Important Notice

UNIVERSITY OF CALGARY

Waveform Inversion for Estimating Subsurface Properties: Phase-encoding Strategies, Optimization Methods, Interparameter Tradeoffs Quantification and Reduction

by

Wenyong Pan

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN GEOLOGY AND GEOPHYSICS

CALGARY, ALBERTA
August, 2017

Abstract

This thesis concerns seismic full-waveform inversion (FWI) techniques for estimating subsurface properties. FWI approaches promise to provide high-resolution estimates of subsurface parameters using full wavefield information. However, FWI also suffers from a series of obstacles including extensive computation requirements, slow convergence rate, cycle-skipping, interparameter tradeoffs, etc. This thesis focuses on developing advanced phase-encoding and optimizations methods for accelerating FWI, quantifying and reducing the interparameter tradeoffs in multiparameter FWI.

Iteratively minimizing the objective function and updating the models gives rise to high computational costs. In this thesis, I have developed phase-encoding approaches for constructing gradient and Hessian diagonals in the τ-p domain for accelerating FWI. Most of FWI applications employ gradient-based methods for updating the models by assuming the Hessian to be an identity matrix, which suffer from slow convergence rate. In this thesis, advanced second-order optimizations (i.e., l-BFGS and Hessian-free methods) are developed for improving the convergence rate. Different preconditioning strategies are examined for accelerating Hessian-free Gauss-Newton FWI.

Simultaneously reconstructing multiple physical parameters suffers from parameter crosstalk, a difficulty arising from inherent ambiguities among different physical parameters. Quantifying the coupling effects of different physical parameters is an essential part of multiparameter FWI. Most parameter resolution studies are based on scattering patterns. Ambiguities appear between different physical parameters if their scattering patterns overlap over a certain range of scattering angles. Scattering patterns of isotropic-elastic parameters with various parameterizations are derived in this thesis. The *interparameter contamination kernels* are introduced to explain the origins of interparameter tradeoffs. A novel inversion strategy with approximate contamination kernels is developed for providing more convincing density estimations in isotropic-elastic FWI. Synthetic examples and realistic seismic dataset examples are given to verify the effectiveness of this inversion strategy. Performances of different parameterizations for isotropic-elastic FWI are also examined.

This thesis also demonstrates that applying inverse multiparameter Hessian to precondi-

tion the gradients is able to suppress the unwanted interparameter contaminations. 3D scattering patterns of elastic constants in general anisotropic media are given. The second-order term in multiparameter Hessian, which accounts for multiparameter second-order scattering, can be constructed with adjoint-state approach. Newton-based methods are applied to reconstruct the elastic constants in anisotropic (HTI) media.

Preface

The PhD thesis is written in manuscript-based format based on three published papers and one submitted paper. All of these work is carried out with the supervision of Dr. Kris Innanen of CREWES project at University of Calgary. Dr. Kris Innanen was involved in all of my research projects as the supervisory author. These papers are republished in this thesis with the permission from the co-authors.

A version of chapter 2 has been published as: Pan, W., Innanen, K. A., Margrave, G. F., and Cao, D., 2015, Efficient pseudo-Gauss-Newton full-waveform inversion in the τ-p domain: Geophysics, 80, R225-R238. I am the lead investigator and the manuscript composer for this paper. Dr. Gary Margrave contributed to the concept of deconvolution imaging condition. Dr. Danping Cao was involved in the theory of inverse problems.

A version of chapter 3 has been published as: Pan, W., Innanen, K. A., and Liao, W., 2017, Accelerating Hessian-free Gauss-Newton full-waveform inversion via l-BFGS preconditioned conjugate-gradient algorithm: Geophysics, 82, R49-R64. I am the lead investigator and the manuscript composer for this paper. Dr. Wenyuan Liao contributed to the iterative methods for solving Newton linear system.

A version of chapter 4 has been submitted for publication: Wenyong Pan, Yu Geng and Kristopher Innanen, Interparameter tradeoff quantification and reduction in isotropic-elastic FWI: synthetic experiments and Hussar dataset application. I am the lead investigator and the manuscript composer for this paper. Dr. Yu Geng contributes the theory of adjoint-state method.

A version of chapter 5 has been published as: Pan, W., Innanen, K. A., Margrave, G. F., Fehler, M. C., Fang, X., and Li, J., 2016, Estimation of elastic constants for HTI media using Gauss-Newton and full-Newton multiparameter full-waveform inversion: Geophysics, 81, R275-R291. Dr. Gary Margrave was involved in understanding the Hessian in FWI. Dr. Mike Fehler contributed to the fractured reservoir characterization and understanding scattering theory. Dr. Xinding Fang and Junxiao Li contributed to the forward modelling problem.

Acknowledgements

First of all and foremost, I would like to express my gratitude to my supervisor Dr. Kris Innanen for his supervision, leadership, and advice during my PhD studies. His courses and lectures significantly enhance my theoretical background in wave physics and scattering theory, which play an important role in my research for inversion sensitivity analysis. His insights and guidance about the parameter crosstalk excited me greatly, which contributed to the development of multiparameter inverse problem a lot. Kris has guided me to investigate the parameter tradeoffs in isotropic-elastic FWI including scattering patterns, various parameterizations, the role of density, etc, which is the essential part of my thesis. Kris also gave me a lot of help and encouragement when I encountered difficulties in my candidacy exam. I also appreciate Kris for his continuous support for me to visit other universities, which greatly expanded my vision and improve my knowledge level. Thanks also to his recommendation and trust, I was able to work as sessional instructor for undergraduate students' courses. I am indebted to him more than he knows. I also gratefully acknowledge the funding support of CREWES sponsors, the National Science and Engineering Research Council of Canada (NSERC, CRDPJ 461179-13), the Eyes High International Doctoral Scholarship and SEG/Chevron Scholarship allowed me to attend conferences and complete my thesis.

I also thanks greatly to Dr. Gary Margrave, who is my co-supervisor for my PhD studies. Gary gave me a lot of guidance to understand the imaging condition in seismic imaging and the relationship between full-waveform inversion and standard inversion methodology in my GOPH 701 project. I worked as a TA for Gary's courses for three years, during which my basic knowledge on time series analysis and seismic data processing were greatly enhanced and improved. After I encountered difficulties in my candidacy exam, Gary encouraged me, corrected my mistakes and helped me practice the questions every week patiently. Thanks to Gary's support and recommendations, I was able to get the position of sessional instructor for GOPH 517 and GOPH 557. It a pity that Gary is not able to attend my PhD defense due to a time conflict. However, Gary is my supervisor forever.

Dr. Larry Lines is a very nice professor. He lent me his text book when I prepared for the candidacy exam and helped me understand the basic concepts in exploration geophysics

more clearly. He gave me a lot of insightful suggestions in my research and helped me understand the reflection coefficients more deeply when I taught GOPH 517. I learned the theory of wave propagation and reflection coefficients from Dr. Edward Krebes, which is very useful when deriving the scattering patterns and implementing FWI algorithms in my research. During my PhD studies, I had a lot of communications with Dr. Wenyuan Liao, which greatly broaden my knowledge in non-linear optimization methods, the algorithms for solving linear systems, etc. Fruitful discussions with Dr. Danping Cao helped a lot at early stage of my research. Thanks also to the valuable comments and suggestions from Dr. Daniel Trad and Dr. Sam Gray when I studied interparameter tradeoffs in isotropic-elastic FWI.

I would like to thanks the staff and students in CREWES sincerely. Thanks specially to Laura Baird, who takes care of the students in CREWES, books the flights for me to attend conferences, organizes the annual meeting and makes everything ready for the students. Thanks also to Kevin Hall and Kevin Bertram, who helped me install the software and provided technical support. Thanks to Joe Wong, who helped me learn the finite-difference modelling codes. Thanks also to my close friends Bona Wu, Tianci Cui, Peng Cheng, Siming (Emma) Lv, Huaizhen Chen and Jean Cui who encouraged me and helped me when I encountered difficulties in my life and research. Thanks also to Junxiao Li, who helped me understand finite-difference modelling more clearly. Thanks also to Dr. Yu Geng, who cooperated with me for developing FWI codes in frequency domain. Thanks also to Faranak Mahmoudian, who gave me a lot of suggestions in my research and helped me revise the thesis. Thanks also to Raul Cova, Scott Keating, Marcelo Andrade, Shaphpoor Moradi, Sergio Romahn, Jian Sun, Kiki Xu, Lei Yang and Zhan Niu for valuable discussions and help in my research.

I also got a lot of help when I visited ERL at M.I.T. in 2014. Dr. Mike Fehler hosted me, helped me learn the scattering theory in anisotropy media, gave me a lot of invaluable suggestions in inversion sensitivity analysis and recommended me for getting the postdoc position. Thanks also to Dr. Xinding Fang for providing forward modelling codes and fruitful suggestions. Thanks greatly to Dr. Frederik Simons, who hosted me when I visited Princeton University in 2016, gave invaluable comments on my research topics and invited

me to give presentation in Solid Earth Brown Bag Seminar at Princeton University. Thanks specially to Dr. Yanhua O. Yuan, who taught me how to do elastic FWI with spectral-element method patiently, supervised me in parameter tradeoffs analysis in isotropic-elastic FWI, etc. Thanks also to Princeton University and Compute Canada for providing parallel computing facilities.

I would not be able to finish my PhD research and thesis without the continuous supports from my family members. I appreciate sincerely my father Jinsheng Pan and mother Huanyin Ma, who worked very hard in China and provide funding support for me to pursue the PhD degree. Thanks greatly to my sister Wenzheng Pan, who provided funding support for me, took good care of me and encouraged me when I encountered difficulties in my life. Finally, I would like to thanks my girl friend Xuan Zhao, who encouraged me to pursue what I wanted and has accompanied me and waited for me since 2013.

Dedication

To my beloved parents (Jinsheng Pan and Huanyin Ma), sister (Wenzheng Pan), and girl friend (Xuan Zhao)

Contents

List of Tables

List of Figures

List of Symbols, Abbreviations

Φ	The misfit function
ϕ	Misfit value
$\|\cdot\|$	l-2 norm
$\mathbf{m}$	Model parameter vector
$\mathbf{x}$	Subsurface position
$\mathbf{d}_{\text{obs}}$	Observed data
$\mathbf{d}_{\text{syn}}$	Synthetic data
$\Delta\mathbf{d}$	Data residual vector
$\nabla_{\mathbf{m}}\Phi, \mathbf{g}$	Gradient of the misfit function
$\mathbf{H}$	Full Hessian
$\mathbf{H}_a$	Gauss-Newton Hessian
$\tilde{\mathbf{H}}_a$	Phase-encoded Gauss-Newton Hessian
$\bar{\mathbf{H}}$	The second-order term in full Hessian
$\mathbf{H}_{pseudo}$	Pseudo-Hessian
$\mathbf{H}_a^{\text{diag}}$	Diagonal Gauss-Newton Hessian
$\mathbf{H}_{pseudo}^{\text{diag}}$	Diagonal pseudo-Hessian
k	Iteration index
$\tilde{k}$	Inner iteration index in Hessian-free Gauss-Newton method
μ_k	Step length in kth iteration
p	Ray parameter
t	Time
$\mathbf{x}_s$	Source position
$\mathbf{x}_g$	Receiver position
ω	Angular frequency
f_s	Source function
$\tilde{f}_s$	Virtual source
$\mathbf{L}$	Impedance matrix
$\mathbf{u}$	Pressure wavefield vector

$\tilde{\mathbf{u}}$	Phase-encoded wavefield vector
∇^2	Laplace operator
δ	Dirac delta function
G	Green's function
G_{ij}	Green's tensor
γ	Phase shift function
c	Top surface velocity
$\tilde{\mathbf{d}}_{\text{obs}}$	Observed data in ray parameter domain
$\tilde{\mathbf{d}}_{\text{syn}}$	Synthetic data in ray parameter domain
$\Delta\mathbf{m}$	Model perturbation
$\nabla_{\mathbf{m}}\mathbf{d}_{\text{syn}}$	Partial derivative wavefield
$A(\omega)$	Real function depending on ω
$\mathbf{D}$	Forward modelled wavefield
$\mathbf{U}$	Backward propagated wavefield
$\mathbf{p^g}$	Ray parameter vector for gradient calculation
$\mathbf{p^H}$	Ray parameter vector for diagonal Hessian calculation
ψ	Encoding function
Δx_s	Source interval
N_p	Number of ray parameter
Δp	Ray parameter spacing
f_{max}	Maximum frequency
$*$	Complex conjugate
$\dagger$	Matrix transpose
$\star$	Complex conjugate transpose
ϵ	Relative least-squares error
$\mathbf{I}$	Identity matrix
$\nabla_{\mathbf{m}}\mathbf{L}$	Partial derivative of impedance matrix
$\mathcal{P}$	Receiver sampling operator

