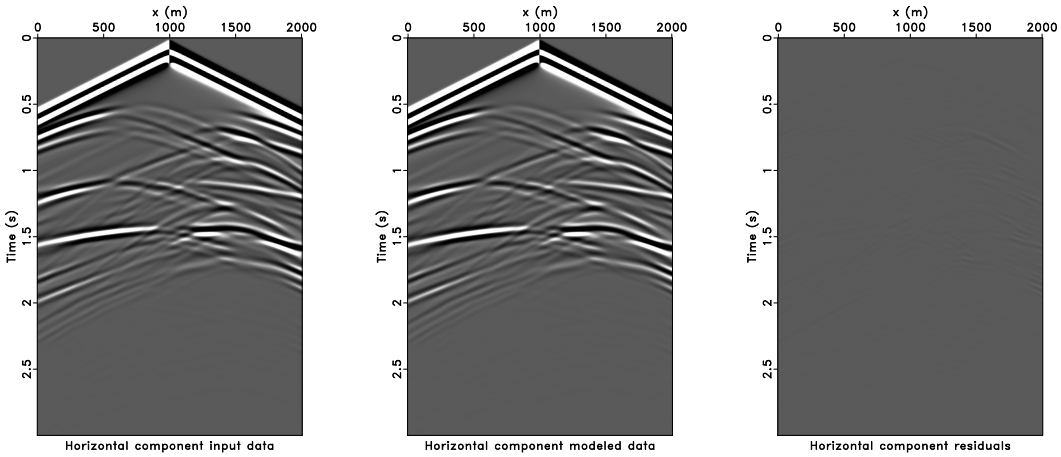
(a) (b)

Figure 5.6: Profiles of inverted P-wave velocity models at (a) $x = 650$ m and (b) $x = 1450$ m. Blue: true model; Red: starting model; Purple: inverted model by elastic NLCG FWI; Green: inverted model by elastic Gauss-Newton FWI.



(a) (b)

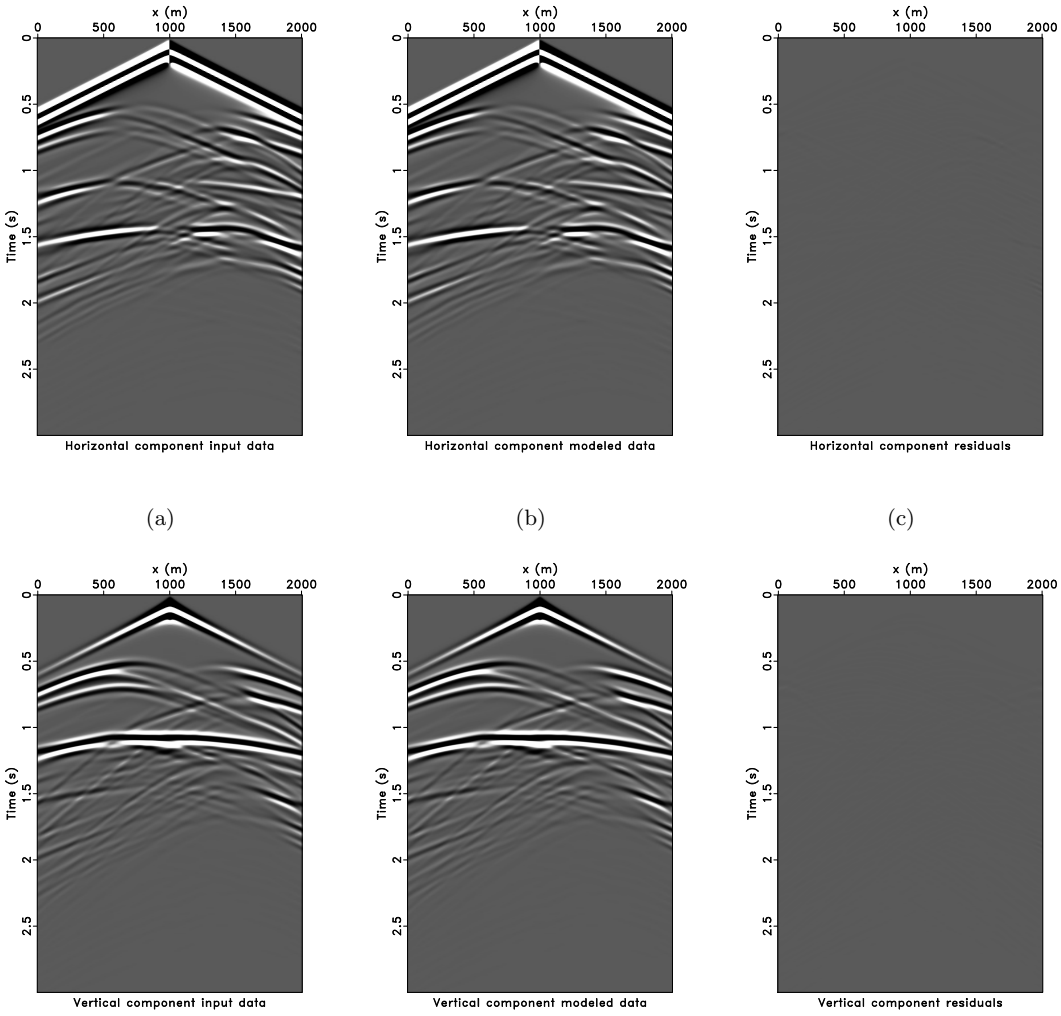
Figure 5.7: Profiles of inverted S-wave velocity models at (a) $x = 650$ m and (b) $x = 1450$ m. Blue: true model; Red: starting model; Purple: inverted model by elastic NLCG FWI; Green: inverted model by elastic Gauss-Newton FWI.



(a) (b) (c)

(d) (e) (f)

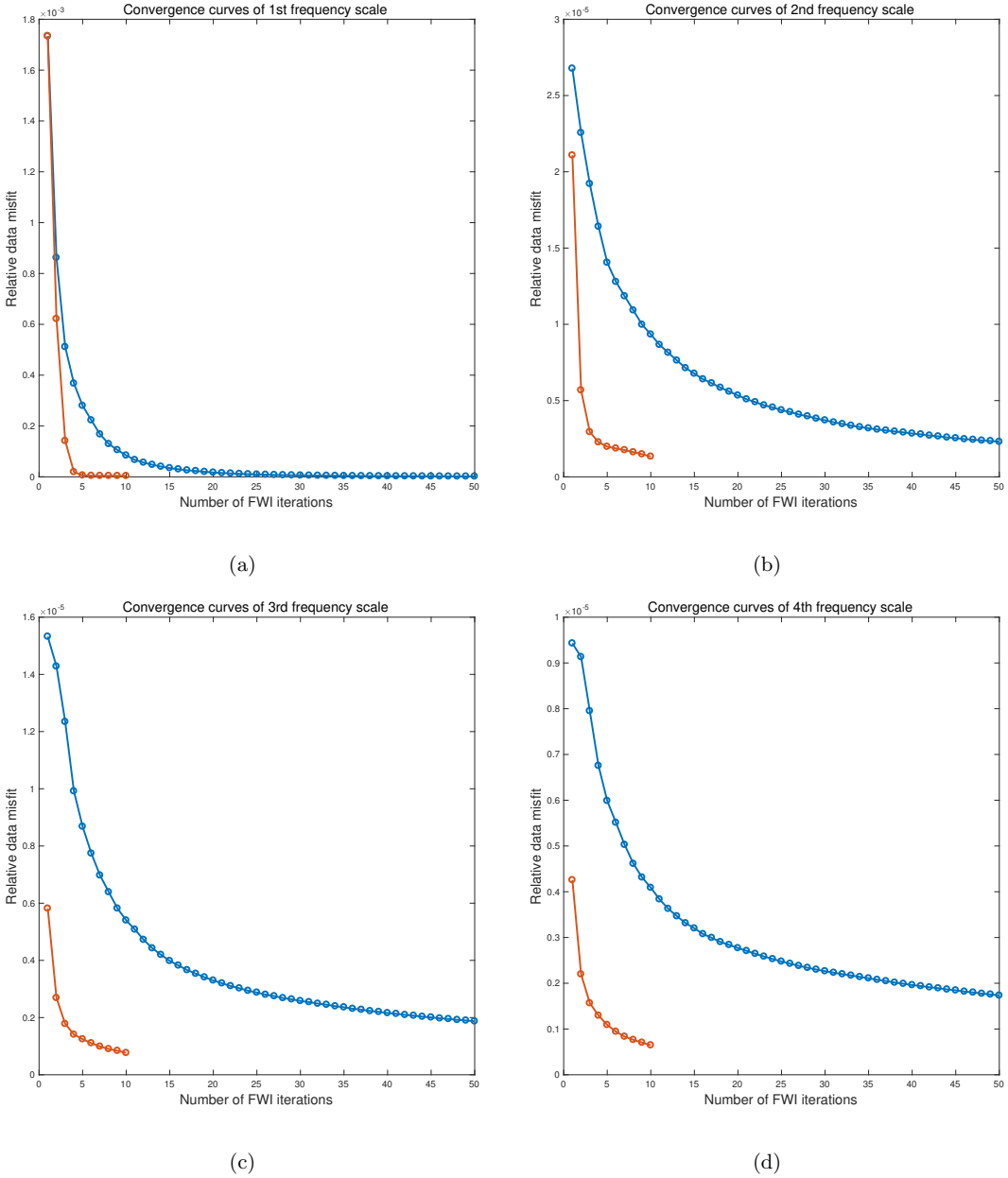
Figure 5.8: Data and residuals for elastic NLCG FWI. (a) Observed horizontal component data of shot at $x = 1000$ m. (b) Horizontal component data modeled by elastic NLCG FWI inverted models. (c) Horizontal component data residual. (d) Observed vertical component data of shot at $x = 1000$ m. (e) Vertical component data modeled by elastic NLCG FWI inverted models. (f) Vertical component data residual.



(a) (b) (c)

(d) (e) (f)

Figure 5.9: Data and residuals for elastic Gauss-Newton FWI. (a) Observed horizontal component data of shot at $x = 1000$ m. (b) Horizontal component data modeled by elastic Gauss-Newton FWI inverted models. (c) Horizontal component data residual. (d) Observed vertical component data of shot at $x = 1000$ m. (e) Vertical component data modeled by elastic Gauss-Newton FWI inverted models. (f) Vertical component data residual.



(a) (b) (c) (d)

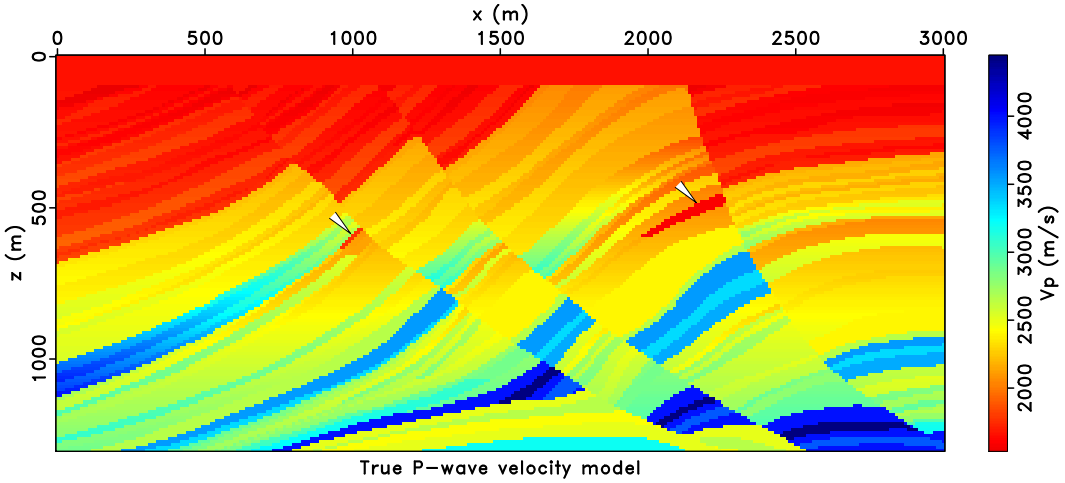
Figure 5.10: Relative data misfit curves for elastic NLCG FWI and elastic Gauss-Newton FWI in frequency band (a) 0 - 5 Hz, (b) 0 - 10 Hz, (c) 0 - 15 Hz, (d) 0 - 30 Hz. Blue: data misfit curves for elastic NLCG FWI. Red: data misfit curves for elastic Gauss-Newton FWI.

whereas their signature is much weaker in the S-wave velocity model. Density is assumed to be constant (2000 kg/m^3). The model has a dimension of 3.0 km in the horizontal axis and 1.3 km in depth with 301×131 grid points. There are 61 shots and 301 receivers deployed along the surface that simulates a fixed-spread survey geometry. The shot interval is 50 m and receiver interval is 10.0 m. A 10 Hz central frequency Ricker wavelet is used to simulate an explosive source. The multicomponent observed data are simulated using our elastic finite-difference code. The forward modelling is second-order accurate in time and twelfth-order accurate in space ($\mathcal{O}(\Delta t^2, \Delta x^{12})$). The observed data cubes are shown in Figure 5.12. The four frequency bands for the multiscale inversion are: 0-2 Hz, 0-5 Hz, 0-10 Hz and 0-30 Hz. Figure 5.13 shows the starting P- and S-wave velocity models for the elastic NLCG FWI and the elastic Gauss-Newton FWI. The initial models are 1D with linearly increasing P and S-wave velocities.