$\boldsymbol{\chi}$	Augmented Lagrangian functional
Λ	Lagrangian multiplier
$\mathcal{H}$	l-BFGS inverse Hessian approximation
$\mathbf{s}$	Model change
$\mathbf{y}$	Gradient change
$\mathbf{H}_a v$	Gauss-Newton Hessian-vector product
$\mathcal{M}$	Preconditioner
$\hat{\epsilon}, \hat{\lambda}$	Stabilization parameters for preconditioning
N_s	Source number
S	Maximum source index
R	Maximum receiver index
T	Maximum recording time
c_{ijkl}	Elastic constant tensor
ρ	Density
κ	Bulk modulus
μ	Shear modulus
α	P-wave velocity
β	S-wave velocity
K_p	Standard sensitivity kernel of model parameter p
$K_{p\leftrightarrow p}$	Correct update kernel for model parameter p
$K_{p\rightarrow q}$	Interparameter contamination kernel from parameter p to parameter q
$\tilde{K}_p$	New update kernel for model parameter p
A_p	Magnitude of the perturbation of model parameter p
$\boldsymbol{\mathfrak{J}}$	Jacobian-vector product
$\boldsymbol{\mathfrak{H}}$	Hessian-vector product
$\tilde{u}$	Adjoint wavefield
a_p	Relative model perturbation of model parameter p
$::$	Sequential contractions over four tensor indices
$\mathcal{N}$	Gaussian distribution
$\mathbf{E}$	Expectation value

$\Sigma_{\mathbf{vv}}$	Variance-covariance matrix of vector $\mathbf{v}$
$\odot$	Element-wise multiplication
$\oslash$	Element-wise division
∂	The first partial derivative
$\mathbf{H}^{-g}$	Generalized inverse of Hessian
$\mathbf{R}$	Resolution matrix
$\prod$	Diagonal matrix with diagonal elements as eigenvalues
$\lambda_{\dot{k}}$	The $\dot{k}$th eigenvalue of Hessian
Ξ	Matrix with columns as eigenvectors of Hessian
$\mathbf{a}_{\dot{k}}$	The $\dot{k}$th eigenvector of Hessian
Eig	Approximate eigenvalue volume
$\mathbf{H}_{qp}$	The sub-block matrix in Hessian corresponding model parameters p and q
IP	P-wave impedance
IS	S-wave impedance
ρ'	Density in velocity-density parameterization
ρ''	Density in impedance-density parameterization
$\boldsymbol{\mathcal{L}}$	Wave operator in isotropic-elastic media
$\boldsymbol{\mathcal{V}}$	Scattering potential operator in displacement space
$\grave{\mathbf{k}}_{\alpha}$	Wavenumber vector for incident P-wave
$\grave{\mathbf{k}}_{\beta}$	Wavenumber vector for incident S-wave
$\acute{\mathbf{k}}_{\beta}$	Wavenumber vector for scattered S-wave
$\acute{\mathbf{k}}_{\alpha}$	Wavenumber vector for scattered P-wave
$\cdot$	Divergence
$\times$	Curl
$\grave{\mathbf{E}}$	Rotation matrix for incident wave
$\acute{\mathbf{E}}$	Rotation matrix for scattered wave
$\grave{\mathbf{p}}_{\mathrm{SV}}$	Unit vector for incident SV-wave
$\grave{\mathbf{p}}_{\mathrm{SH}}$	Unit vector for incident SH-wave
$\acute{\mathbf{p}}_{\mathrm{SV}}$	Unit vector for scattered SV-wave
$\acute{\mathbf{p}}_{\mathrm{SH}}$	Unit vector for scattered SH-wave

$\grave{\mathbf{p}}_{\mathrm{P}}$	Unit vector for incident P-wave
$\acute{\mathbf{p}}_{\mathrm{SV}}$	Unit vector for scattered P-wave
$\tilde{\boldsymbol{\mathcal{V}}}$	Scattering potential operator in wave mode space
$\tilde{\mathcal{V}}_{\mathrm{MN}}$	Scattering potential from M wave to N wave
$\mathbb{R}^{p}_{\mathrm{MN}}$	Scattering coefficient from M wave to N wave due to model perturbation of p
$\mathbf{m}^{\mathrm{est}}$	Estimated model
$\grave{\mathrm{A}}^{p}_{k}$	Weighting coefficient for incident wave
$\acute{\mathrm{A}}^{p}_{j}$	Weighting coefficient for scattered wave
Ω	The whole volume
σ_{ij}	Stress tensor
e_{ij}	Strain tensor
c_{IJ}	Elastic constant tensor following Voigt recipe for indexes
ϑ	Inclination angle departing z axis for incident wave
φ	Azimuthal angle departing x axis for incident wave
θ	Inclination angle departing z axis for scattered wave
ϕ	Azimuthal angle departing x axis for scattered wave
$\Delta\check{\mathbf{M}}$	Equivalent moment tensor source for HTI media
$\Delta\check{\mathbb{M}}$	Reduced equivalent moment tensor source
N_x	Number of grid in x direction
N_z	Number of grid in z direction
U	Amplitude of the incident plane P-wave
U_{SV}	Amplitude of the incident plane SV-wave
U_{SH}	Amplitude of the incident plane SH-wave
$\hat{\mathbf{p}}$	Polarization vector of incident P-wave
$\hat{\mathbf{p}}^{\mathrm{SV}}$	Polarization vector of incident SV-wave
$\hat{\mathbf{p}}^{\mathrm{SH}}$	Polarization vector of incident SH-wave
$\grave{\mathbf{u}}_{\mathrm{P}}$	Incident plane P-wave
$\grave{\mathbf{u}}_{\mathrm{SV}}$	Incident plane SV-wave
$\grave{\mathbf{u}}_{\mathrm{SH}}$	Incident plane SH-wave

FWI	Full-waveform inversion
RFWI	Reflection full-waveform inversion
HTI	Horizontal Transverse Isotropy
VTI	Vertical Transverse Isotropy
AVO	Amplitude versus Offset
VSP	Vertical seismic profile
SD	Steepest-descent
NCG	Non-linear conjugate-gradient
CG	Conjugate-gradient
l-BFGS	Limited-memory Broyden, Fletcher, Goldfarb, Shanno
HF	Hessian-free
GN	Gauss-Newton
FN	Full-Newton
MPSF	Multiparameter point spread function
EMPSF	Extended Multiparameter point spread function
PGN	Pseudo-Gauss-Newton
RTM	Reverse time migration
PML	Perfectly matched layer
LU	Lower upper decomposition
RLSE	Relative least-squares error
SP	Shot-profile
TLPE	Traditional linear phase-encoding
SU	Slant update
FR	Fletcher-Reeves
PR	Polak-Ribière
HS	Hestenes-Stiefel
DY	Dai-Yuan
DPH	Diagonal pseudo-Hessian
DGH	Diagonal Gauss-Newton Hessian
PDGH	Pseudo diagonal Gauss-Newton Hessian

CG-GN	Non-preconditioned Hessian-free Gauss-Newton
PCG	Preconditioned conjugate-gradient method
DPH-GN	Diagonal pseudo-Hessian preconditioned Hessian-free Gauss-Newton method
DGH-GN	Diagonal Gauss Hessian preconditioned Hessian-free Gauss-Newton method
PDGH-GN	Pseudo diagonal Gauss Hessian preconditioned Hessian-free Gauss-Newton method
l-BFGS-GN	l-BFGS preconditioned Hessian-free Gauss-Newton method
SNR	Signal to noise ratio
3C	Three-component
AGC	Automatic gain control

Chapter 1

Introduction

Seismic inversion techniques (i.e., traveltime tomography, impedance inversion, etc.) are longstanding and important issues in exploration geophysics. The main objective of seismic inversion is to recover subsurface properties from the recorded seismic dataset for some practical applied purpose, most often for oil and gas reservoir. In traditional traveltime tomography methods, the traveltime information is picked from the first arrivals to estimate the near-surface velocity structures (Zhu et al., 1992; Zhang and Toksöz, 1998; Zhu et al., 2008). In impedance inversion methods, the estimated reflectivity section is transformed into impedance section with trace integration tying to well log data (Waters, 1978; Russell, 1998; Lloyd, 2013; Cui, 2015; Esmaeili, 2016). AVO (Amplitude-variation-with-offset) inversion techniques are based on linear approximations of the reflection coefficients obtained from the Zoeppritz equations (Aki and Richards, 2002; Krebes, 1987; Lines, 1999; Krebes and Daley, 2007; Innanen, 2011, 2013; Lines et al., 2014; Mahmoudian et al., 2014).

Full-waveform inversion (FWI) techniques were initiated by Lailly (1983) and Tarantola (1984) in the 1980's. Compared to traditional inversion approaches mentioned above, FWI methods employ full wavefield information (traveltime, phase, amplitude, etc.) for inversion and are able to provide high-resolution estimations of model parameters (as shown in Figure 1.1). However, the inversion process involves a large number of numerical simulations of wave equation, which significantly impeded its wide application in 1980's. Benefiting from the vast development of computer technologies in recent decades, full-waveform inversion (FWI) methods have more recently become practical in global- and exploration-scale geophysics for estimating subsurface (an) elastic and anisotropic properties (i.e., compressional wave velocity, density and attenuation) (Pratt et al., 1998b; Sirgue and Pratt, 2004; Tromp et al., 2005; Fichtner et al., 2006; Liu and Tromp, 2006; Virieux and Operto, 2009; Herrmann et al., 2009; Warner et al., 2013; Herrmann et al., 2013).

The basic idea of FWI is to iteratively update the model properties by minimizing a l-2 norm misfit function Φ which measures the differences between seismic observations $\mathbf{d}_{\text{obs}}$ and

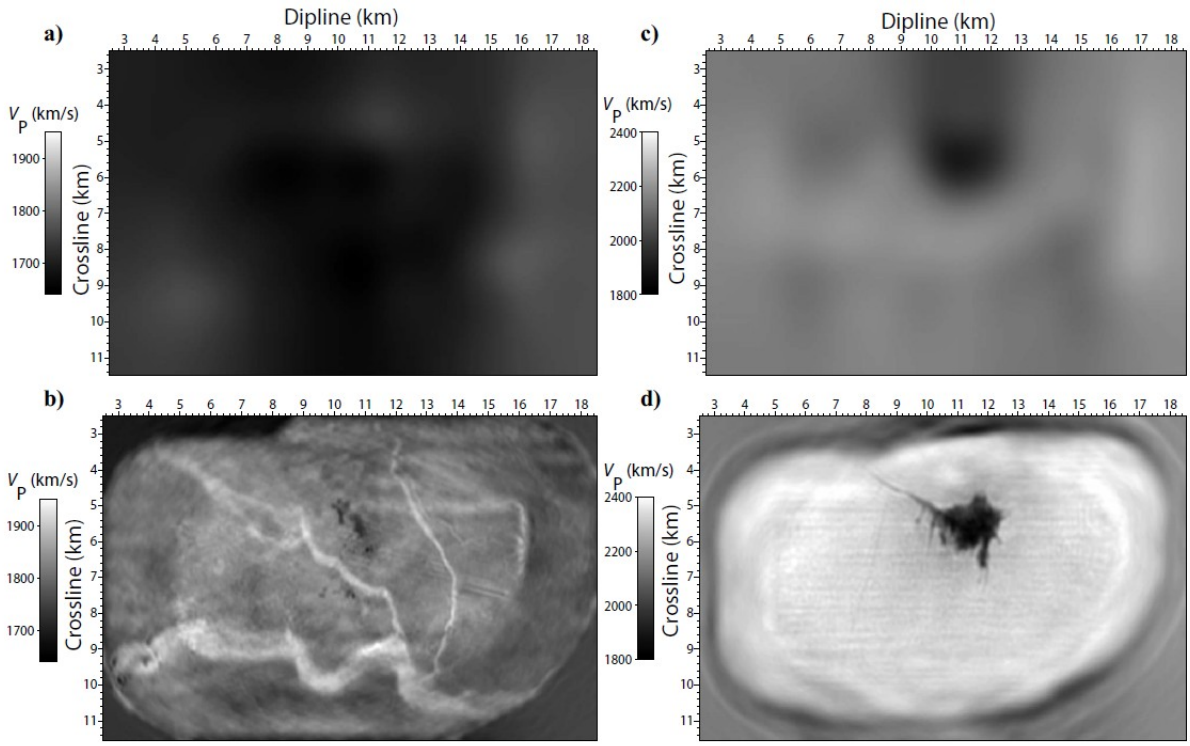

Figure 1.1: (a) and (b) show the 3D slices of velocity models built by ray-based reflection tomography in Valhall filed; (c) and (d) show the 3D slices of the velocity models built by full-waveform inversion method. The inversion results are obtained by L. Sirgue and O. I. Barkved. This figure is adapted from Virieux and Operto (2009).