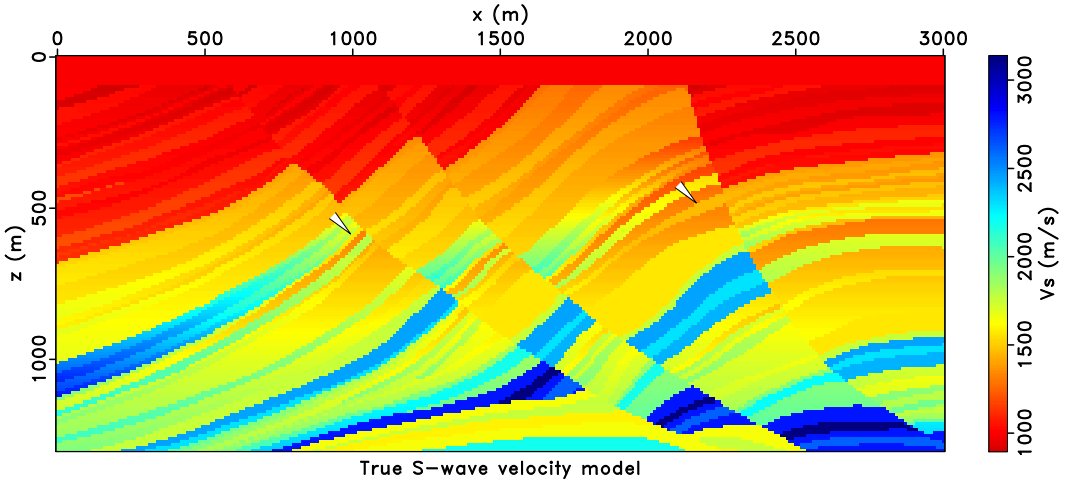
The results of elastic NLCG FWI are shown in Figure 5.14 a and Figure 5.15 a. These results were computed after 50 nonlinear conjugate gradient iterations. The elastic Gauss-Newton FWI has iterated 10 times for outer FWI loop and 20 times for inner CGLS loop (Figure 5.14 b and Figure 5.15 b). The results of elastic Gauss-Newton FWI are slightly better than those of the elastic NLCG FWI. We display the profiles of the inverted P-wave velocity models in Figure 5.16 and the profiles of the inverted S-wave velocity models in Figure 5.17. We also examine the data fitting of the two elastic FWI algorithms in Figure 5.18 and Figure 5.19. Both NLCG and Gauss-Newton elastic FWI predict the observed data well. Figure 5.20 compares the converge curves of the elastic NLCG FWI and the elastic Gauss-Newton FWI for four different frequency bands. We can observe that the elastic Gauss-Newton converges much faster than the elastic NLCG FWI.

5.4 Discussion

The developed time-domain matrix-free elastic Gauss-Newton FWI generates moderately improved results comparing with elastic NLCG FWI. In addition, the elastic Gauss-Newton FWI algorithm is more expensive than the elastic NLCG FWI. This is because the elastic Gauss-Newton FWI algorithm has an additional inner CGLS loop to estimate the search direction. For the numerical examples in this work, the computation time of elastic Gauss-Newton FWI is about three times the computation time of elastic NLCG. There are several strategies to reduce the computational cost of elastic Gauss-Newton FWI. First, one can adopt data contraction strategies such as source encoding, plan-wave synthesis or stochastic source subsampling to reduce the computational cost (Krebs et al., 2009; Vigh and Starr, 2008; van Leeuwen and Herrmann, 2013). Second, it is possible to design an early termination strategy for the inner CGLS iteration in the elastic Gauss-Newton FWI (Metivier

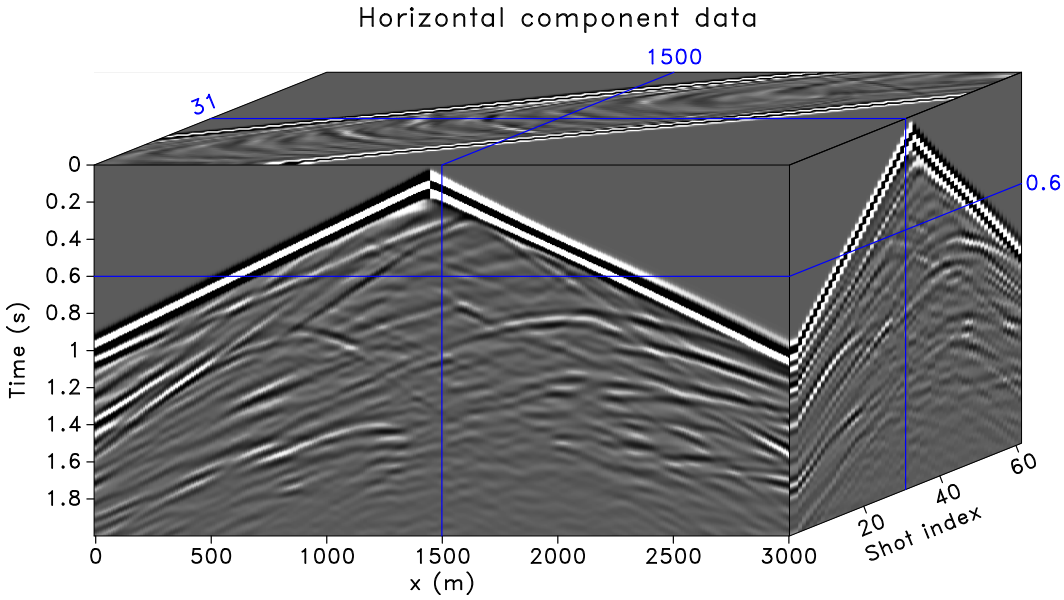


(a)

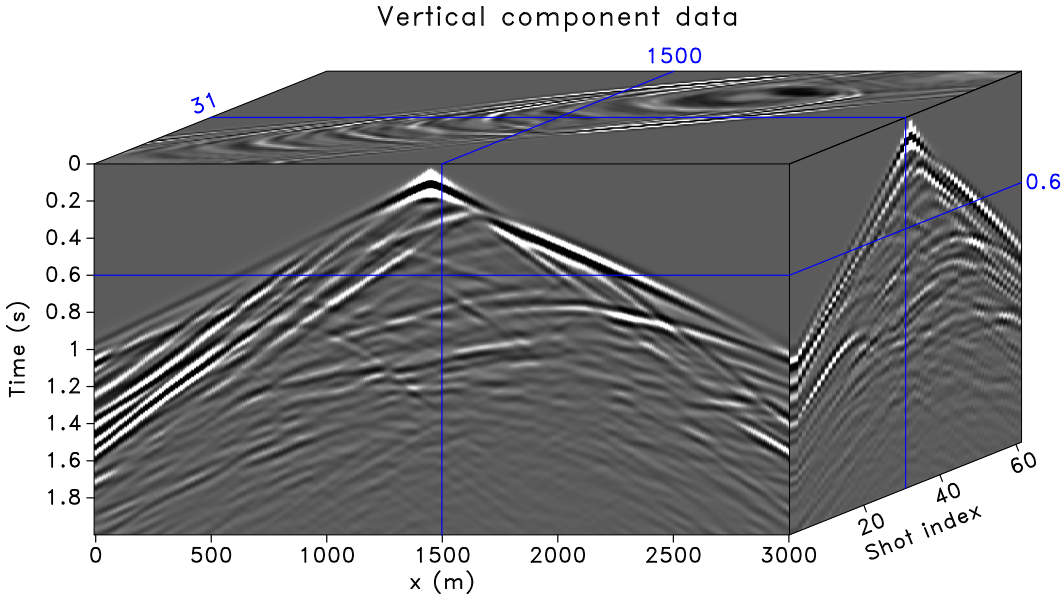


(b)

Figure 5.11: Elastic Marmousi2 model. (a) True P-wave velocity model. (b) True S-wave velocity model.

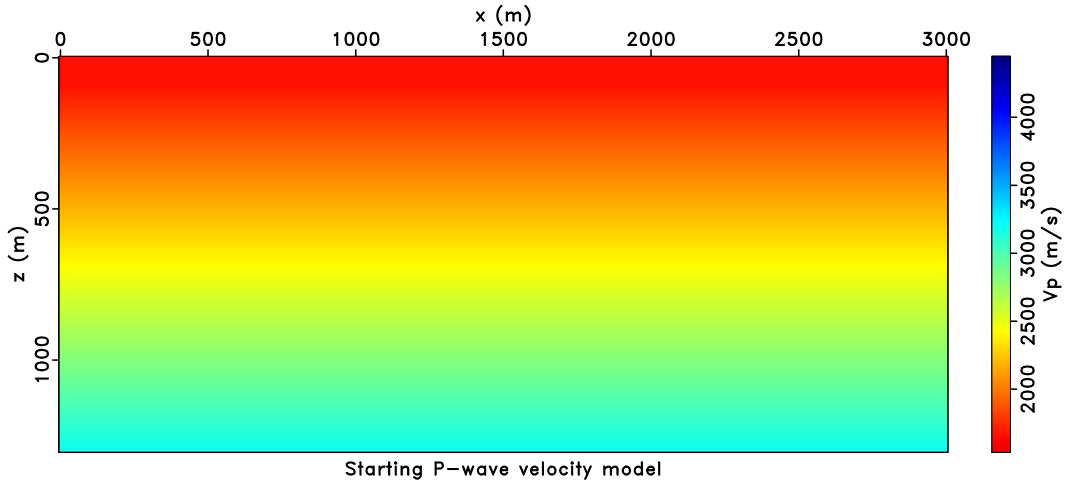


(a)

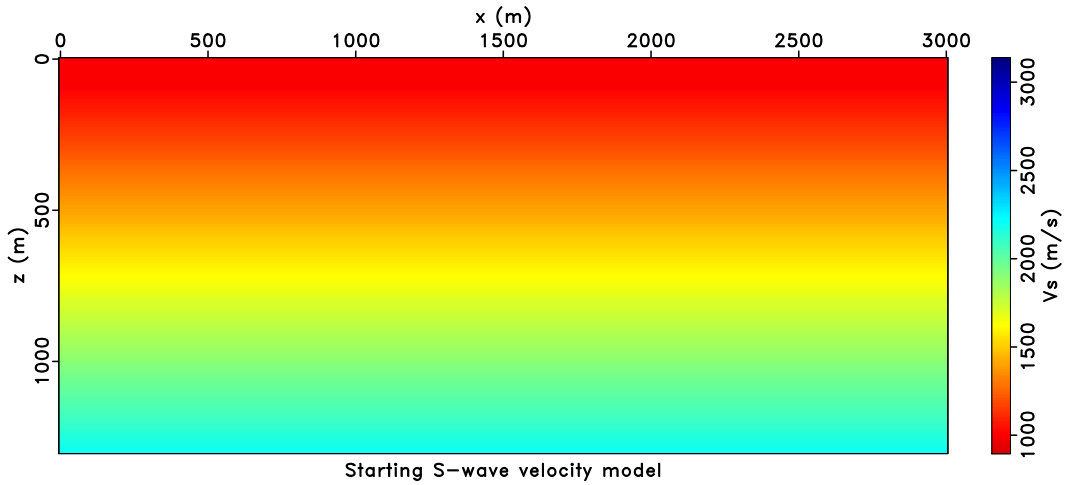


(b)

Figure 5.12: Prestack multicomponent data. (a) Horizontal particle velocity data. (b) Vertical partical velocity data.

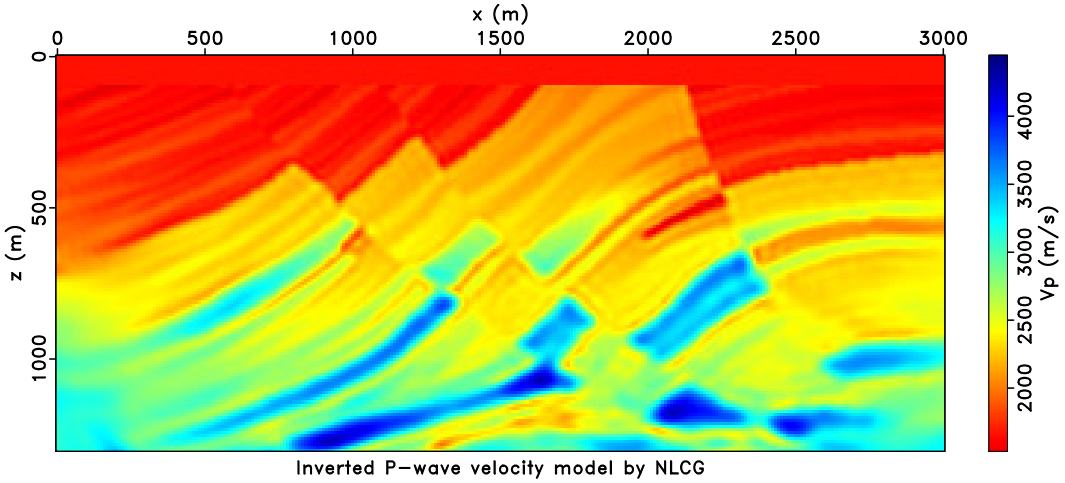


(a)

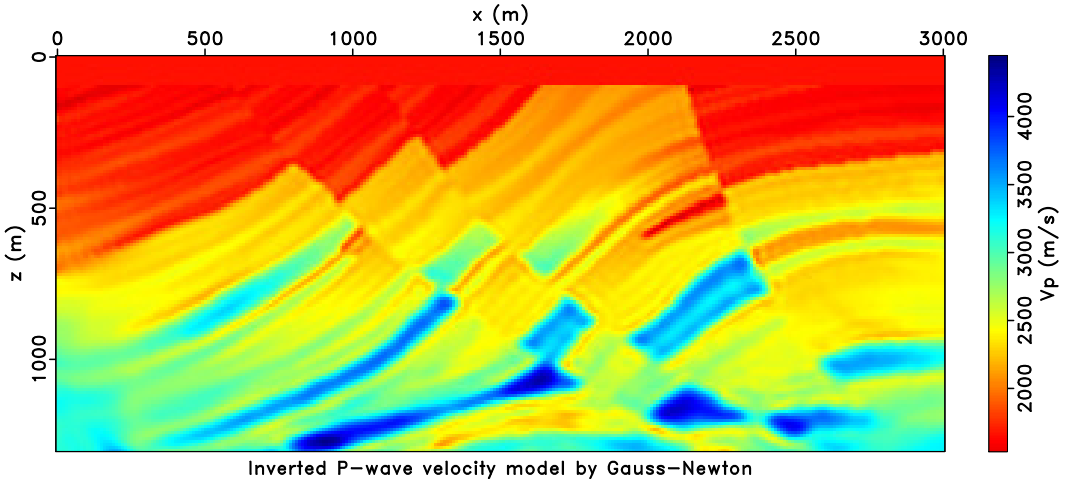


(b)

Figure 5.13: (a) Starting P-wave velocity model. (b) Starting S-wave velocity model.

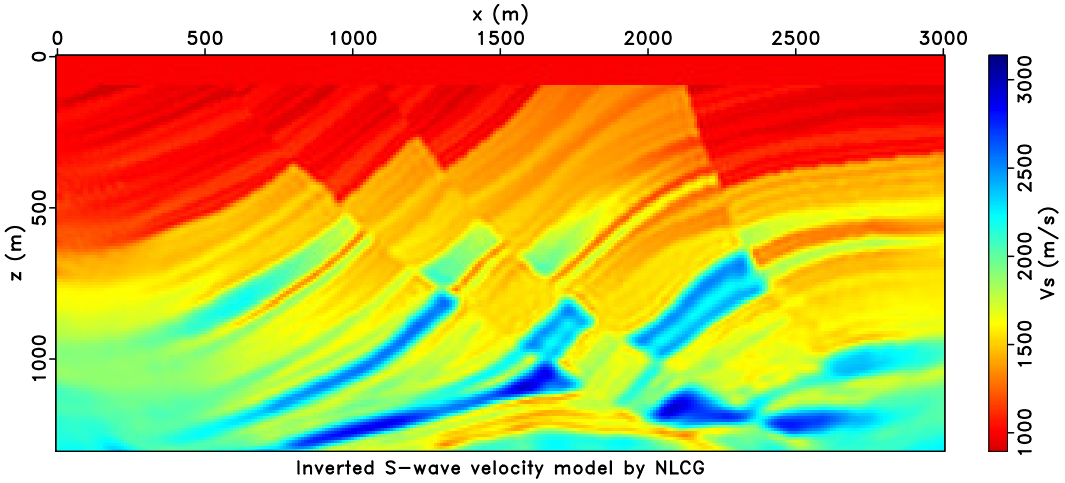


(a)

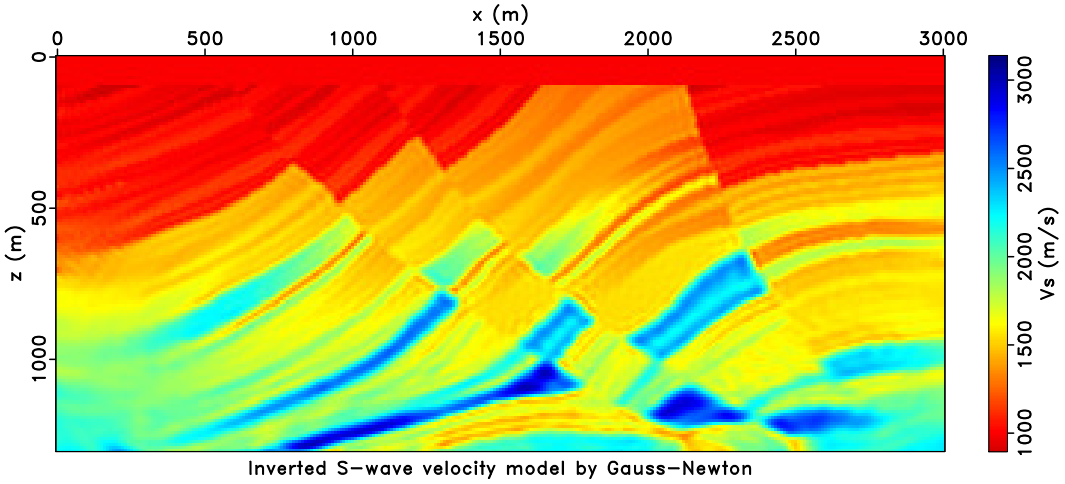


(b)

Figure 5.14: (a) Inverted P-wave velocity model by elastic NLCG FWI. (b) Inverted P-wave velocity model by elastic Gauss-Newton FWI.

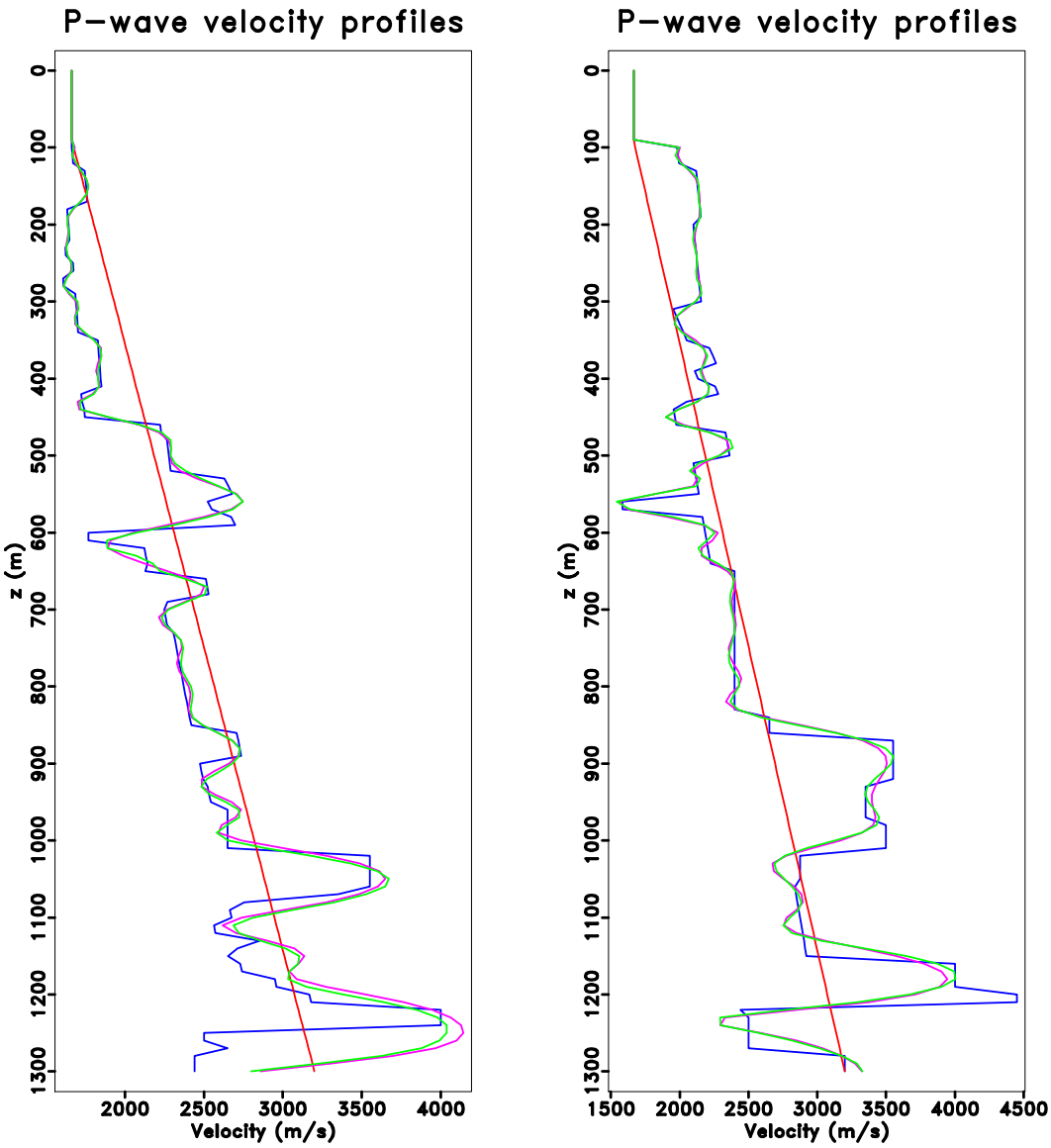


(a)



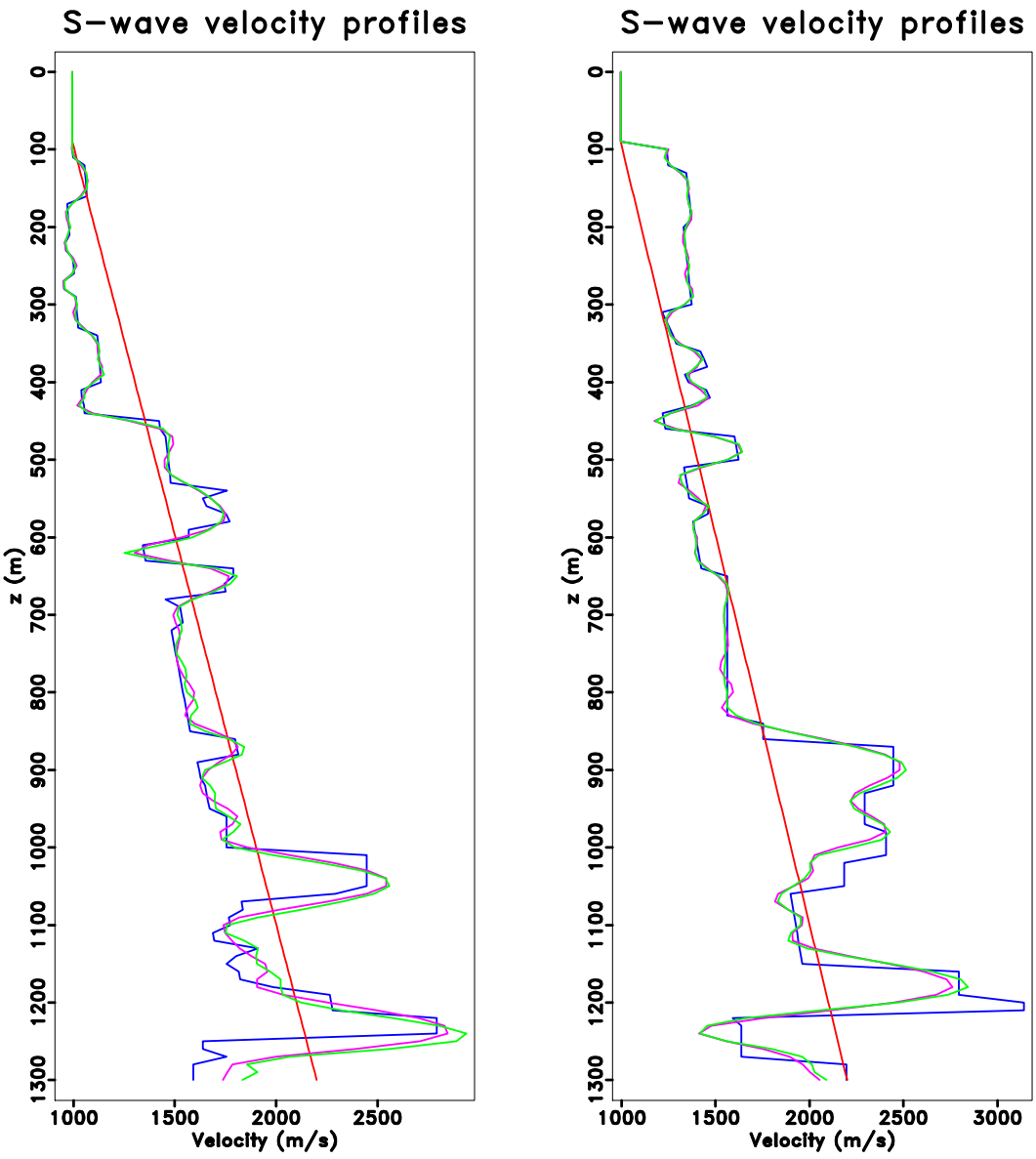
(b)

Figure 5.15: (a) Inverted S-wave velocity model by elastic NLCG FWI. (b) Inverted S-wave velocity model by elastic Gauss-Newton FWI.