the synthetic modelled data $\mathbf{d}_{\text{syn}}$, which can be obtained by numerically solving the wave equation with different approaches (i.e., finite-difference method, spectral-element method, etc). The objective of FWI is to find the model which makes the synthetic data match the observed seismic data. The misfit function Φ is classically solved with local optimization methods (Nocedal and Wright, 2006), in which the gradient $\nabla_{\mathbf{m}}\Phi$, the first derivative of the misfit function, needs to be calculated during the process of iteratively updating the model parameters. The gradient can be calculated by correlating the Fréchet derivative wavefields (Jacobian matrix) with the complex conjugate of data residuals (Virieux and Operto, 2009). However, it is computationally impractical to construct the Jacobian matrix explicitly for large-scale inverse problems. With the adjoint-state method and the Born approximation, the gradient can be calculated by zero-lag crosscorrelating the forward modelled wavefields with the backward propagated data residual wavefields efficiently (Tromp et al., 2005; Plessix, 2006; Fichtner et al., 2006; Geng et al., 2017b,a). Furthermore, the gradient in FWI can also

be calculated using one-way wave equation migration method by extrapolating the wavefields downward using phase-shift method (Margrave et al., 2011a). In different non-linear optimization methods, various inverse Hessian approximations are applied to precondition the gradient for improving the convergence rate (Nocedal and Wright, 2006).

1.1 Issues and inversion strategies in FWI

Full-waveform inversion techniques are promising but also suffer from a number of well-defined obstacles: chief amongst these are extensive computation requirements, slow convergence rate, the phenomenon of cycle-skipping, interparameter tradeoff issue in multiparameter FWI. These challenges make successful FWI applications very difficult. In this section, these problems in FWI are discussed and the potential strategies to overcome them are presented.

1.1.1 Extensive computation requirements

Lailly (1983) and Tarantola (1984) showed that the gradient of the least-squares objective function, when computed with the adjoint-state method (Plessix, 2006), is formed by the zero-lag crosscorrelation between the forward modelled wavefields and the backpropagated wavefields. Unfortunately, what remains still represents a very significant computational burden, especially for large 2D or 3D velocity models. With traditional shot-profile method, the computation cost of calculating gradient is equivalent to $2 \times N_s$ forward simulations (N_s is the source number). Iterative solution of the seismic inverse problem with standard multivariate optimization methods involves a large number of simulations, which make FWI very expensive.

Phase-encoding (or simultaneous-source) strategies, first introduced in pre-stack migration (Morton and Ober, 1998; Romero et al., 2000; Zhang et al., 2005; Liu et al., 2006; Dai and Schuster, 2013), can be employed to reduce the computational cost in gradient calculation. Phase-encoding involves the formation of super-gathers from summation of individual shots (Vigh and Starr, 2008; Krebs et al., 2009; Herrmann et al., 2009; Gao et al., 2010; Ben-Hadj-Ali et al., 2011; Guitton and Díaz, 2012; Tao and Sen, 2013; Castellanos

et al., 2015). Vigh and Starr (2008) originally performed the plane-wave domain FWI with a steepest-descent method. Anagaw and Sacchi (2014) focused on simultaneous-source FWI using random phase-encoding and showed that changing the encoding functions and source positions at every iteration can suppress crosstalk noise effectively. However, the simulation of simultaneous-sources introduces crosstalk artifacts, which negatively influence the inversion results. For random phase-encoding, crosstalk noise becomes stronger with increasing the number of encoded sources (Krebs et al., 2009; Berkhout, 2012). However, in the case of linear phase-encoding strategy, which can be interpreted as a τ-p transform (Zhang et al., 2005), when the sources are distributed densely and regularly over the whole acquisition geometry, the encoded wavefields approach plane wavefields and limited amount of noise will be present (Liu et al., 2006).

1.1.2 Non-linear optimization methods

Traditional optimization methods for FWI in exploration geophysics are gradient-based methods (i.e., steepest-descent (SD) and non-linear conjugate-gradient (NCG) methods). In the SD method, the search direction is simply the negative of the gradient and in NCG method, the search direction is the linear combination of the current gradient and previous search direction (Fletcher and Reeves, 1964; Nocedal and Wright, 2006; Hu et al., 2011). Gradient-based methods are computationally attractive for large-scale inverse problems. However, they suffer from slow convergence rates. The second-order partial derivative of the misfit function (namely the Hessian operator) carries crucial information in the reconstruction process (Santosa and Symes, 1988). The search direction can be significantly enhanced by multiplying the gradient with the inverse Hessian matrix, which serves as a "deconvolution" operator for compensating the geometrical spreading effects and de-blurring the gradient (Pratt et al., 1998b). However, explicit calculation, storage and inversion of the Hessian at each iteration is computationally impractical for large-scale inverse problems. Hence, various approaches have been proposed for approximating the Hessian (Shin et al., 2001a; Plessix and Mulder, 2004; Tang, 2009) or inverse Hessian (Nocedal and Wright, 2006; Demanet et al., 2012).

Instead of constructing the Hessian explicitly, quasi-Newton methods approximate the in-

verse Hessian iteratively by storing the model and gradient changes from previous iterations (Nocedal and Wright, 2006). One popular quasi-Newton method is the BFGS method (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970). However, the storage requirement of the inverse Hessian approximation and computational cost of preconditioning for large-scale inverse problems is still very high. To mitigate this difficulty, a limited-memory BFGS (l-BFGS) method is developed by storing information from a limited number l ($l < 10$) of previous iterations (Nocedal, 1980; Byrd et al., 1995; Nocedal and Wright, 2006). Compared to gradient-based methods, l-BFGS methods provide faster convergence rates for large-scale inverse problems (Brossier et al., 2010; Ma and Hale, 2012).

Hessian-free (HF) optimization methods (truncated-Newton or inexact-Newton method) represent attractive alternatives to the above-described optimization methods (Nash, 1985; Santosa and Symes, 1988; Nash, 2000; Akcelik et al., 2002; Métivier et al., 2014; Li and Demanet, 2016). At each iteration, the search direction is computed by approximately solving the Newton equations using a matrix-free scheme of the conjugate-gradient (CG) algorithm, which is an optimal method for solving a positive definite system (Nash, 1985; Hu et al., 2009). This linear iterative solver only requires the Hessian-vector products instead of forming the Hessian operator explicitly (Métivier et al., 2014). In this thesis, the full Hessian is replaced with the Gauss-Newton Hessian, which is always symmetric and positive semi-definite and which equals to the Hessian when the system is linear. One issue of the HF optimization method is that obtaining the search direction approximately requires a large number of CG iterations, which is still very expensive. Effective preconditioning strategy can be employed to accelerate HF full-waveform inversion (Nash, 2000; Sainath et al., 2013).

1.1.3 Cycle-skipping problem

Another big challenge of FWI is cycle-skipping, which originates from the nonlinearity of the seismic inverse problem. The seismic data relates to the model parameters non-linearly. However, the inversion process is classically solved with local Newton optimization framework and is therefore strongly dependent on the starting model definition. It only locates the global minimum if the starting model is close to the true model. If the initial model is far away from the true model, the solution will tend to become trapped at the local minimum (as

shown in Figure 1.2). One key assumption and criteria in the localized inversion is that the modelled and observed waveforms must be within half a wave-cycle to converge iteratively in the right direction. Theoretically, at low frequencies, there is a high chance that the modelled and observed data match within half a wave-cycle. Hence, the low frequencies are crucial to successful FWI applications (as shown in Figure 1.3). Multiscale approaches were proposed by expanding the frequency bands from low to high (Pratt and Chapman, 1992; Bunks et al., 1995; Sirgue and Pratt, 2004; Vigh and Starr, 2008; Boonyasiriwat et al., 2009). Low frequencies are responsible for recovering long-wavelength components of the model. However, in standard seismic data sets, which are naturally bandlimited, the low frequencies are always missing, which results in the failure of FWI applications.

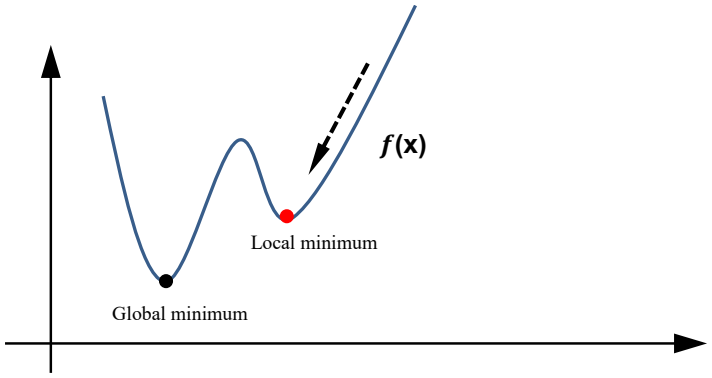

Figure 1.2: The cycle-skipping problem in non-linear inverse problem for minimizing quadratic function $f(x)$ with variable x. The red and black circles indicate local minimum and global minimum.

Different strategies have been proposed to overcome cycle-skipping in FWI. Because traveltime is more linearly related to the velocity structures, crosscorrelation based wave equation traveltime inversion methods can be used to build initial models for full-waveform inversion (Luo and Schuster, 1991; van Leeuwen and Mulder, 2010; Zhang et al., 2011; Ma and Hale, 2013). Xu et al. (2012) developed reflection full-waveform inversion method (RFWI), which uses reflected waves to recover deep parts of background model. Chi et al. (2015) combined RFWI with crosscorrelation based traveltime inversion method to recover long-wavelength of the velocity model. Shin and Cha (2008) proposed a robust waveform inversion algorithm, which is not sensitive to initial model by exploiting the wavefield in the Laplace domain. Wu et al. (2014) developed envelope waveform inversion to estimate

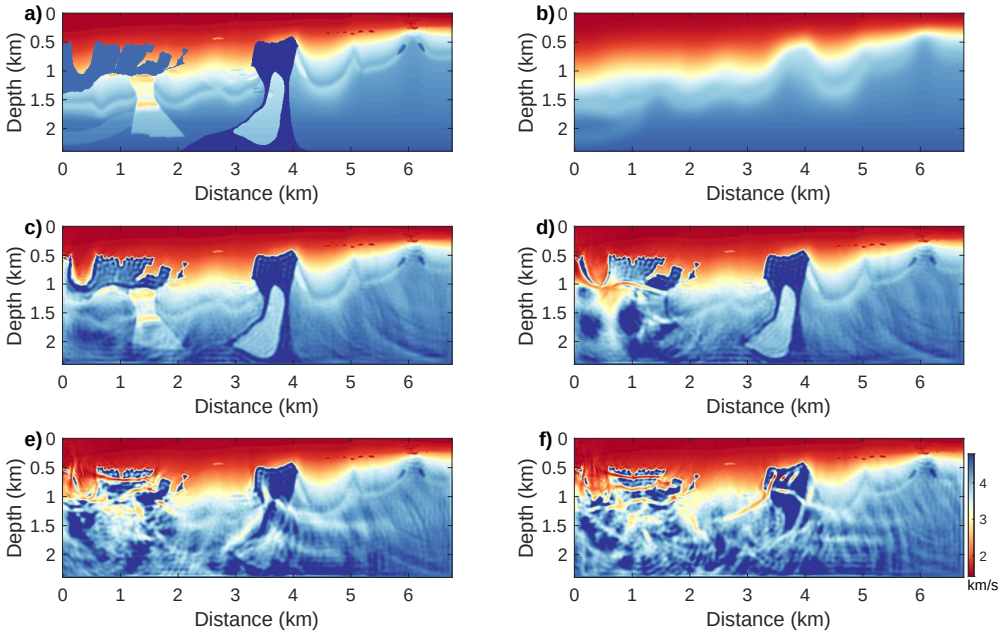

Figure 1.3: (a) and (b) show the true and initial models; (c), (d), (e) and (f) show the inverted models with the frequency bands of [1Hz, 25Hz], [3Hz, 25Hz], [6Hz, 25Hz] and [8Hz, 25Hz].

the long-wavelength velocity structure using the ultra-low frequency carried by envelope fluctuation and decay of the seismic records. Li and Demanet (2015) proposed to use high frequencies in the seismic records to synthesize the low frequencies, which are then used for full-waveform inversion. Esser et al. (2016) and Peters and Herrmann (2017) proposed constrained waveform inversion, which is able to reduce the cycle-skipping problem and build salt structures. Recently, Métivier et al. (2016) measured the FWI misfits between seismograms using an optimal transport distance, which has been proved to be less prone to cycle-skipping.