(a) (b)

Figure 5.16: Profiles of inverted P-wave velocity models at (a) $x = 1000$ m and (b) $x = 2050$ m. Blue: true model; Red: starting model; Purple: inverted model by elastic NLCG FWI; Green: inverted model by elastic Gauss-Newton FWI.



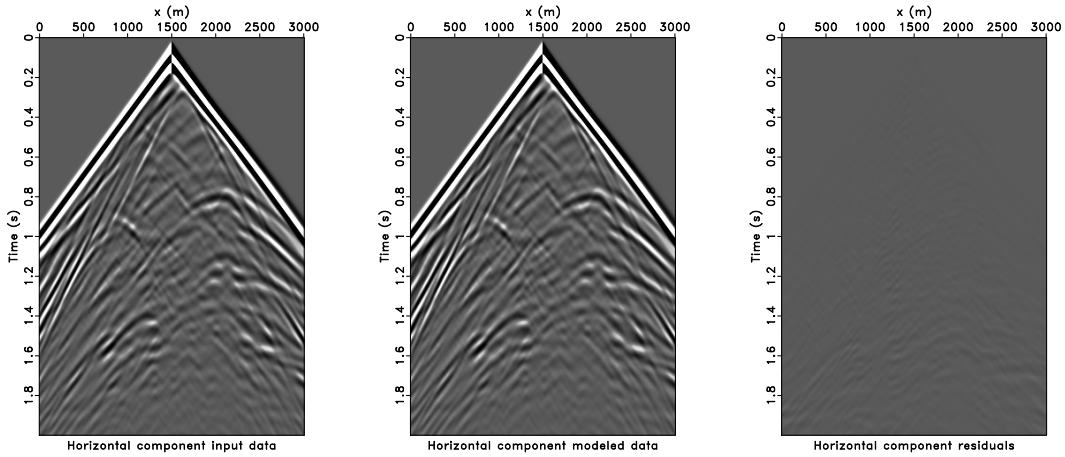
(a) (b)

Figure 5.17: Profiles of inverted S-wave velocity models at (a) $x = 1000$ m and (b) $x = 2050$ m. Blue: true model; Red: starting model; Purple: inverted model by elastic NLCG FWI; Green: inverted model by elastic Gauss-Newton FWI.

et al., 2013). Third, the design of preconditioning operators to accelerate the convergence of the inner CGLS iteration of the elastic Gauss-Newton FWI could lead to important savings in computational time. Research in the area of preconditioning should be prioritized in the future in order to make elastic Gauss-Newton FWI viable for practical applications.

5.5 Conclusions

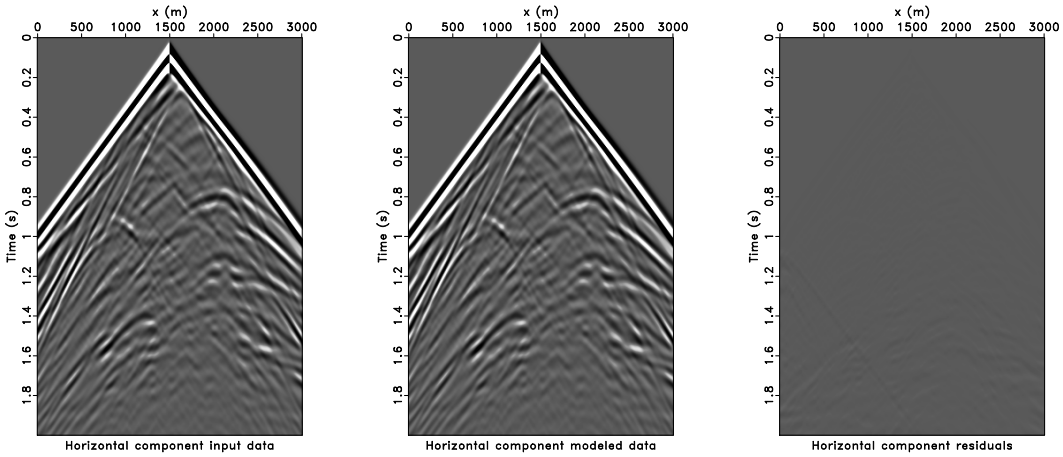
A time-domain matrix-free elastic Gauss-Newton FWI based on the elastic LSRTM algorithm was developed. We formulate the elastic Gauss-Newton FWI as an iterative elastic LSRTM problem. The proposed algorithm consists of two loops of iterations: the outer Gauss-Newton nonlinear iterations and the inner CGLS linear iterations. The outer nonlinear iteration uses a parabolic fitting line search to estimate the step size. The Gauss-Newton search direction in each outer FWI iteration is computed using the matrix-free CGLS algorithm. We point out that this step is equivalent to apply an elastic LSRTM to data residuals with the Fréchet derivative operator as an elastic Born modelling operator and the adjoint of Fréchet derivative operator as the elastic RTM operator. The CGLS algorithm can be safely used for solving the Gauss-Newton search direction because our discretized numerical versions of elastic Born and RTM operators passed the dot-product test. In the inner CGLS linear iterations, the step size is analytically calculated without the need of a line search. The inner CGLS linear iterations are preconditioned using the elastic pseudo-Hessian operator. Our algorithm is matrix-free that only requires the forward Fréchet derivative and adjoint Fréchet derivative operators applied to vectors. The operators are applied on vectors efficiently via the adjoint-state method. We use the proposed algorithm to simultaneously invert for P- and S-wave velocities. The proposed elastic Gauss-Newton FWI generates slightly better-inverted models than the NLCG method based elastic FWI.



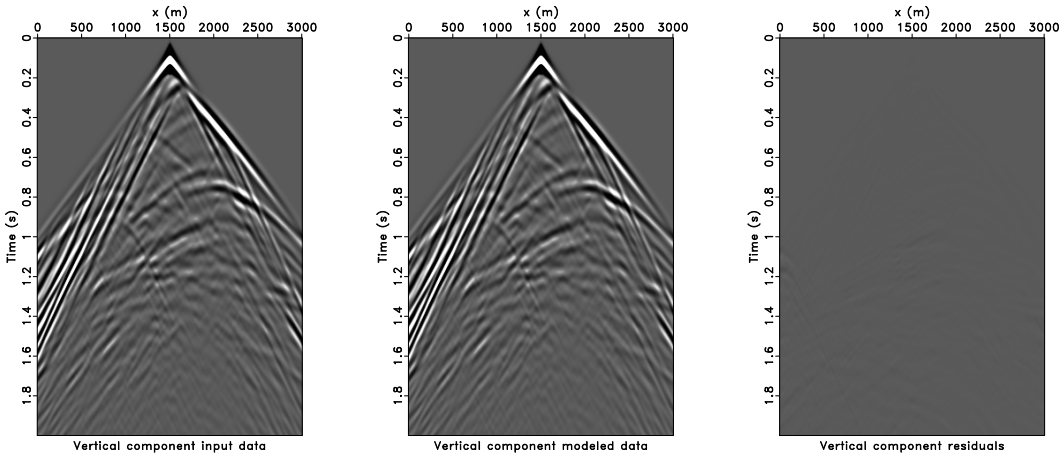
(a) (b) (c)

(d) (e) (f)

Figure 5.18: Data and residuals for elastic NLCG FWI. (a) Observed horizontal component data of shot at $x = 1500$ m. (b) Horizontal component data modeled by elastic NLCG FWI inverted models. (c) Horizontal component data residual. (d) Observed vertical component data of shot at $x = 1500$ m. (e) Vertical component data modelled by elastic NLCG FWI inverted models. (f) Vertical component data residual.



(a) (b) (c)



(d) (e) (f)

Figure 5.19: Data and residuals for elastic Gauss-Newton FWI. (a) Observed horizontal component data of shot at $x = 1500$ m. (b) Horizontal component data modeled by elastic Gauss-Newton FWI inverted models. (c) Horizontal component data residual. (d) Observed vertical component data of shot at $x = 1500$ m. (e) Vertical component data modelled by elastic Gauss-Newton FWI inverted models. (f) Vertical component data residual.

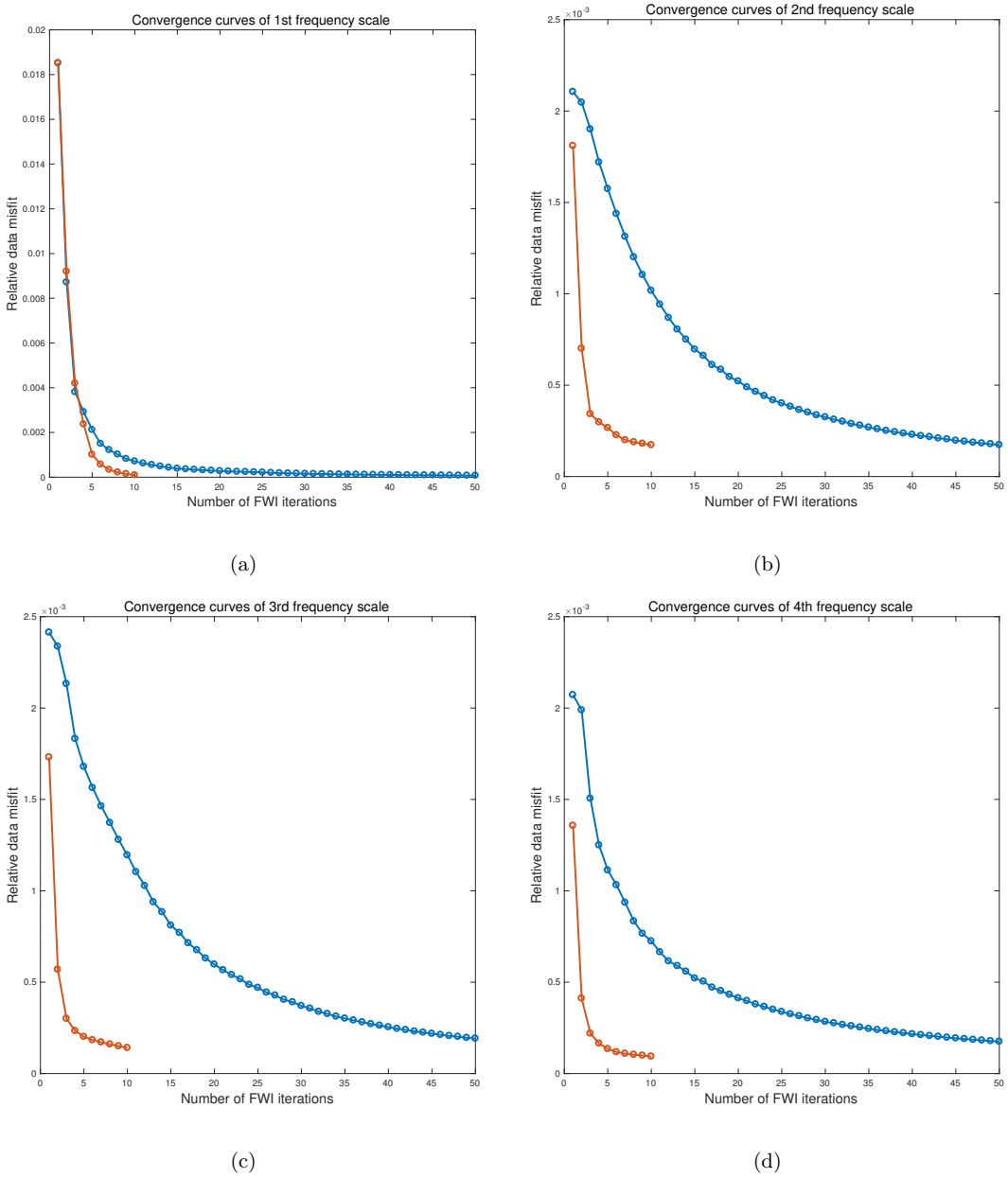


Figure 5.20: Relative data misfit curves for elastic NLCG FWI and elastic Gauss-Newton FWI in frequency band (a) 0 - 2 Hz, (b) 0 - 5 Hz, (c) 0 - 10 Hz, (d) 0 - 30 Hz. Blue: data misfit curves for elastic NLCG FWI. Red: data misfit curves for elastic Gauss-Newton FWI.

CHAPTER 6

Conclusions

With the fast development of computer hardware, imaging and inversion techniques proposed by the exploration geophysics community are moving from simplified methods to more complex methods that honour as far as possible the physics of wave propagation. This thesis deals with imaging techniques based on two-way wave-equation operators and elastic physics. Multiparameter imaging and inversion often suffer from crosstalk artifacts of parameters. In this thesis, I investigate the role of Hessian for decoupling elastic parameters in elastic imaging and inversion. Moreover, I develop a matrix-free elastic Gauss-Newton FWI method based on elastic LSRTM code. It may be an alternative to popular gradient-based elastic FWI methods.

In Chapter 1, I have reviewed the topics of seismic imaging and inversion.

In Chapter 2, I draw connections between migration, least-squares migration and full-waveform inversion.

In Chapter 3, I formulate the elastic Born approximation and elastic reverse time migration (RTM) operators with a first-order velocity-stress isotropic elastic wave-equation system. I incorporate the two operators into a least-squares inversion scheme. I name this new method as elastic least-squares reverse time migration. The forward modelling engine of my proposed method adopts the time-domain staggered-grid finite-difference and the convolutional perfectly matched layers boundary condition. My discretized computer codes of the elastic Born operator and its adjoint (elastic RTM operator) pass the dot-product test. I use the CGLS algorithm to solve the least-squares optimization problem. The Hessian operator for elastic LSRTM is implicitly inverted via a matrix-free algorithm that only requires the action of forward and adjoint operators on vectors. Note that I do not explicitly compute the Hessian or store the Hessian in memory. I observe that the elastic LSRTM provides higher

resolution images with fewer artifacts and a superior balance of amplitudes when compared to elastic RTM. The elastic LSRTM can also mitigate cross-talk between the P-wave and S-wave images.

In Chapter 4, I investigate the influence of including density in elastic LSRTM. In conventional elastic LSRTM algorithms, density is either assumed to be constant or known. In other words, the density perturbation is not part of the least-squares inversion formulation. I find that neglecting density in elastic LSRTM may lead to crosstalk artifacts in the P- and S-wave images. I introduce a time-domain three-parameter elastic LSRTM algorithm to simultaneously invert for density, P- and S-wave images. The proposed three-parameter elastic LSRTM is able to reduce the multiparameter crosstalk among density, P- and S-wave images. It provides an estimation of the density perturbation image. Moreover, I observe that including density image in the elastic LSRTM inversion can improve the convergence of the least-squares inversion. The proposed three-parameter elastic LSRTM does not entirely remove the multiparameter crosstalk from the images. The radiation patterns of the density perturbation and velocity perturbations overlap which indicates that one has little hope in precisely recovering density. The short wavelength of the density model may be resolved only partially from P- and S-wave images. A prior information about the density contrast should be be incorporated to compute reliable multiparameter estimates. One possibility is to adopt a Bayesian approach to integrate information about correlations that might exist between elastic parameters. This approach will entail having access to sufficient information from well log data analysis, rock physics models and geostatistical information to estimate covariance matrices that can help the recovery of the elastic parameters. The Bayesian approach is prevalent in linearized AVO analysis (Ulrych et al., 2001; Buland and Omre, 2003; Alemie and Sacchi, 2011) and has also been used for poststack inversion by many researchers (Bongajum et al., 2013), one could expect similar developments in multiparameter imaging and inversion.

In Chapter 5, I introduce a time-domain matrix-free elastic Gauss-Newton FWI algorithm. I point out that the elastic Gauss-Newton FWI is an iterative elastic LSRTM problem. In this framework, different LSM methods can be used to design a matrix-free Gauss-Newton FWI algorithm. My proposed algorithm consists of two loops of iterations: the outer Gauss-Newton nonlinear iterations and the inner CGLS linear iterations. The Gauss-Newton search direction in each outer FWI iteration is computed using the CGLS method, which is equivalent to apply an elastic LSRTM on data residual. The advantage of my proposed algorithm is that it is matrix-free and only requires the forward and adjoint Fréchet derivative operators applied to vectors. Memory requirements are of the same order of the elastic NLCG FWI method. I use the proposed algorithm to simultaneously invert for P- and S-wave velocities. The proposed elastic Gauss-Newton FWI generates slightly better-inverted models than the elastic NLCG FWI. The elastic Gauss-Newton FWI converges

faster than the elastic NLCG FWI. The proposed matrix-free Gauss-Newton FWI has the potential to be used to suppress crosstalk in multiparameter FWI. The disadvantage of the matrix-free elastic Gauss-Newton FWI is that it is more expensive than the elastic NLCG FWI. The elastic Gauss-Newton FWI algorithm has an additional inner loop for CGLS iteration. For the numerical examples presented in this thesis, the run time of elastic Gauss-Newton FWI is approximately three times of the run time of elastic NLCG. Future research is needed to design algorithms that reduce the computational cost of the matrix-free elastic Gauss-Newton FWI. For example, one can utilize data contraction strategies such as source encoding, plan-wave synthesis or stochastic source subsampling to reduce the computational cost. Another route to reduce computational cost of the time-domain elastic Gauss-Newton FWI is to design an intelligent criterion for terminating the inner CGLS iteration early. Moreover, preconditioning strategies need to be investigated to accelerate the convergence of the inner CGLS loop in the elastic Gauss-Newton FWI algorithm. With the advancement of the development of high-performance computing and the acceleration strategies, the time-domain matrix-free elastic Gauss-Newton FWI will become a practical alternative to conventional gradient-based optimization method. Last, it is important to point out that adopting second-order optimization methods like the Gauss-Newton FWI offers a significant challenge if one wishes to invert for large 3D subsurface models.

Bibliography

Akcelik, V., G. Biros, and O. Ghattas, 2002, Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation: Supercomputing, ACM/IEEE 2002 Conference, 41–41.

Aki, K., and P. G. Richards, 2002, Quantitative seismology: University Science Books.

Albertin, U., P. Shen, A. Sekar, T. Johnsen, C. Wu, K. Nihei, and K. Bube, 2016, 3D orthorhombic elastic full-waveform inversion in the reflection domain from hydrophone data: 86th Annual International Meeting, SEG, Expanded Abstracts, 1094–1098.

Alemie, W., and M. Sacchi, 2016, Joint reparametrized time-lapse full-waveform inversion: 86th Annual International Meeting, SEG, Expanded Abstracts, 1309–1314.

Alemie, W., and M. D. Sacchi, 2011, High-resolution three-term AVO inversion by means of a Trivariate Cauchy probability distribution: Geophysics, **76**, R43–R55.

Anagaw, A. Y., and M. D. Sacchi, 2012, Full waveform inversion with simultaneous sources using the full Newton method: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–5.

———, 2014, Comparison of multifrequency selection strategies for simultaneous-source full-waveform inversion: Geophysics, **79**, R165–R181.

Anikiev, D., B. Kashtan, and W. A. Mulder, 2013, Decoupling of elastic parameters with iterative linearized inversion: 83rd Annual International Meeting, SEG, Expanded Abstracts, 3185–3190.

Backus, G., and F. Gilbert, 1970, Uniqueness in the inversion of inaccurate gross Earth data: Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, **266**, 123–192.

Bamberger, A., G. Chavent, C. Hemon, and P. Lailly, 1982, Inversion of normal incidence seismograms: Geophysics, **47**, 757–770.

Bamberger, A., G. Chavent, and P. Lailly, 1979, About the stability of the inverse problem in 1-D wave equations - application to the interpretation of seismic profiles: Applied Mathematics and Optimization, **5**, 1–47.

Baysal, E., D. D. Kosloff, and J. W. C. Sherwood, 1983, Reverse time migration: Geophysics, **48**, 1514–1524.

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

———, 2011, An efficient frequency-domain full waveform inversion method using simultaneous encoded sources: Geophysics, **76**, R109–R124.

Berkhout, A. J., and D. J. Verschuur, 2005, Removal of internal multiples with the common-focus-point (CFP) approach: Part 1 - explanation of the theory: Geophysics, **70**, V45–V60.

Beydoun, W. B., and M. Mendes, 1989, Elastic ray-Born l2-migration/inversion: Geophysical Journal International, **97**, 151–160.

Beylkin, G., and R. Burridge, 1990, Linearized inverse scattering problems in acoustics and elasticity: Wave Motion, **12**, 15 – 52.

Biondi, B., and A. Almomin, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, WA129.

Bjorck, A., 1996, Numerical methods for least squares problems: Society for Industrial and Applied Mathematics.

Bongajum, E., J. Boisvert, and M. Sacchi, 2013, Bayesian linearized seismic inversion with locally varying spatial anisotropy: Journal of Applied Geophysics, **88**, 31 – 41.

Borisov, D., R. Modrak, F. Gao, and J. Tromp, 2018, 3D elastic full-waveform inversion of surface waves in the presence of irregular topography using an envelope-based misfit function: Geophysics, **83**, R1–R11.

Borisov, D., and S. C. Singh, 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.

Borzi, A., and V. Schulz, 2012, Computational optimization of systems governed by partial differential equations: Society for Industrial and Applied Mathematics. SIAM, Philadelphia, PA.

Bourgeois, A., B. F. Jiang, and P. Lailly, 1989, Linearized inversion: a significant step beyond pre-stack migration: Geophysical Journal International, **99**, 435–445.

Bozdag, E., J. Trampert, and J. Tromp, 2011, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: Geophysical Journal International, **185**, 845–870.

Brenders, A. J., and R. G. Pratt, 2007, Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model: Geophysical Journal International, **168**, 133–151.

Brossier, R., S. Operto, and J. Virieux, 2009a, Robust elastic frequency-domain full-waveform inversion using the L1 norm: Geophysical Research Letters, **36**.

———, 2009b, Seismic imaging of complex onshore structures by 2D elastic frequency-domain

full-waveform inversion: Geophysics, **74**, WCC105–WCC118.

———, 2010, Which data residual norm for robust elastic frequency-domain full waveform inversion?: Geophysics, **75**, R37–R46.

———, 2015, Velocity model building from seismic reflection data by full-waveform inversion: Geophysical Prospecting, **63**, 354–367.

Buland, A., and H. Omre, 2003, Bayesian linearized AVO inversion: Geophysics, **68**, 185–198.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Burstedde, C., and O. Ghattas, 2009, Algorithmic strategies for full waveform inversion: 1D experiments: Geophysics, **74**, WCC37–WCC46.

Chang, W., and G. A. McMechan, 1986, Reverse-time migration of offset vertical seismic profiling data using the excitation-time imaging condition: Geophysics, **51**, 67–84.

———, 1987, Elastic reverse-time migration: Geophysics, **52**, 1365–1375.

Chavent, G., F. Clement, and S. Gomez, 1994, Automatic determination of velocities via migration-based traveltime waveform inversion: A synthetic data example: 1179–1182.

Chen, K., and M. D. Sacchi, 2015, Robust reduced-rank filtering for erratic seismic noise attenuation: Geophysics, **80**, V1–V11.

———, 2016, Elastic least-squares reverse time migration via linearized elastic full waveform inversion with pseudo-Hessian preconditioning: Signal Analysis and Imaging Group, Report 17, University of Alberta, 274–310.

———, 2017a, Elastic least-squares reverse time migration via linearized elastic full-waveform inversion with pseudo-Hessian preconditioning: Geophysics, **82**, S341–S358.

———, 2017b, Robust f-x projection filtering for simultaneous random and erratic seismic noise attenuation: Geophysical Prospecting, **65**, 650–668.

———, 2017c, Time-domain elastic Gauss-Newton full-waveform inversion via matrix-free adjoint-state method: Signal Analysis and Imaging Group, Report 18, University of Alberta, 166–204.

———, 2018, The importance of including density in elastic least-squares reverse time migration: multiparameter crosstalk and convergence: Geophysical Journal International, *under revision*, **N**, P1–P2.

Chen, P., and E.-J. Lee, 2015, Full-3D Seismic Waveform Inversion: Springer Geophysics.