1.1.4 Interparameter tradeoff (or parameter crosstalk) in multiparameter FWI

Elastic and anisotropic parameters are important for reservoir characterization. Simultaneously reconstructing multiple physical parameters suffers from parameter crosstalk problem arising from the inherent ambiguities among these parameters, which increases the nonlinearity and uncertainty of the inverse problems significantly (Tarantola, 1986; Köhn et al., 2012; Operto et al., 2013; Alkhalifa and Plessix, 2014; Innanen, 2014a). Resolving abilities

of different parameterizations within FWI, even within fundamentally the same model (e.g., P-wave impedance, S-wave impedance and density versus P-wave velocity, S-wave velocity and density in isotropic- elastic FWI) are quite different. In recent years, researchers have devoted intensive efforts to the parameter resolution studies based on analytic solutions of Fréchet derivative wavefields ("scattering" or "radiation" patterns) for different parameter classes (Tarantola, 1986; Gholami et al., 2013b; Alkhalifa and Plessix, 2014; Podgornova et al., 2015). Coupling effects appear between two different physical parameters, if the scattered wavefileds due to the model perturbations overlap at certain range of scattering angle (Tarantola, 1986). These analyses are important for determining optimal parameterization, acquisition geometry and inversion strategies for multiparameter FWI. A high-resolution parameterization should have scattering patterns as different as possible and posterior uncertainties as uncorrelated as possible (Tarantola, 1986). Gholami et al. (2013b) investigated the scattering patterns of parameters resulting from various parameterizations of multiparameter acoustic FWI. Alkhalifa and Plessix (2014) emphasized the power of horizontal P-wave velocity in reducing the number of parameters for VTI FWI. Podgornova et al. (2015) analyzed the resolution limits of multiparameter anisotropic FWI based on the singular value decomposition of far-field linearized inversion operator. Overlapping the scattering patterns due to different physical parameters only represents an asymptotic approximation of Gauss-Newton Hessian (Operto et al., 2013; Alkhalifa and Plessix, 2014), which means that they do not provide complete information from which to understand the features of the interparameter contaminations (i.e., finite-frequency effects).

Reducing the uncertainties introduced by the interparameter tradeoffs is becoming essential for multiparameter FWI. Mode decomposition is potential strategy for mitigating the interparameter tradeoffs by isolating P and S wavefields but may also be limited in reducing the contaminations in density updates and multiparameter acoustic FWI (Wang and Cheng, 2017). Subspace optimization methods mitigate the interparameter tradeoffs by scaling different physical parameters but do not prevent their occurrence (Kennett et al., 1988; Bernauer et al., 2014). Newton-based optimization methods are promising for the capabilities of inverse multiparameter Hessian in suppressing the unwanted parameter crosstalk artifacts (Innanen, 2014a; Wang et al., 2016; Yang et al., 2016). Explicitly constructing and invert-

ing multiparameter Hessian for large-scale inverse problems is in practice computationally impracticable. Truncated-Newton (or Hessian-free) optimization methods represent affordable strategies for multiparameter FWI, in which the Newton equation is solved iteratively with matrix-free scheme of conjugate-gradient algorithm (Métivier et al., 2013; Boehm and Ulbrich, 2014; Métivier et al., 2015; Liu et al., 2015). However, with small number of inner iterations, the effectiveness of removing interparameter mappings may also not be obvious (Baumstein, 2014).

1.2 Thesis objectives and organization

The objectives of this thesis are: investigating efficient phase-encoding approaches for accelerating FWI, developing advanced non-linear optimization methods for improving the convergence rate of FWI, and quantifying and reducing the interparameter tradeoffs (or parameter crosstalk) in multiparameter FWI. The thesis is organized as follows:

Chapter 2 presents an efficient τ-p domain waveform inversion method aiming at reducing the computational burden of FWI with phase-encoding techniques. It was published in Geophysics in 2015 (Pan et al., 2015a). The gradient is constructed using linear phase-encoding and a slant update strategy is introduced to further reduce the computational burden. Poorly scaled and blurred gradient updates can be enhanced using exact or approximate versions of the inverse Hessian, which leads to a faster convergence rate. It is also revealed that preconditioning the gradient with diagonal pseudo-Hessian resembles a deconvolution imaging condition. The phase-encoding approach is also employed to construct diagonal Hessian, which serves as effective preconditioner for accelerating FWI.

Chapter 3 presents various non-linear optimization methods for full-waveform inversion. It was published in Geophysics in 2017 (Pan et al., 2017a). Most FWI applications employ gradient-based optimization methods by assuming the Hessian as an identity matrix, which is computationally attractive for large-scale inverse problems but suffer from slow convergence rates. Newton-based methods provide quadratic convergence but involve the computationally intractable problem of calculating and inverting the Hessian matrix explicitly in large-scale inverse problems. Instead of constructing the Hessian explicitly, the limited-memory BFGS

(l-BFGS) approximates the inverse Hessian iteratively using the gradient and model changes from a limited number l ($l < 10$) of previous iterations. In Hessian-free (HF) optimization methods (truncated-Newton or inexact-Newton method), the search direction is obtained by approximately solving the Newton equations using linear conjugate-gradient (CG) algorithm. I develop different preconditioning strategies for the inner CG algorithm and found that the l-BFGS inverse Hessian preconditioner works best for accelerating the Hessian-free Gauss-Newton FWI. Different non-linear optimization methods are finally applied to reconstruct velocity and density parameters in multiparameter acoustic FWI for comparison.

Chapter 4 focuses on the interparameter tradeoff issue in isotropic-elastic FWI. The material in this chapter has recently been submitted for publication. When simultaneously inverting multiple physical parameters, perturbation of one parameter will be mapped into the update of another, which presents unwanted parameter crosstalk artifacts in the inverted models. In this chapter, I give the scattering patterns within various parameterizations for interparameter tradeoff analysis in isotropic-elastic FWI and also define *interparameter contamination kernels*, which measure the parameter leakage among different physical parameters directly. Products of the multiparameter Hessian with an arbitrary vector can be calculated with adjoint-sate approach efficiently. To reduce interparameter contaminations in the inversion process, I develop a novel inversion strategy based on approximate contamination kernels, which is able to provide more convincing density estimations. Both synthetic data examples and realistic example with Hussar seismic dataset are given. I also examine the performances of three parameterizations (velocity-density, modulus-density and impedance-density parameterizations) in isotropic-elastic FWI. I found that the velocity-density parameterization still represents a better choice to reconstruct isotropic-elastic parameters than modulus-density and impedance-density parameterizations.

Chapter 5 presents the Newton-based (Gauss-Newton and full-Newton) methods for recovering elastic constants in anisotropic HTI media. This was published in Geophysics in 2016 (Pan et al., 2016). I derive the 3D scattering patterns for elastic constants in general anisotropic-elastic media for interparameter tradeoffs analysis. It is shown that by applying inverse multiparameter Hessian to precondition the gradient updates is able to suppress the interparameter contaminations effectively. The second-order term of multiparameter

Hessian, which accounts for multiparameter second-order scattering effects, can be calculated with adjoint-state method more efficiently. Furthermore, numerical examples are given to show that the second-order term of multiparameter Hessian is able to suppress the artifacts due to multiparameter second-order scattering effects in the gradient updates.

In **Chapter 6**, I summarize the thesis' contributions and innovations in phase-encoding approaches, non-linear optimization methods, interparameter tradeoffs analysis and reduction in multiparameter FWI. Plans for further studies on multiparameter 3D FWI are given at the end.

1.3 Thesis contributions

Contributions of this thesis are summarized as follows:

- Developing an efficient full-waveform inversion strategy with phase-encoding approaches in τ-p domain. I reveal that diagonal pseudo-Hessian is equivalent to source illumination and preconditioning the gradient with diagonal pseudo-Hessian resembles deconvolution imaging condition

- Developing various non-linear optimization methods including steepest-descent (SD), non-linear conjugate-gradient (NCG), l-BFGS and (preconditioned) Hessian-free Gauss-Newton methods.

- Analyzing the interparameter tradeoffs of isotropic-elastic parameters within various parameterizations in isotropic-elastic FWI based upon scattering patterns and multiparameter Hessian-vector products. Developing inversion strategy with approximate contamination kernels to reduce the influences of interparameter contaminations in the inversion process. The coupling effects of isotropic-elastic parameters within various parameterizations are studied. The inversion performances of various parameterizations are also evaluated with numerical examples.

- The 3D scattering patterns of elastic stiffness coefficients in general anisotropic-elastic media for analyzing the coupling effects of different parameters in anisotropic media are derived. Numerical examples are given to verify the role of multiparameter Hessian

in reducing the unwanted parameter crosstalk artifacts. The second-order term in multiparameter Hessian is able to removed the artifacts due to multiparameter second-order scattering effects. Newton-based methods are finally applied to inverted elastic constants in a 2D HTI media.

Chapter 2

Phase-encoding strategies

2.1 Summary

In this chapter, an efficient τ-p domain waveform inversion with a phase-encoding (or simultaneous-source) technique is developed aiming at reducing the computational burden of FWI. The gradient is constructed in the τ-p domain using linear phase-encoding, and a slant update strategy further reduces the computational burden. Poorly scaled and blurred gradient updates can be enhanced using exact or approximate versions of the inverse Hessian, which leads to a faster convergence rate. I develop a new chirp phase-encoding strategy for diagonal Hessian construction. Preconditioning the gradient using the diagonal phase-encoded approximate Hessian forms what I refer to as a pseudo-Gauss-Newton (PGN) step. To test the effectiveness of the τ-p domain FWI, the strategies were enacted on a modified Marmousi model. The effectiveness of the proposed phase-encoding strategies in frequency domain is also examined.

2.2 Introduction

In full-waveform inversion (FWI), one seeks to iteratively estimate subsurface properties by minimizing the difference between observed and numerically modelled seismic data (Lailly, 1983; Tarantola, 1984; Pratt et al., 1998b; Virieux and Operto, 2009; Herrmann et al., 2013; Warner et al., 2013). Industrial applications of FWI suffer from a series of well-defined obstacles: chief amongst these are extensive computation requirements per iteration, slow convergence rate, and the phenomenon of cycle-skipping. Aiming at reducing the computational burden, a pseudo-Gauss-Newton method carried out in the τ-p domain is proposed in this chapter.

For the traditional shot-profile method, iterative solution of the seismic inverse problem with standard multivariate optimization methods involves an impractical number of simu-

lations. Lailly (1983) and Tarantola (1984) showed that the gradient of the least-squares objective function, when computed with the adjoint-state method (Plessix, 2006), is formed by the zero-lag correlation between the forward modelled wavefields and the backpropagated wavefields. Unfortunately, what remains still represents a very significant computational burden, for large 2D or 3D velocity models. Strategies based on phase-encoding, first introduced in pre-stack migration (Morton and Ober, 1998; Romero et al., 2000; Zhang et al., 2005; Liu et al., 2006; Dai and Schuster, 2013), can be employed to reduce the computational cost of the gradient calculation. The phase-encoding technique involves the formation of super-gathers from summation of individual shots (Vigh and Starr, 2008; Krebs et al., 2009; Herrmann et al., 2009; Gao et al., 2010; Ben-Hadj-Ali et al., 2011; Guitton and Díaz, 2012; Tao and Sen, 2013; Castellanos et al., 2015). However, the simulation of simultaneous-sources introduces crosstalk artifacts, which negatively influence the inversion results. For random phase-encoding, crosstalk noise becomes stronger with increasing the number of encoded sources (Krebs et al., 2009; Berkhout, 2012). Anagaw and Sacchi (2014) focused on simultaneous-source FWI using random phase-encoding and showed that changing the encoding functions and source positions at every iteration can suppress crosstalk noise effectively. However, in the case of linear phase-encoding strategy, which can be interpreted as a τ-p transform (Zhang et al., 2005), when the sources are distributed densely and regularly over the whole acquisition geometry, the encoded wavefields approach plane wavefields and limited amount of noise will be present (Liu et al., 2006). Plane-wave migration still suffers from undesired artifacts due to the simultaneous extrapolation of multiple shots (Tang, 2009), which can be mitigated by stacking a set of ray parameters. Vigh and Starr (2008) originally performed the plane-wave domain FWI with a steepest-descent method. This chapter focuses on constructing the gradient for FWI using a linear phase-encoding strategy. To reduce the computational cost further, a slant update strategy is introduced, in which the model is updated using the slant gradient with one single ray parameter, but the p value is changed as the iteration proceeds. Thus, the bias in any update can be balanced during later iterations; this allows the mitigation of artifacts and aliasing. This strategy enables a similar quality inversion result, but with a considerably reduced computational burden. To illustrate these aspects of the candidate strategy, the computational cost of the τ-p domain

method with traditional methods are compared and the quality of the inversion results are evaluated.