Chen, Y., H. Chen, K. Xiang, and X. Chen, 2017, Preserving the discontinuities in least-squares reverse time migration of simultaneous-source data: Geophysics, **82**, S185–S196.

Cheng, J., N. Kazemi, and M. Sacchi, 2016, Least-squares migration via a gradient projection method - application to seismic data deblending: 78th Annual International Conference and Exhibition, EAGE, Extended Abstracts.

Choi, Y., D.-J. Min, and C. Shin, 2008a, Frequency-domain elastic full waveform inversion

using the new pseudo-Hessian matrix: Experience of elastic Marmousi-2 synthetic data: Bulletin of the Seismological Society of America, **98**, 2402.

———, 2008b, Two-dimensional waveform inversion of multi-component data in acoustic-elastic coupled media: Geophysical Prospecting, **56**, 863–881.

Claerbout, J. F., 1970, Coarse grid calculations of waves in inhomogeneous media with application to delineation of complicated seismic structure: Geophysics, **35**, 407–418.

———, 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481.

———, 1985, Imaging the earth's interior: Blackwell Scientific Publications, Inc.

———, 1992, Earth soundings analysis: Processing versus inversion: Blackwell Scientific Publications Cambridge, Massachusetts, USA, **6**.

Claerbout, J. F., and F. Muir, 1973, Robust modelling with erratic data: Geophysics, **38**, 826–844.

Clement, F., G. Chavent, and S. Gomez, 2001, Migration-based traveltime waveform inversion of 2-D simple structures: A synthetic example: Geophysics, **66**, 845–860.

Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic nonlinear waveform inversion: Application to real data: Geophysics, **55**, 527–538.

Crase, E., C. Wideman, M. Noble, and A. Tarantola, 1992, Nonlinear elastic waveform inversion of land seismic reflection data: Journal of Geophysical Research: Solid Earth, **97**, 4685–4703.

Dahlen, F. A., S.-H. Hung, and G. Nolet, 2000, Fréchet kernels for finite-frequency traveltimes - I. theory: Geophysical Journal International, **141**, 157–174.

Dai, W., P. Fowler, and G. T. Schuster, 2012, Multi-source least-squares reverse time migration: Geophysical Prospecting, **60**, 681–695.

Dai, W., and G. T. Schuster, 2013, Plane-wave least-squares reverse-time migration: Geophysics, **78**, S165–S177.

Dellinger, J., and J. Etgen, 1990, Wavefield separation in two-dimensional anisotropic media: Geophysics, **55**, 914–919.

Dong, S., J. Cai, M. Guo, S. Suh, Z. Zhang, B. Wang, and Z. Li, 2012, Least-squares reverse time migration: towards true amplitude imaging and improving the resolution: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–5.

Douma, H., D. Yingst, I. Vasconcelos, and J. Tromp, 2010, On the connection between artifact filtering in reverse-time migration and adjoint tomography: Geophysics, **75**, S219–S223.

Du, Q., Y. Zhu, and J. Ba, 2012, Polarity reversal correction for elastic reverse time migration: Geophysics, **77**, S31–S41.

Duan, L., L. Peng, and Y. Zhang, 2014, Band-limited impedance perturbation inversion using cross-correlative least-squares RTM: 84th Annual International Meeting, SEG, Expanded Abstracts, 3720–3725.

Duan, Y., A. Guitton, and P. Sava, 2017, Elastic least-squares reverse time migration: Geophysics, **82**, S315–S325.

Duan, Y., and P. Sava, 2015, Scalar imaging condition for elastic reverse time migration: Geophysics, **80**, S127–S136.

Dussaud, E., W. W. Symes, P. Williamson, L. Lemaistre, P. Singer, B. Denel, and A. Cherrett, 2008, Computational strategies for reverse-time migration: 78th Annual International Meeting, SEG, Expanded Abstracts, 2267–2271.

Epanomeritakis, I., V. Akcelik, O. Ghattas, and J. Bielak, 2008, A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion: Inverse Problems, **24**, 034015.

Etgen, J., S. H. Gray, and Y. Zhang, 2009, An overview of depth imaging in exploration geophysics: Geophysics, **74**, WCA5–WCA17.

Etgen, J. T., 1986, Prestack reverse time migration of shot profiles: SEP report, **50**, 151–169.

Feng, Z., and G. T. Schuster, 2017, Elastic least-squares reverse time migration: Geophysics, **82**, S143–S157.

Fichtner, A., 2010, Full Seismic Waveform Modelling and Inversion: Springer, Heidelberg.

Fichtner, A., B. L. N. Kennett, H. Igel, and H.-P. Bunge, 2008, Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain: Geophysical Journal International, **175**, 665–685.

Fichtner, A., and J. Trampert, 2011, Hessian kernels of seismic data functionals based upon adjoint techniques: Geophysical Journal International, **185**, 775–798.

Fletcher, R. P., D. Nichols, R. Bloor, and R. T. Coates, 2016, Least-squares migration data domain versus image domain using point spread functions: The Leading Edge, **35**, 157–162.

Forgues, E., and G. Lambare, 1997, Parameterization study for acoustic and elastic ray + Born inversion: Journal of Seismic Exploration, **6**, 253–277.

French, W. S., 1975, Computer migration of oblique seismic reflection profiels: Geophysics, **40**, 961–980.

Fu, L., and W. W. Symes, 2017, An adaptive multiscale algorithm for efficient extended waveform inversion: Geophysics, **82**, R183.

Gauthier, O., J. Virieux, and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results: Geophysics, **51**, 1387–1403.

Gelis, C., J. Virieux, and G. Grandjean, 2007, Two-dimensional elastic full waveform inversion using Born and Rytov formulations in the frequency domain: Geophysical Journal International, **168**, 605–633.

Gu, B., Z. Li, P. Yang, W. Xu, and J. Han, 2017, Elastic least-squares reverse time migration with hybrid l1/l2 misfit function: Geophysics, **82**, S271–S291.

Guasch, L., M. Warner, T. Nangoo, J. Morgan, A. Umpleby, I. Stekl, and N. Shah, 2012, Elastic 3D full-waveform inversion: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–5.

Guitton, A., 2004, Amplitude and kinematic corrections of migrated images for nonunitary imaging operators: Geophysics, **69**, 1017–1024.

Guo, P., and G. A. McMechan, 2018, Compensating Q effects in viscoelastic media by adjoint-based least-squares reverse time migration: Geophysics, **83**, S151–S172.

Hansen, P., 1998, Rank-deficient and discrete ill-posed problems: Society for Industrial and Applied Mathematics.

Hardage, B., M. DeAngelo, P. Murray, and D. Sava, 2011, Multicomponent seismic technology: Society of Exploration Geophysicists.

Hestenes, M. R., and E. Stiefel, 1952, Methods of conjugate gradients for solving linear systems: Journal of Research of the National Bureau of Standards, **49**, 409–436.

Hou, J., and W. W. Symes, 2016, Accelerating extended least-squares migration with weighted conjugate gradient iteration: Geophysics, **81**, S165–S179.

Huang, Y., R. Nammour, and W. Symes, 2016, Flexibly preconditioned extended least-squares migration in shot-record domain: Geophysics, **81**, S299–S315.

Huber, P., 1981, Robust statistics: John Wiley & Sons, Inc.

Jackson, D. D., 1979, The use of a priori data to resolve non-uniqueness in linear inversion: Geophysical Journal of the Royal Astronomical Society, **57**, 137–157.

Jeong, W., H.-Y. Lee, and D.-J. Min, 2012, Full waveform inversion strategy for density in the frequency domain: Geophysical Journal International, **188**, 1221–1242.

Ji, J., 2009, An exact adjoint operation pair in time extrapolation and its application in least-squares reverse-time migration: Geophysics, **74**, H27–H33.

Jin, S., R. Madariaga, J. Virieux, and G. Lambar, 1992, Two-dimensional asymptotic iterative elastic inversion: Geophysical Journal International, **108**, 575–588.

Kamei, R., and D. Lumley, 2017, Full waveform inversion of repeating seismic events to estimate time-lapse velocity changes: Geophysical Journal International, **209**, 1239–1264.

Kaplan, S. T., P. S. Routh, and M. D. Sacchi, 2010, Derivation of forward and adjoint operators for least-squares shot-profile split-step migration: Geophysics, **75**, S225–S235.

Kazemi, N., and M. D. Sacchi, 2015, Block row recursive least-squares migration: Geophysics, **80**, A95–A101.

Keys, R. G., and A. B. Weglein, 1983, Generalized linear inversion and the first Born theory for acoustic media: Journal of Mathematical Physics, **24**, 1444–1449.

Kohn, D., D. De Nil, A. Kurzmann, A. Przebindowska, and T. Bohlen, 2012, On the influence of model parametrization in elastic full waveform tomography: Geophysical Journal International, **191**, 325–345.

Kolb, P., F. Collino, and P. Lailly, 1986, Pre-stack inversion of a 1-D medium: Proceedings

of the IEEE, **74**, 498–508.

Komatitsch, D., and R. Martin, 2007, An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation: Geophysics, **72**, SM155–SM167.

Kotsi, M., and A. Malcolm, 2017, A statistical comparison of three 4D full-waveform inversion schemes: 87th Annual International Meeting, SEG, Expanded Abstracts, 1434–1438.

Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysics, **74**, WCC177–WCC188.

Kuehl, H., and M. D. Sacchi, 2003, Least-squares wave-equation migration for AVP/AVA inversion: Geophysics, **68**, 262–273.

Kwon, T., S. J. Seol, and J. Byun, 2015, Efficient full-waveform inversion with normalized plane-wave data: Geophysical Journal International, **201**, 53–60.

Lailly, P., 1983, The seismic inverse problem as a sequence of before-stack migrations: Conference on Inverse Scattering: Theory and Applications. SIAM, Philadelphia, 206–220.

Lambare, G., J. Virieux, R. Madariaga, and S. Jin, 1992, Iterative asymptotic inversion in the acoustic approximation: Geophysics, **57**, 1138–1154.

Le, H., 2016, Anisotropic full-waveform inversion using the second-order pseudo-acoustic wave equations: 86th Annual International Meeting, SEG, Expanded Abstracts, 408–412.

Levander, A. R., 1988, Fourth-order finite-difference P-SV seismograms: Geophysics, **53**, 1425–1436.

Lions, J. L., 1971, Optimal control of systems governed by partial differential equations: Springer Berlin Heidelberg.

Liu, Y., and M. K. Sen, 2009, An implicit staggered-grid finite-difference method for seismic modelling: Geophysical Journal International, **179**, 459–474.

Loewenthal, D., and I. R. Mufti, 1983, Reversed time migration in spatial frequency domain: Geophysics, **48**, 627–635.

Lu, R., P. Traynin, and J. E. Anderson, 2009, Comparison of elastic and acoustic reverse-time migration on the synthetic elastic Marmousi-II OBC dataset: 79th Annual International Meeting, SEG, Expanded Abstracts, 2799–2803.

Lu, S., X. Li, A. Valenciano, N. Chemingui, and C. Cheng, 2017, Broadband least-squares wave-equation migration: 87th Annual International Meeting, SEG, Expanded Abstracts, 4422–4426.

Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: 81st Annual International Meeting, SEG, Expanded Abstracts, 2788–2792.

Luo, Y., Y. Ma, Y. Wu, H. Liu, and L. Cao, 2016, Full-traveltime inversion: Geophysics, **81**, R261–R274.

Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**,

645–653.

Luo, Y., H. Zhu, T. Nissen-Meyer, C. Morency, and J. Tromp, 2009, Seismic modeling and imaging based upon spectral-element and adjoint methods: The Leading Edge, **28**, 568–574.

Martin, G. S., R. Wiley, and K. J. Marfurt, 2006, Marmousi2: An elastic upgrade for Marmousi: The Leading Edge, **25**, 156–166.

Matharu, G., and M. Sacchi, 2017, Feasibility testing of simultaneous source elastic full-waveform inversion: 87th Annual International Meeting, SEG, Expanded Abstracts, 1573–1577.

McMechan, G. A., 1983, Migration by extrapolation of time-dependent boundary values: Geophysical Prospecting, **31**, 413–420.

Metivier, L., F. Bretaudeau, R. Brossier, S. Operto, and J. Virieux, 2014, Full waveform inversion and the truncated newton method: quantitative imaging of complex subsurface structures: Geophysical Prospecting, **62**, 1353–1375.

Metivier, L., R. Brossier, Q. Merigot, E. Oudet, and J. Virieux, 2016, Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion: Geophysical Journal International, **205**, 345–377.

Metivier, L., R. Brossier, J. Virieux, and S. Operto, 2013, Full waveform inversion and the truncated Newton method: SIAM Journal on Scientific Computing, **35**, B401–B437.

Mora, P., 1987a, Elastic wavefield inversion: Ph.D. thesis, Stanford University.

———, 1987b, Nonlinear two-dimensional elastic inversion of multioffset seismic data: Geophysics, **52**, 1211–1228.