Given sparsely sampled sources, the linear phase-encoding technique generates strong crosstalk noise, arising from the undesired interactions between unrelated source and receiver wavefields. It is important therefore to analyze the influence of source spacing on τ-p domain FWI as carried out with the slant update strategy. Generally, different ray parameters are responsible for illuminating and updating of subsurface layers with different dip angles. The ray parameter range could potentially be determined by the geological structures in the target area. So, a proper ray parameter range should be designed to balance the updates and guarantee the convergence rate. The sensitivity of τ-p domain FWI to the ray parameter range setting is discussed.

The gradient in FWI is equivalent to a reverse time migration (RTM) image constructed using a crosscorrelation imaging condition. This means a steepest-descent update is blurred and poorly-scaled, having neglected amplitude loss during forward and backward propagation (Shin et al., 2001a,b). Poor scaling increases the number of updates needed to attain the global minimum. This can be significantly improved by multiplying the gradient with the inverse Hessian. The Hessian operator acts like a nonstationary "deconvolution" operator to compensate for geometrical spreading and de-blur the gradient (Pratt et al., 1998b; Pan et al., 2014a). The resulting improvement in scaling increases the convergence rate, but, unfortunately, the calculation, storage and inversion of the Hessian greatly increases the per iteration computational burden, tending to reverse any savings. Shin et al. (2001a) introduced the pseudo-Hessian, which allows one to compute an approximate diagonal Hessian at the same cost as the gradient by replacing the partial derivative wavefields with the virtual sources during the autocorrelation process. Tang (2009) introduced a receiver-side phase-encoded Hessian by constructing the receiver-side Green's functions using random phase-encoding. In this chapter, a chirp phase-encoding method is introduced, which combines the linear and random phase-encoding strategies, for diagonal Hessian construction within the Gauss-Newton approximation. This chirp phase-encoding strategy is similar to the mixed shot-encoding scheme for wave equation migration by Perrone and Sava (2012). Preconditioning the gradient with the diagonal phase-encoded approximate Hessian in τ-p

domain forms what I refer to as pseudo-Gauss-Newton (PGN) method. The phenomenon of cycle-skipping originates from the nonlinearity of the seismic inverse problem and the use of an inaccurate initial velocity model. This, coupled with the lack of low-frequency data, results in the global minimum being missed in favour of some nearby local minimum. In this chapter, to improve the chance of avoiding cycle-skipping and attaining the global minimum, a multi-scale approach in time domain is implemented by low-pass filtering the data residuals (Pratt and Chapman, 1992; Bunks et al., 1995; Sirgue and Pratt, 2004; Vigh and Starr, 2008; Boonyasiriwat et al., 2009).

This chapter is organized as follows. First, the basic theory for least-squares inverse problem is reviewed and the slant update strategy with phase-encoding and varied p values is introduced. Then, the construction of diagonal Hessian using phase-encoding method combining with several numerical examples in the numerical section is explained. The proposed strategies are then practiced on a modified Marmousi model to verify the slant update strategy, to analyze the influences of ray parameter range and source spacing, and to compare different preconditioning strategies. I then implement the proposed phase-encoding strategies in frequency domain and numerical examples based on Marmousi model are also given for illustration.

2.3 Theory and Methods

In this section, the basic principle of forward modelling and least-squares inverse problem with linear phase-encoding are reviewed, and the slant update strategy with varied p values is introduced. Then how to construct the diagonal Hessian using phase-encoding method is described.

2.3.1 Forward modelling in τ-p (or frequency-ray parameter) domain

Forward modelling of wave propagation is one important step for FWI. The acoustic wave equation with constant-density in the time domain is (Marfurt, 1984):

$$\nabla^2 \bar{u}(\mathbf{x}, \mathbf{x}_s, t) - m(\mathbf{x}) \frac{\partial^2 \bar{u}(\mathbf{x}, \mathbf{x}_s, t)}{\partial t^2} = \bar{f}_s(\mathbf{x}_s, t), \tag{2.1}$$

where $\mathbf{x} = (x, y, z)$ denotes the subsurface location with Cartesian coordinates, ∇^2 is the Laplace operator, $m(\mathbf{x})$ is the square of the slowness, and $\bar{u}(\mathbf{x}, \mathbf{x}_s, t)$ denotes the pressure wavefields at position $\mathbf{x}$ and time t due to the source $\bar{f}_s(\mathbf{x}_s, t)$ excited at position $\mathbf{x}_s$. To obtain the forward solution $\bar{u}(\mathbf{x}, \mathbf{x}_s, t)$ of equation (2.1), an explicit finite-difference method with fourth-order accuracy in space and second-order accuracy in time is used (Virieux, 1986; Levander, 1988). The perfectly matched layer (PML) boundary condition is applied on all of the boundaries of the model (Berenger, 1994). In frequency domain, equation (2.1) becomes the Helmholtz equation which can be written as (Marfurt, 1984)

$$\mathbf{L}\left(\mathbf{x}, \omega; \mathbf{m}\right) \mathbf{u}\left(\mathbf{x}, \mathbf{x}_s, \omega\right) = f_s(\omega)\delta\left(\mathbf{x} - \mathbf{x}_s\right), \tag{2.2}$$

where $\mathbf{m}$ is the model parameter vector, $\mathbf{u}(\mathbf{x}, \mathbf{x}_s, \omega)$ is the discrete pressure wavefield vector. $\mathbf{L}\left(\mathbf{x}, \omega; \mathbf{m}\right) = [\nabla^2 + \omega^2 \mathbf{m}(\mathbf{x})]$ is the discretized impedance matrix. ω is angular frequency and $\delta\left(\mathbf{x} - \mathbf{x}_s\right)$ is the Dirac delta function. Note that the FWI algorithms are implemented in both of time domain and frequency domain. The notations in the following part of this section are expressed in frequency domain for sake of compactness. The Green's function $G\left(\mathbf{x}, \mathbf{x}_s, \omega\right)$ is defined as the solution of equation (2.2) due to a point source:

$$\mathbf{L}\left(\mathbf{x}, \omega; \mathbf{m}\right) G\left(\mathbf{x}, \mathbf{x}_s, \omega\right) = \delta\left(\mathbf{x} - \mathbf{x}_s\right). \tag{2.3}$$

The solution of equation (2.2) can be written as:

$$\mathbf{u}\left(\mathbf{x}, \mathbf{x}_s, \omega\right) = f_s\left(\omega\right) G\left(\mathbf{x}, \mathbf{x}_s, \omega\right). \tag{2.4}$$

Traditional shot-profile RTM can provide high quality images but typically at high computational cost (Romero et al., 2000). Plane-wave source migration with slant stacking (or delayed-shot migration) has been introduced to reduce the computational burden by applying different phase shifts at densely distributed sources (Morton and Ober, 1998; Romero et al., 2000; Zhang et al., 2005; Dai and Schuster, 2013). Godwin and Sava (2013) compared different phase-encoding schemes for wave equation migration. The linear phase-encoding is performed by applying linear phase shifts (or time delays in the time domain) at densely distributed sources (see Figure 2.1). The phase shift function $\gamma\left(\mathbf{x}_s, p, \omega\right) = \omega p\left(\mathbf{x}_s - \mathbf{x}_s^0\right)$ is associated with ray parameter (or slant parameter) p and source location $\mathbf{x}_s$. A common-

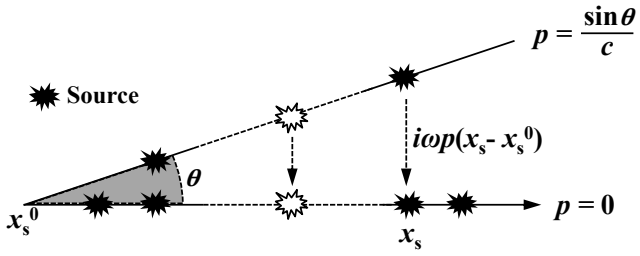

Figure 2.1: Linear phase-encoding strategy. The phase shift $\omega p(x_s - x_s^0)$ is controlled by ray parameter p and source position x_s. The ray parameter p is a function of take-off angle θ and top surface velocity c.

receiver gather can be transformed into ray parameter super-gather from a line source wavefield, through a τ-p transform (Zhang et al., 2005; Trad et al., 2002, 2003):

$$\tilde{\mathbf{u}}\left(\mathbf{x}_g, p, \omega\right) = \sum_{\mathbf{x}_s} \mathbf{u}\left(\mathbf{x}_g, \mathbf{x}_s, \omega\right) \mathbf{exp}\left(i\omega p(\mathbf{x}_s - \mathbf{x}_s^0)\right), \tag{2.5}$$

where $\mathbf{x}_s^0$ indicates the location of the initial source and $\mathbf{x}_g$ indicate the receiver locations. The synthetic ray parameter gather in frequency domain is generated with a line source by applying phase shifts at the source locations. The corresponding wave equation is given by:

$$\mathbf{L}\left(\mathbf{x}, \omega\right) \tilde{\mathbf{u}}(\mathbf{x}, p, \omega) = \sum_{\mathbf{x}_s} \mathbf{exp}\left(i\omega p\left(\mathbf{x}_s - \hat{\mathbf{x}}_s\right)\right) f_s(\omega)\delta\left(\mathbf{x} - \mathbf{x}_s\right), \tag{2.6}$$

when $p \geq 0$, $\hat{\mathbf{x}}_s$ indicates the location of initial source $\mathbf{x}_s^0$, if $p < 0$, $\hat{\mathbf{x}}_s$ indicates the location of the right most source $\mathbf{x}_s^{\text{end}}$. The solution of equation (2.6) with a line source can be written as:

$$\tilde{\mathbf{u}}(\mathbf{x}, p, \omega) = \sum_{\mathbf{x}_s} f_s(\omega) G\left(\mathbf{x}, \mathbf{x}_s, \omega\right) \mathbf{exp}\left(i\omega p\left(\mathbf{x}_s - \hat{\mathbf{x}}_s\right)\right). \tag{2.7}$$

In frequency domain, the linear equation (2.6) is solved with a direct solver based on multifrontal Lower Upper (LU) decomposition (Davis and Duff, 1997), which is efficient for a multi-source problem with forward and backward substitutions (Tao and Sen, 2013). In equation (2.6), because the number of ray parameters is generally much smaller than the number of sources, the number of backward substitutions is considerably reduced (Wu et al., 2015).

2.3.2 Formulating non-linear least-squares inverse problem

As a non-linear least-squares optimization problem, FWI seeks to estimate the subsurface parameters through an iterative process by minimizing the difference between the synthetic data $\tilde{\mathbf{d}}_{\text{syn}}$ and observed data $\tilde{\mathbf{d}}_{\text{obs}}$ (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009). In frequency-ray parameter domain, the misfit function Φ with least-squares norm can be written as:

$$\Phi\left(\mathbf{m}\right)=\frac{1}{2}\sum_{\mathbf{p}}\sum_{\mathbf{x}_g}\sum_{\omega}\|\tilde{\mathbf{d}}_{\text{obs}}\left(\mathbf{x}_g,p,\omega\right)-\tilde{\mathbf{d}}_{\text{syn}}\left(\mathbf{x}_g,p,\omega;\mathbf{m}\right)\|^2, \tag{2.8}$$

where $\|\cdot\|$ means the ℓ-2 norm, $\mathbf{p}$ is the ray parameter vector and observed data $\tilde{\mathbf{d}}_{\text{obs}}\left(\mathbf{x}_g,p,\omega\right)$ can be obtained with τ-p transform:

$$\tilde{\mathbf{d}}_{\text{obs}}\left(\mathbf{x}_g,p,\omega\right)=\sum_{\mathbf{x}'_s}\mathbf{d}_{\text{obs}}\left(\mathbf{x}_g,\mathbf{x}'_s,\omega\right)\mathbf{exp}\left(i\omega p(\mathbf{x}'_s-\hat{\mathbf{x}}_s)\right), \tag{2.9}$$

where $\mathbf{d}_{\text{obs}}\left(\mathbf{x}_g,\mathbf{x}'_s,\omega\right)$ is the observed data in source-receiver domain corresponding to source position $\mathbf{x}'_s$. The Newton optimization approach is developed based on the second-order Taylor-Lagrange expansion of the misfit function Φ:

$$\Phi\left(\mathbf{m}+\Delta\mathbf{m}\right)\approx\Phi\left(\mathbf{m}\right)+\mathbf{g}^{\dagger}\Delta\mathbf{m}+\frac{1}{2}\Delta\mathbf{m}^{\dagger}\mathbf{H}\Delta\mathbf{m}, \tag{2.10}$$

where the symbol "$\dagger$" means transpose, $\Delta\mathbf{m}$ is the search direction, $\mathbf{g}=\nabla_{\mathbf{m}}\Phi\left(\mathbf{m}\right)$ and $\mathbf{H}=\nabla_{\mathbf{m}}\nabla_{\mathbf{m}}\Phi\left(\mathbf{m}\right)$ indicate gradient and Hessian respectively. To minimize the quadratic approximation of the misfit function, the updated model at the $(k+1)$th iteration can be written as the sum of the model at the kth iteration and the search direction $\Delta\mathbf{m}_k$:

$$\mathbf{m}_{k+1}=\mathbf{m}_k+\mu_k\Delta\mathbf{m}_k, \tag{2.11}$$

where μ_k is the step length, a scalar constant calculated through a line search method (Gauthier et al., 1986; Pica et al., 1990; Nocedal and Wright, 2006). Within a Newton optimization framework, the search direction $\Delta\mathbf{m}_k$ is the solution of the Newton linear system:

$$\mathbf{H}_k\Delta\mathbf{m}_k=-\mathbf{g}_k. \tag{2.12}$$

The gradient is the first-order partial derivative of the misfit function with respect to the model parameter and it indicates the direction in which the misfit function is increasing most

rapidly (Pratt et al., 1998b). Gradient vectors can be constructed by zero-lag correlation between the Fréchet derivative wavefield with complex conjugate of the data residuals. The gradient vector for square of slowness $\mathbf{m}$ can be written as:

$$\mathbf{g}\left(\mathbf{x}\right)=\sum_{\mathbf{p}}\sum_{\mathbf{x}_g}\sum_{\omega}\Re\left(\nabla_{\mathbf{m(x)}}\tilde{\mathbf{d}}_{\text{syn}}^{\dagger}\left(\mathbf{x}_g,p,\omega;\mathbf{m}\right)\Delta\tilde{\mathbf{d}}^{*}\left(\mathbf{x}_g,p,\omega\right)\right), \tag{2.13}$$

where the symbol "$*$" means complex conjugate, $\Re\left(\cdot\right)$ denotes the real part, and $\tilde{\mathbf{d}}\left(\mathbf{x}_g,p,\omega\right)$ denotes the data residual vector with ray parameter p:

$$\begin{aligned}\Delta\tilde{\mathbf{d}}\left(\mathbf{x}_g,p,\omega\right)&=\sum_{\mathbf{x}'_s}\left(\mathbf{d}_{\text{obs}}\left(\mathbf{x}_g,\mathbf{x}'_s,\omega\right)-\mathbf{d}_{\text{syn}}\left(\mathbf{x}_g,\mathbf{x}'_s,\omega;\mathbf{m}\right)\right)\mathbf{exp}\left(i\omega p(\mathbf{x}'_s-\hat{\mathbf{x}}_s)\right),\\&=\sum_{\mathbf{x}'_s}\Delta\mathbf{d}\left(\mathbf{x}_g,\mathbf{x}'_s,\omega\right)\mathbf{exp}\left(i\omega p(\mathbf{x}'_s-\hat{\mathbf{x}}_s)\right),\end{aligned} \tag{2.14}$$

where $\Delta\mathbf{d}\left(\mathbf{x}_g,\mathbf{x}'_s,\omega\right)$ can be considered as the data residual vector in source-receiver domain. In equation (2.13), $\nabla_{\mathbf{m(x)}}\tilde{\mathbf{d}}_{\text{syn}}\left(\mathbf{x}_g,p,\omega;\mathbf{m}\right)$ represents the Fréchet derivative wavefield due to perturbations of model parameter $\mathbf{m}$ at position x. With the Born approximation, it can be expressed as:

$$\nabla_{\mathbf{m(x)}}\tilde{\mathbf{d}}_{\text{syn}}\left(\mathbf{x}_g,p,\omega;\mathbf{m}\right)\approx-\omega^2\sum_{\mathbf{x}_s}f_s(\omega)G\left(\mathbf{x}_s,\mathbf{x},\omega\right)G(\mathbf{x},\mathbf{x}_g,\omega)\mathbf{exp}\left(i\omega p\left(\mathbf{x}_s-\hat{\mathbf{x}}_s\right)\right). \tag{2.15}$$

Inserting equations (2.14) and (2.15) into equation (2.13), the gradient vector is obtained as:

$$\begin{aligned}\tilde{\mathbf{g}}\left(\mathbf{x}\right)=-\sum_{\mathbf{p^g}}\sum_{\mathbf{x}_g}\sum_{\omega}\Re\Bigg(&\omega^2\sum_{\mathbf{x}_s}f_s(\omega)G\left(\mathbf{x}_s,\mathbf{x},\omega\right)G(\mathbf{x},\mathbf{x}_g,\omega)\mathbf{exp}\left(i\omega p_j^{\mathbf{g}}\left(\mathbf{x}_s-\hat{\mathbf{x}}_s\right)\right)\\&\times\sum_{\mathbf{x}'_s}\left|A\left(\omega\right)\right|^2\Delta\mathbf{d}^{*}\left(\mathbf{x}_g,\mathbf{x}'_s,\omega\right)\mathbf{exp}\left(i\omega p_j^{\mathbf{g}}(\hat{\mathbf{x}}_s-\mathbf{x}'_s)\right)\Bigg),\end{aligned} \tag{2.16}$$

where $A\left(\omega\right)$ represents a weighting function depending on angular frequency ω (Liu et al., 2006), $\mathbf{p^g}$ indicates the ray parameter vector for gradient calculation and $p_j^{\mathbf{g}}$ indicates one single ray parameter with index of j. This expression for gradient calculation in ray parameter domain can also be derived by crosscorrelating forward modelled wavefields $\mathbf{D}$ with backward propagated data residual wavefields $\mathbf{U}$:

$$\mathbf{D}\left(\mathbf{x},p_j^{\mathbf{g}},\omega\right)=\sum_{\mathbf{x}_s}f_s(\omega)G\left(\mathbf{x},\mathbf{x}_s,\omega\right)A\left(\omega\right)\mathbf{exp}\left(i\omega p_j^{\mathbf{g}}(\mathbf{x}_s-\hat{\mathbf{x}})\right), \tag{2.17}$$

$$\mathbf{U}^{*}\left(\mathbf{x}, p_j^{\mathbf{g}}, \omega\right) = \sum_{\mathbf{x}_s'} G\left(\mathbf{x}, \mathbf{x}_g, \omega\right) \Delta\mathbf{d}^{*}\left(\mathbf{x}_g, \mathbf{x}_s', \omega\right) A^{*}\left(\omega\right) \mathbf{exp}\left(-i\omega p_j^{\mathbf{g}}(\mathbf{x}_s' - \hat{\mathbf{x}})\right). \tag{2.18}$$

When $p_j^{\mathbf{g}} \geq 0$, $\hat{\mathbf{x}} = \mathbf{x}_s^0$, with $\mathbf{x}_s^0$ being the position of the initial source. When $p_j^{\mathbf{g}} < 0$, $\hat{\mathbf{x}} = \mathbf{x}_s^{\text{end}}$, with $\mathbf{x}_s^{\text{end}}$ being the position of the right most source. The gradient with slant parameter $p_j^{\mathbf{g}}$ can be expressed as:

$$\begin{aligned}
\tilde{\mathbf{g}}\left(\mathbf{x}, p_j^{\mathbf{g}}\right) &= -\sum_{\omega} \Re\left(\omega^2 \mathbf{D}(\mathbf{x}, p_j^g, \omega)\mathbf{U}^{*}(\mathbf{x}, p_j^g, \omega)\right), \\
&= -\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_s'}\sum_{\omega} \Re\left(\omega^2 f_s(\omega) G\left(\mathbf{x}, \mathbf{x}_s, \omega\right) G\left(\mathbf{x}, \mathbf{x}_g, \omega\right) \Delta\mathbf{d}^{*}\left(\mathbf{x}_g, \mathbf{x}_s', \omega\right)\right. \\
&\times \mid A\left(\omega\right) \mid^2 \mathbf{exp}\left(i\omega p_j^{\mathbf{g}}(\mathbf{x}_s - \mathbf{x}_s')\right)\Big),
\end{aligned} \tag{2.19}$$

where the factor ω^2 comes from the time differentiation (Sirgue and Pratt, 2004; Margrave et al., 2010). In equation (2.19), when $\mathbf{x}_s = \mathbf{x}_s'$, the slant gradient $\tilde{\mathbf{g}}(\mathbf{x}, p_j^{\mathbf{g}})$ is equal to the conventional shot-profile gradient $\mathbf{g}\left(\mathbf{x}\right)$. When $\mathbf{x}_s \neq \mathbf{x}_s'$, the slant gradient $\tilde{\mathbf{g}}\left(\mathbf{x}, p_j^{\mathbf{g}}\right)$ becomes the crosstalk term $\mathbf{g}_{cross}$. The slant gradient with ray parameter $p_j^{\mathbf{g}}$ can be written as the sum of the conventional shot-profile gradient and the crosstalk term:

$$\tilde{\mathbf{g}}\left(\mathbf{x}, p_j^{\mathbf{g}}\right) = \mathbf{g}\left(\mathbf{x}\right) + \mathbf{g}_{cross}. \tag{2.20}$$

To suppress the crosstalk term in the above equation, the phase-encoded gradient is constructed by summing over a set of ray parameters:

$$\begin{aligned}
\tilde{\mathbf{g}}\left(\mathbf{x}\right) = \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_s'}\sum_{\omega}\sum_{\tilde{\mathbf{p}}^{\mathbf{g}}} \Re\left(\omega^2 f_s(\omega) G\left(\mathbf{x}, \mathbf{x}_s, \omega\right) G\left(\mathbf{x}, \mathbf{x}_g, \omega\right) \Delta\mathbf{d}^{*}\left(\mathbf{x}_g, \mathbf{x}_s', \omega\right)\right. \\
\left.\times \mid A\left(\omega\right) \mid^2 \mathbf{exp}\left(i\omega\tilde{\mathbf{p}}^{\mathbf{g}} \cdot (\mathbf{x}_s - \mathbf{x}_s')\right)\right),
\end{aligned} \tag{2.21}$$

where $\tilde{\mathbf{p}}^{\mathbf{g}}$ represents the ray parameter vector for phase-encoding. Reorganizing equation (2.21) gives:

$$\tilde{\mathbf{g}}\left(\mathbf{x}\right) = -\sum_{\mathbf{x}_g}\sum_{\mathbf{x}_s'}\sum_{\mathbf{x}_s}\sum_{\omega} \Re\left(\omega^2 f_s(\omega) G\left(\mathbf{x}_s, \mathbf{x}, \omega\right) G(\mathbf{x}, \mathbf{x}_g, \omega) \Delta\mathbf{d}^{*}\left(\mathbf{x}_g, \mathbf{x}_s', \omega\right) \psi\left(\mathbf{x}_s, \mathbf{x}_s', \omega\right)\right), \tag{2.22}$$

where $\psi\left(\mathbf{x}_s, \mathbf{x}_s', \omega\right)$ is the encoding function:

$$\psi\left(\mathbf{x}_s, \mathbf{x}_s', \omega\right) = \sum_{\tilde{\mathbf{p}}^{\mathbf{g}}} A^2\left(\omega\right) \mathbf{exp}\left(i\omega\tilde{\mathbf{p}}^{\mathbf{g}} \cdot (\mathbf{x}_s - \mathbf{x}_s')\right). \tag{2.23}$$

If the number of ray parameters in vector $\tilde{\mathbf{p}}^{\mathbf{g}}$ is sufficient large, integrating over ray parameters yields a Dirac delta function in space (Liu et al., 2006; Tang, 2009):

$$\psi\left(\mathbf{x}_s, \mathbf{x}'_s, \omega\right) = \frac{A^2\left(\omega\right)}{\omega^2}\delta\left(\mathbf{x}_s - \mathbf{x}'_s\right). \tag{2.24}$$