———, 1988, Elastic wave-field inversion of reflection and transmission data: Geophysics, **53**, 750–759.

———, 1989, Inversion = migration + tomography: Geophysics, **54**, 1575–1586.

Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221.

Nocedal, J., and S. J. Wright, 2006, Numerical Optimization: Springer.

Oh, J.-W., M. Kalita, and T. Alkhalifah, 2018, 3D elastic full-waveform inversion using P-wave excitation amplitude: Application to ocean bottom cable field data: Geophysics, **83**, R129–R140.

Operto, S., Y. Gholami, V. Prieux, A. Ribodetti, R. Brossier, L. Metivier, and J. Virieux, 2013, A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice: The Leading Edge, **32**, 1040–1054.

Operto, S., A. Miniussi, R. Brossier, L. Combe, L. Metivier, V. Monteiller, A. Ribodetti, and J. Virieux, 2015, Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to valhall in the visco-acoustic vertical transverse isotropic approximation: Geophysical Journal International, **202**, 1362–1391.

Operto, S., J. Virieux, J.-X. Dessa, and G. Pascal, 2006, Crustal seismic imaging from multifold ocean bottom seismometer data by frequency domain full waveform tomography: Application to the eastern Nankai trough: Journal of Geophysical Research: Solid Earth, **111**.

Ostmo, S., W. A. Mulder, and R.-. Plessix, 2002, Finite-difference iterative migration by linearized waveform inversion in the frequency domain: 72nd Annual International Meeting, SEG, Expanded Abstracts, 1384–1387.

Pageot, D., S. Operto, M. Valle, R. Brossier, and J. Virieux, 2013, A parametric analysis of two-dimensional elastic full waveform inversion of teleseismic data for lithospheric imaging: Geophysical Journal International.

Paige, C. C., and M. A. Saunders, 1982, LSQR: An algorithm for sparse linear equations and sparse least squares: ACM Trans. Math. Softw., **8**, 43–71.

Pan, W., K. A. Innanen, G. F. Margrave, M. C. Fehler, X. Fang, and J. Li, 2016, Estimation of elastic constants for HTI media using Gauss-Newton and full-Newton multiparameter full-waveform inversion: Geophysics, **81**, R275–R291.

Pica, A., J. P. Diet, and A. Tarantola, 1990, Nonlinear inversion of seismic reflection data in a laterally invariant medium: Geophysics, **55**, 284–292.

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

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model: Geophysics, **64**, 888–901.

Pratt, R. G., C. Shin, and G. J. Hick, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, 341–362.

Pratt, R. G., and R. M. Shipp, 1999, Seismic waveform inversion in the frequency domain, Part 2: Fault delineation in sediments using crosshole data: Geophysics, **64**, 902–914.

Pratt, R. G., Z.-M. Song, P. Williamson, and M. Warner, 1996, Two-dimensional velocity models from wide-angle seismic data by wavefield inversion: Geophysical Journal International, **124**, 323–340.

Qu, Y., J. Li, J. Huang, and Z. Li, 2018, Elastic least-squares reverse time migration with velocities and density perturbation: Geophysical Journal International, **212**, 1033–1056.

Raknes, E. B., and B. Arntsen, 2014, Time-lapse full-waveform inversion of limited-offset seismic data using a local migration regularization: Geophysics, **79**, WA117–WA128.

Ravaut, C., S. Operto, L. Improta, J. Virieux, A. Herrero, and P. Dell'Aversana, 2004, Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: application to a thrust belt: Geophysical Journal International, **159**, 1032–1056.

Ren, Z., Y. Liu, and M. K. Sen, 2017, Least-squares reverse time migration in elastic media: Geophysical Journal International, **208**, 1103–1125.

Rickett, J. E., 2003, Illumination-based normalization for wave-equation depth migration: Geophysics, **68**, 1371–1379.

Romdhane, A., G. Grandjean, R. Brossier, F. Rejiba, S. Operto, and J. Virieux, 2011, Shallow-structure characterization by 2D elastic full-waveform inversion: Geophysics, **76**, R81–R93.

Routh, P., G. Palacharla, I. Chikichev, and S. Lazaratos, 2012, Full wavefield inversion of time-lapse data for improved imaging and reservoir characterization: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–6.

Santosa, F., and W. W. Symes, 1988, Computation of the Hessian for least-squares solutions of inverse problems of reflection seismology: Inverse Problems, **4**, 211.

Schiemenz, A., and H. Igel, 2013, Accelerated 3-D full-waveform inversion using simultaneously encoded sources in the time domain: application to valhall ocean-bottom cable data: Geophysical Journal International, **195**, 1970–1988.

Schuster, G. T., 1993, Least-squares cross-well migration: 110–113.

Sears, T., S. Singh, and P. Barton, 2008, Elastic full waveform inversion of multi-component OBC seismic data: Geophysical Prospecting, **56**, 843–862.

Sears, T. J., P. J. Barton, and S. C. Singh, 2010, Elastic full waveform inversion of multicomponent ocean-bottom cable seismic data: Application to Alba field, U. K. North sea: Geophysics, **75**, R109–R119.

Sheen, D.-H., K. Tuncay, C.-E. Baag, and P. J. Ortoleva, 2006, Time domain Gauss-Newton seismic waveform inversion in elastic media: Geophysical Journal International, **167**, 1373–1384.

Shin, C., S. Jang, and D.-J. Min, 2001a, Improved amplitude preservation for prestack depth migration by inverse scattering theory: Geophysical Prospecting, **49**, 592–606.

Shin, C., K. Yoon, K. J. Marfurt, K. Park, D. Yang, H. Y. Lim, S. Chung, and S. Shin, 2001b, Efficient calculation of a partial-derivative wavefield using reciprocity for seismic imaging and inversion: Geophysics, **66**, 1856–1863.

Singh, S., T. Sears, M. Roberts, A. Gosselet, G. Royle, and P. Baton, 2008, Full elastic waveform inversion: Future of quantitative seismic imaging: 78th Annual International Meeting, SEG, Expanded Abstracts, 1905–1909.

Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.

Song, Z., P. R. Williamson, and R. G. Pratt, 1995, Frequency-domain acoustic-wave modeling and inversion of crosshole data: Part II - inversion method, synthetic experiments and real-data results: Geophysics, **60**, 796–809.

Stanton, A., and M. Sacchi, 2015, Least squares wave equation migration of elastic data:

77th Annual International Conference and Exhibition, EAGE, Extended Abstracts.

Stanton, A., and M. D. Sacchi, 2017, Elastic least-squares one-way wave-equation migration: Geophysics, **82**, S293–S305.

Sun, M., L. Dong, Y. Liu, and J. Yang, 2017, The least-squares elastic reverse-time migration with density variation: 79th Annual International Conference and Exhibition, EAGE, Extended Abstracts.

Sun, R., and G. A. McMechan, 1986, Pre-stack reverse-time migration for elastic waves with application to synthetic offset vertical seismic profiles: Proceedings of the IEEE, **74**, 457–465.

———, 2001, Scalar reverse-time depth migration of prestack elastic seismic data: Geophysics, **66**, 1519–1527.

Sun, R., G. A. McMechan, H.-. Hsiao, and J. Chow, 2004, Separating P- and S-waves in prestack 3D elastic seismograms using divergence and curl: Geophysics, **69**, 286–297.

Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.

Tang, Y., 2009, Target-oriented wave-equation least-squares migration/inversion with phase-encoded Hessian: Geophysics, **74**, WCA95–WCA107.

Tarantola, A., 1984a, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

———, 1984b, Linearized inversion of seismic reflection data: Geophysical Prospecting, **32**, 998–1015.

———, 1986, A strategy for nonlinear elastic inversion of seismic reflection data: Geophysics, **51**, 1893–1903.

———, 1988, Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation: Scattering and Attenuations of Seismic Waves, Part I, 365–399.

———, 2005, Inverse problem theory and methods for model parameter estimation: Society for Industrial and Applied Mathematics.

Tarantola, A., and B. Valette, 1982, Generalized nonlinear inverse problems solved using the least squares criterion: Reviews of Geophysics, **20**, 219–232.

Trad, D., 2017, Challenges for least-squares Kirchhoff depth migration: 87th Annual International Meeting, SEG, Expanded Abstracts, 4545–4549.

Tromp, J., C. Tape, and Q. Liu, 2005, Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels: Geophysical Journal International, **160**, 195–216.

Tura, M. A. C., and L. R. Johnson, 1993, A stable method for linearized inversion of elastic parameters: Geophysical Journal International, **115**, 1–13.

Ulrych, T. J., M. D. Sacchi, and A. Woodbury, 2001, A Bayes tour of inversion: A tutorial: Geophysics, **66**, 55–69.

van Leeuwen, T., and F. J. Herrmann, 2013, Fast waveform inversion without source-

encoding: Geophysical Prospecting, **61**, 10–19.

Verschuur, D. J., A. J. Berkhout, and C. P. A. Wapenaar, 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, 1166–1177.

Vigh, D., K. Jiao, D. Watts, and D. Sun, 2014, Elastic full-waveform inversion application using multicomponent measurements of seismic data collection: Geophysics, **79**, R63–R77.

Vigh, D., and E. W. Starr, 2008, 3D prestack plane-wave, full-waveform inversion: Geophysics, **73**, VE135–VE144.

Vigh, D., E. W. Starr, and J. Kapoor, 2009, Developing earth models with full waveform inversion: The Leading Edge, **28**, 432–435.

Virieux, J., 1986, P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **51**, 889–901.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.

Wang, J., H. Kuehl, and M. D. Sacchi, 2005, High-resolution wave-equation AVA imaging: Algorithm and tests with a data set from the Western Canadian Sedimentary Basin: Geophysics, **70**, S91–S99.

Wang, J., and M. D. Sacchi, 2007, High-resolution wave-equation amplitude-variation-with-ray-parameter (AVP) imaging with sparseness constraints: Geophysics, **72**, S11–S18.

Wang, Y., L. Dong, Y. Liu, and J. Yang, 2016, 2D frequency-domain elastic full-waveform inversion using the block-diagonal pseudo-Hessian approximation: Geophysics, **81**, R247–R259.

Warner, M., and L. Guasch, 2016, Adaptive waveform inversion: Theory: Geophysics, **81**, R429–R445.

Weglein, A. B., 1999, Multiple attenuation: an overview of recent advances and the road ahead (1999): The Leading Edge, **18**, 40–44.

Whitmore, N. D., 1983, Iterative depth migration by backward time propagation: 53rd Annual International Meeting, SEG, Expanded Abstracts, 382–385.

Wong, M., B. L. Biondi, and S. Ronen, 2015, Imaging with primaries and free-surface multiples by joint least-squares reverse time migration: Geophysics, **80**, S223–S235.

Wu, R., and K. Aki, 1985, Scattering characteristics of elastic waves by an elastic heterogeneity: Geophysics, **50**, 582–595.

Wu, R.-S., J. Luo, and B. Wu, 2014, Seismic envelope inversion and modulation signal model: Geophysics, **79**, WA13–WA24.

Xu, L., and M. D. Sacchi, 2018, Preconditioned acoustic least-squares two-way wave-equation migration with exact adjoint operator: Geophysics, **83**, S1–S13.

Xu, L., A. Stanton, and M. Sacchi, 2016, Elastic least-squares reverse time migration: 86th Annual International Meeting, SEG, Expanded Abstracts, 2289–2293.

Xu, S., D. Wang, F. Chen, G. Lambare, and Y. Zhang, 2012, Inversion on reflected seismic wave: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–7.

Xue, Z., Y. Chen, S. Fomel, and J. Sun, 2016, Seismic imaging of incomplete data and simultaneous-source data using least-squares reverse time migration with shaping regularization: Geophysics, **81**, S11–S20.

Yan, J., and P. Sava, 2008, Isotropic angle-domain elastic reverse-time migration: Geophysics, **73**, S229–S239.

Yang, D., F. Liu, S. Morton, A. Malcolm, and M. Fehler, 2016a, Time-lapse full-waveform inversion with ocean-bottom-cable data: Application on Valhall field: Geophysics, **81**, R225–R235.

Yang, J., Y. Liu, and L. Dong, 2016b, Least-squares reverse time migration in the presence of density variations: Geophysics, **81**, S497–S509.

Yang, Y., B. Engquist, J. Sun, and B. F. Hamfeldt, 2018, Application of optimal transport and the quadratic wasserstein metric to full-waveform inversion: Geophysics, **83**, R43–R62.

Yao, G., and H. Jakubowicz, 2016, Least-squares reverse-time migration in a matrix-based formulation: Geophysical Prospecting, **64**, 611–621.

Youn, O. K., and H. Zhou, 2001, Depth imaging with multiples: Geophysics, **66**, 246–255.

Zhang, D., and G. T. Schuster, 2014, Least-squares reverse time migration of multiples: Geophysics, **79**, S11–S21.