Plugging equation (2.24) into equation (2.22) gives:

$$\tilde{\mathbf{g}}\left(\mathbf{x}\right) = -\sum_{\mathbf{x}_g}\sum_{\mathbf{x}'_s}\sum_{\mathbf{x}_s}\sum_{\omega}\Re\left(\omega^2 f_s(\omega)G\left(\mathbf{x}_s, \mathbf{x}, \omega\right)G(\mathbf{x}, \mathbf{x}_g, \omega)\Delta\mathbf{d}^*\left(\mathbf{x}_g, \mathbf{x}'_s, \omega\right)\frac{A^2\left(\omega\right)}{\omega^2}\delta\left(\mathbf{x}_s - \mathbf{x}'_s\right)\right). \tag{2.25}$$

Making use of the sifting property of the delta function and setting $\mid A\left(\omega\right) \mid^2$ satisfy to $\mid A\left(\omega\right) \mid^2 = \mid \omega \mid^2$, the phase-encoded gradient is equivalent to the shot-profile gradient:

$$\tilde{\mathbf{g}}\left(\mathbf{x}\right) = \mathbf{g}\left(\mathbf{x}\right) = \sum_{\mathbf{x}_s}\sum_{\omega}\Re\left(\omega^2 f_s(\omega)G\left(\mathbf{x}, \mathbf{x}_s, \omega\right)G\left(\mathbf{x}, \mathbf{x}_g, \omega\right)\Delta\mathbf{d}^*\left(\mathbf{x}_g, \mathbf{x}_s, \omega\right)\right). \tag{2.26}$$

Generally, $2N_p^g$ simulations are required in one FWI iteration with slant stacking, where N_p^g is the number of the ray parameters for calculating the gradient. To reduce the computational cost further, I propose a slant update strategy. In this approach, the slant gradient with single slant parameter is used in place of the phase-encoded gradient with slant stacking. But, the slant parameter is changed iteration by iteration. This reduces the required number of simulations to 2 per iteration, while allowing us to mitigate bias by scheduling p values as updating proceeds.

2.3.3 Source spacing: anti-aliasing and crosstalk noise

The number of ray parameters N_p and ray parameter spacing Δp for linear phase-encoding have been studied by many researchers (Stork and Kapoor, 2004; Etgen, 2005; Zhang et al., 2005; Gray, 2013). Assuming that the source sampling interval Δx_s is small enough and the spread length is great enough, Zhang et al. (2005) gave the rule for determining N_p, when considering equation (2.5) as a Fourier transform. The anti-aliasing rule given by Zhang et al. (2005) is independent of source spacing Δx_s. While recalling the anti-aliasing rule of Radon transform:

$$\Delta x_s \leq \frac{1}{N_p \Delta p f_{max}}, \tag{2.27}$$

where f_{max} is the maximum frequency in the seismic data. If the ray parameter range $N_p\Delta p$ has been determined by the geological structures in the target area and the maximum frequency f_{max} has also been defined for multi-scale FWI, the source spacing Δx_s should be small enough to avoid aliasing (Vigh and Starr, 2008; Kwon et al., 2015).

Furthermore, considering the crosstalk artifacts in the gradient calculation, when the sources are distributed densely and regularly across the whole acquisition geometry, the linear phase-encoding technique generates only a limited amount of noise (Liu et al., 2006). However, when the shots are sparsely sampled, the crosstalk noise arising from undesired interaction between the unrelated source and receiver wavefields becomes very serious. Note that this is different from the condition in random phase-encoding, in which the crosstalk noise becomes stronger with increasing the number of encoded sources (Krebs et al., 2009; Anagaw and Sacchi, 2014). If the ray parameter range in linear phase-encoding is fixed, stacking over sufficient ray parameters at each iteration is required to reduce the artifacts caused by sparse source experiments. On the other hand, for densely distributed sources, sparse p sampling can be used to create high-quality image or gradient. Hence, it can be concluded that to mitigate crosstalk artifacts, the source spacing Δx_s is proportional to the number of ray parameters $\tilde{N}_p$ for stacking at each iteration:

$$\Delta x_s \propto \tilde{N}_p. \tag{2.28}$$

2.3.4 Hessian approximations

The full Hessian operator $H(\mathbf{x}, \mathbf{x}')$, which is the second-order partial derivative of the misfit function with respect to the model parameters, can be written as the summation of a second-order term $\bar{H}(\mathbf{x}, \mathbf{x}')$ and a first-order term $H_a(\mathbf{x}, \mathbf{x}')$:

$$H(\mathbf{x}, \mathbf{x}') = \bar{H}(\mathbf{x}, \mathbf{x}') + H_a(\mathbf{x}, \mathbf{x}'). \tag{2.29}$$

The second-order term is a correlation between the complex conjugate of the data residuals and the second-order partial derivative wavefields. This second-order term works as a demultiple operator to suppress the second-order scattering effects in the gradient. It can be accounted for through truncated-Newton strategy by employing the second-order adjoint-state method for computing Hessian-vector products (Métivier et al., 2013, 2014; Pan et al.,

2017a, 2018). When assuming that the initial model is close to the real model, the second-order term in the full Hessian operator can be neglected. The full Hessian under these circumstances can be replaced by the approximate Hessian, the first-order term, which acts as a preconditioner for the gradient in Gauss-Newton method. The approximate Hessian $H_a(\mathbf{x}, \mathbf{x}')$ is expressible as:

$$H_a(\mathbf{x}, \mathbf{x}') \approx \sum_{\mathbf{x}_s} \sum_{\mathbf{x}_g} \sum_{\omega} \Re \left(\omega^4 G\left(\mathbf{x}_g, \mathbf{x}', \omega\right) G\left(\mathbf{x}', \mathbf{x}_s, \omega\right) G^*\left(\mathbf{x}_g, \mathbf{x}, \omega\right) G^*\left(\mathbf{x}, \mathbf{x}_s, \omega\right) \right). \quad (2.30)$$

Each element in the approximate Hessian is the scalar product of two partial derivative wavefields (Pratt et al., 1998b), which is equivalent to the zero-lag crosscorrelation computation in time domain. Within the Born approximation, the two partial derivative wavefields can be simplified to four Green's functions, which forms the approximate Hessian expression in equation (2.30) (Plessix and Mulder, 2004). In the high-frequency limit and assuming a reference model with relatively smooth impedance variations, the two partial derivative wavefields are largely uncorrelated with each other but perfectly self-correlated (Pratt et al., 1998b; Shin et al., 2001b; Tang, 2009), which means that the approximate Hessian is diagonally dominant. Off-diagonal elements of the approximate Hessian, the crosscorrelation of the partial derivative wavefields, can for this reason be neglected. Thus, the diagonal elements of the approximate Hessian, the autocorrelation between the source-side Green's functions and receiver-side Green's functions, can act as a preconditioner for the gradient. The diagonal approximate Hessian can be expressed as when setting $\mathbf{x} = \mathbf{x}'$:

$$H_a^{\text{diag}}(\mathbf{x}) = \sum_{\mathbf{x}_s} \sum_{\mathbf{x}_g} \sum_{\omega} \Re \left(\omega^4 G\left(\mathbf{x}_g, \mathbf{x}, \omega\right) G\left(\mathbf{x}, \mathbf{x}_s, \omega\right) G^*\left(\mathbf{x}_g, \mathbf{x}, \omega\right) G^*\left(\mathbf{x}, \mathbf{x}_s, \omega\right) \right). \quad (2.31)$$

Under the assumption of infinite receiver coverage, the influence of the receiver-side Green's functions in the Hessian can be approximated as a constant scalar (Plessix and Mulder, 2004). This approximation to Hessian matrix is equivalent to the pseudo-Hessian proposed by Shin et al. (2001b). The faint gradient or image can be enhanced effectively by multiplying the inverse of the diagonal pseudo-Hessian (Jang et al., 2009), which can be written as:

$$H_{pseudo}^{\text{diag}}(\mathbf{x}) = \sum_{\mathbf{x}_s} \sum_{\omega} \Re \left(\omega^4 G\left(\mathbf{x}, \mathbf{x}_s, \omega\right) G^*\left(\mathbf{x}, \mathbf{x}_s, \omega\right) \right). \quad (2.32)$$

It is also observed that the diagonal pseudo-Hessian is actually equivalent to the source illumination (as shown in Appendix A) and the gradient preconditioned by the diagonal pseudo-Hessian resembles a deconvolution imaging condition.

2.3.5 Phase-encoded Hessian

The Green's functions in approximate Hessian (equation (2.30)) can also be constructed using phase-encoding technique, which reduces the computational cost considerably, but unfortunately also introduces strong crosstalk artifacts. The crosstalk artifacts can be reduced with sufficient source and receiver ray parameters. The receiver-side linear phase-encoded Hessian can be written as:

$$\begin{aligned}\tilde{H}_a\left(\mathbf{x},\mathbf{x}'\right) = &\sum_{\mathbf{x}_s}\sum_{\omega}\Re\left(\omega^4 G(\mathbf{x}',\mathbf{x}_s,\omega)G^*(\mathbf{x},\mathbf{x}_s,\omega)\right)\\ &\times\sum_{\mathbf{x}'_g}\sum_{\omega}\sum_{\mathbf{p}_g^{\mathbf{H}}}\Re\left(G(\mathbf{x}',\mathbf{x}'_g,\omega)A_g\left(\omega\right)\mathbf{exp}\left(i\omega\mathbf{p}_g^{\mathbf{H}}\cdot\left(\mathbf{x}'_g-\hat{\mathbf{x}}_g\right)\right)\right)\\ &\times\sum_{\mathbf{x}_g}\sum_{\omega}\sum_{\mathbf{p}_g^{\mathbf{H}}}\Re\left(G^*(\mathbf{x},\mathbf{x}_g,\omega)A_g^*\left(\omega\right)\mathbf{exp}\left(-i\omega\mathbf{p}_g^{\mathbf{H}}\cdot\left(\mathbf{x}_g-\hat{\mathbf{x}}_g\right)\right)\right),\end{aligned} \tag{2.33}$$

where $\mathbf{p}_g^{\mathbf{H}}$ indicates the receiver-side ray parameters vector for diagonal Hessian construction, $A_g\left(\omega\right)$ is the weighting function for receiver-side phase-encoding, the subscript g means receiver and $\hat{\mathbf{x}}_g$ means the position of initial receiver ($p_{g,j}^{\mathbf{H}}\geq 0$) or right most receiver ($p_{g,j}^{\mathbf{H}}<0$). Extracting the encoding function $\psi(\mathbf{x}'_g,\mathbf{x}_g,\omega)$ from equation (2.33):

$$\psi(\mathbf{x}'_g,\mathbf{x}_g,\omega)=\sum_{\mathbf{p}_g^{\mathbf{H}}}\mid A_g\left(\omega\right)\mid^2\mathbf{exp}\left(i\omega\mathbf{p}_g^{\mathbf{H}}\cdot\left(\mathbf{x}'_g-\mathbf{x}_g\right)\right). \tag{2.34}$$

It is noticed that the crosstalk noise arises from undesired interactions between the receiver-side Green's functions corresponding to different receiver positions $\mathbf{x}_g$ and $\mathbf{x}'_g$. Note that the crosstalk production regime is a little different from that in the phase-encoded gradient (equation (2.21)). So, in equation (2.33), when $\mathbf{x}'_g=\mathbf{x}_g$, the phase-encoded Hessian becomes the non-encoded approximate Hessian $\mathbf{H}_a$. When $\mathbf{x}'_g\neq\mathbf{x}_g$, crosstalk artifacts term $\mathbf{H}_{cross}$ remains. Thus, equation (2.33) can be written as the summation of the non-encoded Hessian and crosstalk artifacts (Tang, 2009; Tao and Sen, 2013):

$$\tilde{\mathbf{H}}_a=\mathbf{H}_a+\mathbf{H}_{cross}. \tag{2.35}$$

Considering three dimensions and stacking over sufficient receiver-side ray parameters, the encoding function $\psi(\mathbf{x}'_g, \mathbf{x}_g, \omega)$ can be reduced to:

$$\psi(\mathbf{x}'_g, \mathbf{x}_g, \omega) = \frac{\mid A_g(\omega) \mid^2}{\mid \omega \mid^2} \delta(\mathbf{x}'_g - \mathbf{x}_g). \tag{2.36}$$

Making $\mid A_g(\omega) \mid^2 = \mid \omega \mid^2$, the crosstalk term in equation (2.35) can be dispersed completely (Tang, 2009).