Zhang, Y., L. Duan, and Y. Xie, 2015, A stable and practical implementation of least-squares reverse time migration: Geophysics, **80**, V23–V31.

Zhou, W., R. Brossier, S. Operto, and J. Virieux, 2015, Full waveform inversion of diving and reflected waves for velocity model building with impedance inversion based on scale separation: Geophysical Journal International, **202**, 1535–1554.

Zhu, H., Y. Luo, T. Nissen-Meyer, C. Morency, and J. Tromp, 2009, Elastic imaging and time-lapse migration based on adjoint methods: Geophysics, **74**, WCA167–WCA177.

APPENDIX A

Software developed for this thesis: Elastic least-squares reverse time migration (ELSRTM)

My software for the proposed elastic LSRTM method is written in C programming language and parallelized using the message passing interface (MPI). This section provide a brief documentation of the main functions used to develop this thesis.

- `elsrtm.c`: Main function. It defines variables and allocate dynamic memory for arrays used in elastic LSRTM. It calls the function `elastic_lsrtm.c`.
- `read_par.c`: Function reads input parameters from a text file.
- `read_model.c`: Function reads background density, P- and S-wave velocity models from hard drive.
- `compute_fdc.c`: Function computes staggered-grid finite-difference (FD) stencil coefficients from input FD order (Liu and Sen, 2009).
- `check_cfl.c`: Function checks the stability condition (CFL condition) of the finite-difference time stepping.
- `check_dispersion.c`: Function checks the numerical dispersion of the FD modeling.
- `pad_model.c`: Function pads the input models by the width of absorbing boundary and the width of half the FD stencil length.
- `vel_to_lame.c`: Function calculates Lamé parameters from density, P- and S- wave velocities.

- `extend_rho_mu.c`: Function extends right and bottom sides of density and shear modulus model for one more layer to deal with the staggered pattern when calculating the effective media parameters.
- `effective_par.c`: Function calculates the effective media parameters on the staggered-grids.
- `initialize_PML.c`: Function constructs the damping profile in the unsplit Convolutional Perfectly Matched Layer (C-PML) boundary (Komatitsch and Martin, 2007).
- `split_shots.c`: Function divides the computational assignments by shots and distributes to different CPU cores.
- `elastic_lsrtm.c`: Function conducts elastic LSRTM via the conjugate gradient least squares (CGLS) algorithm (Paige and Saunders, 1982). It calls the function `elastc_born.c` as forward operator and the function `elastic_rtm.c` as adjoint operator.
- `read_data.c`: Function reads two-component observed data from hard drive.
- `elastic_rtm.c`: Function conducts elastic RTM. It calls function `wavefield_simulation.c` twice for forward and adjoint (backward) wave propagations, and computing kernels (elastic images).
- `wavefield_simulation.c`: Function conducts forward and adjoint (backward) wave propagation simulations using time domain elastic staggered-grid FD scheme (Virieux, 1986). The spatial FD order is selectable (can be an arbitrary order) and is an input parameter from `read_par.c`. The corresponding FD stencil coefficients are automatically calculated using function `compute_fdc.c`. The computation of the elastic RTM kernels requires access to the forward and backward wavefields at same time step. However, these two wavefields are computed in the reverse time direction. I adopt the source wavefield reconstruction method (Gauthier et al., 1986; Dussaud et al., 2008). During the forward simulation of the source-side wavefield, only the wavefield within the depth of half of the spatial finite difference operator length on boundaries and the final time snapshots are saved in memory. Then, the source-side wavefield is recomputed from the saved wavefield by backward propagation while simultaneously computing the backward receiver-side wavefield. In the source-side wavefield reconstruction, the time-stepping runs in time-reversal direction and the source function (Ricker wavelet) is subtracted from the wavefield. The function `wavefield_simulation.c` is responsible for forward simulation, adjoint simulation and source wavefield reconstruction. The different modes of the function are controlled by a switcher `int FORWARD` (`= 1` or `= 0`). It calls the following functions.

- `update_velocity_interior.c`: Function computes particle velocity fields using the staggered-grid FD stencil (Virieux, 1986). It updates the wavefield in the interior of the computational domain that excludes the C-PML boundaries.
- `update_stress_interior.c`: Function computes stress tensor fields using the staggered-grid FD stencil. It updates the wavefield in the interior of the computational domain that excludes the C-PML boundaries.
- `update_velocity_boundary.c`: Function computes particle velocity fields in the computational boundaries using the staggered-grid FD stencil and C-PML boundary condition (Komatitsch and Martin, 2007).
- `update_stress_boundary.c`: Function computes stress tensor fields in the computational boundaries using the staggered-grid FD stencil and C-PML boundary condition.
- `explosive_source.c`: Function adds the source term into (during forward simulation) or subtracts the source term from (during source-wavefield reconstruction) the wavefield during FD time stepping. The two modes are controlled by a switcher `int DIRECTION` (`= 1` or `= 0`).
- `wavelet.c`: Source wavelet function.
- `add_adjoint_src.c`: Function adds the adjoint source term into the wavefield during adjoint (backward) FD time stepping.
- `store_wavefield_boundary.c`: Function saves the particle velocity and stress tensor wavefields along the computational boundaries within the depth of half of the spatial FD stencil length during the forward simulation. The saved wavefields are used to reconstruct the source-side wavefield during adjoint simulation (back propagation).
- `store_last_snapshot.c`: Function saves the particle velocity and stress tensor wavefields at the last time step during the forward simulation. The saved wavefields are used to reconstruct the source-side wavefield during adjoint simulation (back propagation).
- `restore_last_snapshot.c`: Function re-injects the saved wavefields for source wavefield reconstruction.
- `restore_velocity_boundary.c`: Function re-injects the saved particle velocity wavefields along boundaries for source wavefield reconstruction.
- `restore_stress_boundary.c`: Function re-injects the saved stress tensor wavefields along boundaries for source wavefield reconstruction.

- `compute_gradients.c`: Function computes the elastic RTM images via the interaction of the forward and adjoint (backward) wavefields.
- `elastic_born.c`: Function conducts the elastic Born approximation modeling. The source-side wavefield modeling time stepping and the Born modeling time stepping are put in the same time loop to avoid the need of saving source-side wavefield.
- `add_scatter_source.c`: Functions adds the Born secondary source into the wavefields during the elastic Born modeling time stepping.

APPENDIX B

Software developed for this thesis: time-domain matrix-free elastic Gauss-Newton FWI (EGNFWI)

The software of the proposed time-domain matrix-free elastic Gauss-Newton method is also written in the C programming language and parallelized using the message passing interface (MPI). The development of this software package is based on the elastic LSRTM software package.

- `efwi.c`: Main function. It defines variables and allocate dynamic memory for arrays used in elastic Gauss-Newton FWI. It calls the function `elastic_gn_fwi.c` for computation.
- `read_par.c`: Function reads input parameters from a text file.
- `read_frequencies.c`: Function reads frequency bands used in multiscale method from a text file.
- `read_model.c`: Function reads background density, P- and S-wave velocity models from hard drive.
- `compute_fdc.c`: Function computes staggered-grid FD stencil coefficients from input FD order (Liu and Sen, 2009).
- `check_cfl.c`: Function checks the stability condition (CFL condition) of the FD time stepping.
- `check_dispersion.c`: Function checks the numerical dispersion of the FD modeling.
- `initialize_PML.c`: Function constructs the damping profile in the unsplit C-PML boundary (Komatitsch and Martin, 2007).

- `split_shots.c`: Function divides the computational assignments by shots and distributes to different CPU cores.
- `elastic_gn_fwi.c`: Function conducts elastic Gauss-Newton FWI. It calls function `forward_modeling.c` for simulating data, `elastic_lsrtm.c` for computing Gauss-Newton search direction, `line_search.c` for estimating the step size, and `update_model_test.c` for updating the model.
- `forward_modeling.c`: Function simulates wavefield using current model and calculates data residuals. It calls function `wavefield_simulation.c`.
- `band_pass_filter_1d.c`: Function filters the source wavelet using a time-domain Butterworth filter.
- `band_pass_filter_2d.c`: Function filters the observed seismic data using a time-domain Butterworth filter.
- `wavefield_simulation.c`: Function conducts forward and adjoint (backward) wave propagation simulations using time domain elastic staggered-grid FD scheme (Virieux, 1986). The spatial FD order is a user-defined parameter. During backward propagation, the function adopts the source wavefield reconstruction method (Gauthier et al., 1986; Dussaud et al., 2008) to recompute the source-side wavefield. The function is responsible for forward simulation, adjoint simulation and source wavefield reconstruction. The different modes of the function are controlled by a switcher `int FORWARD` (`= 1` or `= 0`).
- `elastic_lsrtm.c`: Function computes Gauss-Newton search direction using elastic LSRTM code. It uses the CGLS algorithm for solving a linear system of equations. It calls functions `elastic_rtm.c` and `elastic_born.c` a forward and adjoint operator pair.
- `read_data.c`: Function reads two-component observed data from hard drive.
- `elastic_rtm.c`: Function conducts elastic RTM. It calls function `wavefield_simulation.c` twice for forward and adjoint (backward) wave propagations, and computing FWI kernels.
- `update_velocity_interior.c`: Function computes particle velocity fields in the interior of the computational domain using the staggered-grid FD stencil (Virieux, 1986).
- `update_stress_interior.c`: Function computes stress tensor fields in the interior of the computational domain using the staggered-grid FD stencil.

- `update_velocity_boundary.c`: Function computes particle velocity fields in the computational boundaries using the staggered-grid FD stencil and C-PML boundary condition (Komatitsch and Martin, 2007).
- `update_stress_boundary.c`: Function computes stress tensor fields in the computational boundaries using the staggered-grid FD stencil and C-PML boundary condition.
- `explosive_source.c`: Function adds the source term into (during forward simulation) or subtracts the source term from (during source-wavefield reconstruction) the wavefield during FD time stepping. The two modes are controlled by a switcher `int DIRECTION` (`= 1` or `= 0`).
- `wavelet.c`: Source wavelet function.
- `add_adjoint_src.c`: Function adds the adjoint source term into the wavefield during adjoint (backward) FD time stepping.
- `store_wavefield_boundary.c`: Function saves the particle velocity and stress tensor wavefields along the computational boundaries within the depth of half of the spatial FD stencil length during the forward simulation. The saved wavefields are used to reconstruct the source-side wavefield during adjoint simulation (back propagation).
- `store_last_snapshot.c`: Function saves the particle velocity and stress tensor wavefields at the last time step during the forward simulation. The saved wavefields are used to reconstruct the source-side wavefield during adjoint simulation (back propagation).
- `restore_last_snapshot.c`: Function re-injects the saved wavefields for source wavefield reconstruction.
- `restore_velocity_boundary.c`: Function re-injects the saved particle velocity wavefields along boundaries for source wavefield reconstruction.
- `restore_stress_boundary.c`: Function re-injects the saved stress tensor wavefields along boundaries for source wavefield reconstruction.
- `compute_gradients.c`: Function computes the elastic FWI gradient kernels via the interaction of the forward and adjoint (backward) wavefields.
- `elastic_born.c`: Function conducts the elastic Born approximation modeling. The source-side wavefield modeling time stepping and the Born modeling time stepping are put in the same time loop to avoid the need of saving source-side wavefield.
- `add_scatter_source.c`: Function adds the Born secondary source into the wavefields during the elastic Born modeling time stepping.

- `line_search.c`: Function performs line search via parabola fitting approach. It alls functions `update_model_test.c`, `forward_modeling.c` and `calculate_step_length.c`.
- `pad_model.c`: Function pads the input models by the width of absorbing boundary and the width of half the FD stencil length.
- `vel_to_lame.c`: Function calculates Lamé parameters from density, P- and S- wave velocities.
- `extend_rho_mu.c`: Function extends right and bottom sides of density and shear modulus model for one more layer to deal with the staggered pattern when calculating the effective media parameters.
- `effective_par.c`: Function calculates the effective media parameters on the staggered-grids.
- `calculate_step_length.c`: Function solves coefficients of the parabola in parabola fitting line search and returns the step size.
- `update_model_test.c`: Function updates the model in line search trial step or updates the model using an estimated step size. The two modes are controlled by a switcher `int TEST` (`= 1` or `= 0`).

APPENDIX C

High-performance computing implementation

The numerical examples in this thesis are computed on the high-performance computing infrastructure provided by Compute Canada. The main systems used are Jasper, Parallel, Cedar (GP2) and Graham (GP3) of Compute Canada. Among them, Jasper and Parallel adopt the Portable Batch System (PBS) job scheduler, Cedar and Graham adopt the SLURM job scheduler. For the numerical examples of elastic LSRTM, I requested 101 CPU cores. For numerical examples of elastic Gauss-Newton FWI, I requested 61 CPU cores.