In this research, a chirp phase-encoding strategy is introduced, which is a combination of linear phase-encoding and random phase-encoding, to calculate the receiver-side Green's functions. A random factor is added into the phase delay term of equation (2.33), which gives:

$$\begin{aligned} \tilde{H}_a(\mathbf{x}, \mathbf{x}', \varepsilon) = & \sum_{\mathbf{x}_s} \sum_{\omega} \Re\left(\omega^4 G(\mathbf{x}', \mathbf{x}_s, \omega) G^*(\mathbf{x}, \mathbf{x}_s, \omega)\right) \\ & \times \sum_{\mathbf{x}_g} \sum_{\mathbf{x}'_g} \sum_{\omega} \Re\left(G(\mathbf{x}', \mathbf{x}'_g, \omega) G^*(\mathbf{x}, \mathbf{x}_g, \omega)\right) \\ & \times \sum_{\mathbf{x}_g} \sum_{\mathbf{x}'_g} \sum_{\omega} \sum_{\mathbf{p}_g^{\mathbf{H}}} \Re\left(\mid A_g(\omega) \mid^2 \mathbf{exp}\left(i\omega(\mathbf{p}_g^{\mathbf{H}} + \varepsilon \Delta \mathbf{p}_g) \cdot \left(\mathbf{x}'_g - \mathbf{x}_g\right)\right)\right), \end{aligned} \tag{2.37}$$

where ε is the parameter used to control phase shift dither and $\Delta \mathbf{p}_g$ is the ray parameter dithering vector and its element is defined as:

$$\Delta p_{g,j} = \zeta p_{g,j}^{\mathbf{H}}, \tag{2.38}$$

where $p_{g,j}^{\mathbf{H}}$ is the jth ray parameter element in $\mathbf{p}_g^{\mathbf{H}}$ and ζ is small random value in $[-1, 1]$. Chirp phase-encoding strategy is expected to reduce the crosstalk noise more efficiently than the linear phase-encoding, with the same computational cost. The explicit expressions for the diagonal approximate Hessian with both source-side and receiver-side phase-encoding and diagonal pseudo-Hessian with only source-side encoding are written as:

$$\begin{aligned} \tilde{H}_a^{\text{diag}}(\mathbf{x}, \varepsilon) = & \sum_{\mathbf{x}_s} \sum_{\mathbf{x}'_s} \sum_{\omega} \sum_{\mathbf{p}_s^H} \Re\left(\omega^4 G(\mathbf{x}, \mathbf{x}'_s, \omega) G^*(\mathbf{x}, \mathbf{x}_s, \omega)\right. \\ & \left. \times \mid A_s(\omega) \mid^2 \mathbf{exp}\left(i\omega(\mathbf{p}_s^{\mathbf{H}} + \varepsilon \Delta \mathbf{p}_s) \cdot (\mathbf{x}'_s - \mathbf{x}_s)\right)\right) \\ & \times \sum_{\mathbf{x}_g} \sum_{\mathbf{x}'_g} \sum_{\omega} \sum_{\mathbf{p}_g^H} \Re\left(G(\mathbf{x}, \mathbf{x}'_g, \omega) G^*(\mathbf{x}, \mathbf{x}_g, \omega)\right. \\ & \left. \times \mid A_g(\omega) \mid^2 \mathbf{exp}\left(i\omega(\mathbf{p}_g^{\mathbf{H}} + \varepsilon \Delta \mathbf{p}_g) \cdot (\mathbf{x}'_g - \mathbf{x}_g)\right)\right), \end{aligned} \tag{2.39}$$

$$\tilde{H}^{\text{diag}}_{pseudo}(\mathbf{x},\varepsilon) = \sum_{\mathbf{x}_s}\sum_{\mathbf{x}'_s}\sum_{\omega}\sum_{\mathbf{p}_s^H} \Re\left(\omega^4 G(\mathbf{x},\mathbf{x}'_s,\omega)G^*(\mathbf{x},\mathbf{x}_s,\omega)\right. \\ \left. \times \mid A_s(\omega) \mid^2 \exp\left(i\omega(\mathbf{p}_s^{\mathbf{H}} + \varepsilon\Delta\mathbf{p}_s)\cdot(\mathbf{x}'_s - \mathbf{x}_s)\right)\right), \tag{2.40}$$

where $\mathbf{p}_s^{\mathbf{H}}$ is the source-side ray parameter vector, $\Delta\mathbf{p}_s$ is the source-side dithering vector and $A_s(\omega)$ is the weighting function for source-side phase-encoding. If the parameter ε in equations (2.39) and (2.40) is equal to zero, the chirp phase-encoding becomes linear phase-encoding.

2.3.6 Pseudo-Gauss-Newton step

As mentioned above, to reduce the computational cost further, updating the velocity model using the slant gradient with varied p values as the iteration proceeds is proposed. Preconditioning the gradient using the phase-encoded diagonal approximate Hessian forms one pseudo-Gauss-Newton step, as indicated by equation (2.41):

$$\Delta\mathbf{m}(\mathbf{x}) = -\frac{\tilde{\mathbf{g}}\left(\mathbf{x}, p_j^{\mathbf{g}}\right)}{\tilde{H}_a^{\text{diag}}(\mathbf{x},\varepsilon) + \breve{\lambda}\mathbf{I}}, \tag{2.41}$$

where $\breve{\lambda}\mathbf{I}$ is the damping term, $\breve{\lambda}$ is a small constant value and $\mathbf{I}$ is an identity matrix. The relative least-squares error (RLSE) ϵ is used to evaluate the quality of the inverted model (Moghaddam et al., 2013):

$$\epsilon = \frac{\parallel \mathbf{m}_k - \mathbf{m}_{\text{true}} \parallel^2}{\parallel \mathbf{m}_0 - \mathbf{m}_{\text{true}} \parallel^2}, \tag{2.42}$$

where $\mathbf{m}_k$, $\mathbf{m}_{\text{true}}$ and $\mathbf{m}_0$ indicate the inverted model, true model and initial model respectively. If the model is inverted completely, RLSE ϵ approaches 0. The RLSE ϵ of the inversion result can be normalized by RLSE ϵ_0 of the initial model, which gives normalized RLSE $\tilde{\epsilon}$.

2.3.7 Comparison of computational cost

The computational cost for a traditional Gauss-Newton (TGN) method is compared with that of a source encoded Gauss-Newton method (SEGN), and that of a pseudo-Gauss-Newton (PGN) method, in Table 2.1. Note that for SEGN method, I focus on gradient construction with linear phase-encoding and in each SEGN iteration, stacking over a set of ray parameters is required to get the gradient comparable to that from shot-profile method. The PGN method is more efficient than TGN and SEGN methods for gradient calculation, because

Methods	Gradient	$\hat{H}_a(\mathbf{x},\mathbf{x}')$	$H_a(\mathbf{x},\mathbf{x}')$	$H_a^{\text{diag}}(\mathbf{x})$	Computational Cost
TGN	N_m or $2N_s$	N_m	N_s+N_g	\	N_m or $(2N_s+N_g)$
SEGN	$2N_p^{\mathbf{g}}$	\	\	$N_{ps}^{\mathbf{H}}+N_{pg}^{\mathbf{H}}$	$\left(2N_p^{\mathbf{g}}+N_{ps}^{\mathbf{H}}+N_{pg}^{\mathbf{H}}\right)$
PGN	2	\	\	$N_{ps}^{\mathbf{H}}+N_{pg}^{\mathbf{H}}$	$N_{ps}^{\mathbf{H}}+N_{pg}^{\mathbf{H}}+2$

Table 2.1: Computational cost comparison for different strategies.

it only requires 2 simulations to calculate the gradient in one iteration. The number of simulations needed to construct the gradient in TGN and SEGN methods are $2N_s$ and $2N_p^{\mathbf{g}}$ respectively, where N_s and $N_p^{\mathbf{g}}$ are the number of sources and ray parameters.

The approximate Hessian within a Gauss-Newton framework consists of two first-order partial derivative wavefields. In this research, it is named as original approximate Hessian $\hat{H}_a(\mathbf{x},\mathbf{x}')$ in Table 2.1. To calculate the partial derivative wavefields directly, N_m simulations are needed, where N_m is the number of model parameters. The gradient can also be obtained directly using equation (2.13). Thus, for the TGN method, N_m simulations are needed for each iteration, which is extremely expensive. When constructing the non-encoded approximate Hessian $H_a(\mathbf{x},\mathbf{x}')$ in equation (2.30), number of N_s+N_g simulations considering reciprocity is needed. In this situation, for each TGN step, $2N_s+N_g$ simulations are needed when storing the source-side Green's functions in gradient calculation for approximate Hessian construction. Preconditioning the gradient using diagonal approximate Hessian in TGN method requires the same computational cost with the latter case. In SEGN and PGN methods, only $N_{ps}^{\mathbf{H}}+N_{pg}^{\mathbf{H}}$ simulations are needed to calculate the diagonal part of the phase-encoded Hessian, where $N_{ps}^{\mathbf{H}}$ and $N_{pg}^{\mathbf{H}}$ are the numbers of source-side and receiver-side ray parameters respectively. For SEGN method, the phase-encoded Green's functions can also be stored for diagonal Hessian construction if $N_{ps}^{\mathbf{H}}=N_p^{\mathbf{g}}$.

2.4 Numerical Experiments

In this section, I first illustrate with numerical examples the possibility of constructing diagonal Hessian using phase-encoding technique. Then the proposed strategies on a modified Marmousi model are applied to verify the slant update strategy. Computational cost of

different encoding schemes is compared. The influence of source spacing and different preconditioning strategies are analyzed. Numerical examples with proposed phase-encoding strategy in frequency domain are also illustrated.

2.4.1 Constructing Hessian approximations with phase-encoding technique

In this section, the possibility of constructing the approximate Hessian and its diagonal components using phase-encoding method is examined. A homogeneous model with a constant velocity of 2500 m/s is used for Hessian calculation. It consists of $50 \times 50 = 2500$ grid cells with 5 m grid interval in both horizontal and vertical dimensions. The source function is a Ricker wavelet with a 25 Hz dominant frequency. Nine sources are arranged from 25 m to 255 m with a source interval of 25 m departing from the top surface 5 m. A number of 27 receivers with an interval of 25 m along the left, right and bottom boundaries of the model are deployed.

Figures 2.2a and b illustrate the non-encoded approximate Hessian (equation (2.30)) and its diagonal part (equation (2.31)) constructed by the shot-profile method. First, it is noticed that the approximate Hessian is diagonally dominant. The high-amplitude points in Figure 2.2b correspond to sources and receivers' locations, which was also discussed by Virieux and Operto (2009). Figure 2.2c shows the receiver-side linear phase-encoded Hessian contaminated by crosstalk artifacts (equation (2.33)). It can be observed that its energy is concentrated in the center part of the matrix. Figure 2.2d is the receiver-side linear phase-encoded approximate Hessian (equation (2.33)) with ray parameter range [-0.3 s/km, 0.3 s/km] and spacing $\Delta p = 0.05$ s/km. Comparison between Figures 4a and d shows that the receiver-side linear phase-encoded Hessian matches closely the non-encoded one. So, the phase-encoding method can reduce the crosstalk noise and re-distribute the energy along the diagonal band of the matrix.

I next turn to the question of to what extent the phase-encoding method can disperse crosstalk artifacts in the diagonal Hessian. The velocity model is 2 km in width and 1 km in depth with a constant velocity of 2500 m/s. A single source is placed at (1 km, 0 km) and four receivers are deployed on both sides of the source. It is shown that the diagonal part of the non-encoded approximate Hessian (equation (2.31)) arising from this model in

Figure 2.3a. Figure 2.3b shows the diagonal Hessian constructed using single receiver-side ray parameter $p = 0$ s/km and it is contaminated by the crosstalk artifacts, as indicated by the vertical stripes. Figure 7c illustrates the diagonal receiver-side phase-encoded Hessian (equation (2.39)), constructed using 7 simulations with ray parameter range [-0.3 s/km, 0.3 s/km] and spacing $\Delta p = 0.1$ s/km. Figure 2.3d illustrates the diagonal receiver-side phase-encoded Hessian constructed using 13 simulations with ray parameter range [-0.3 s/km, 0.3 s/km] and spacing $\Delta p = 0.05$ s/km. Therefore it is concluded that with increasing the number of simulations, the crosstalk artifacts can be dispersed effectively.

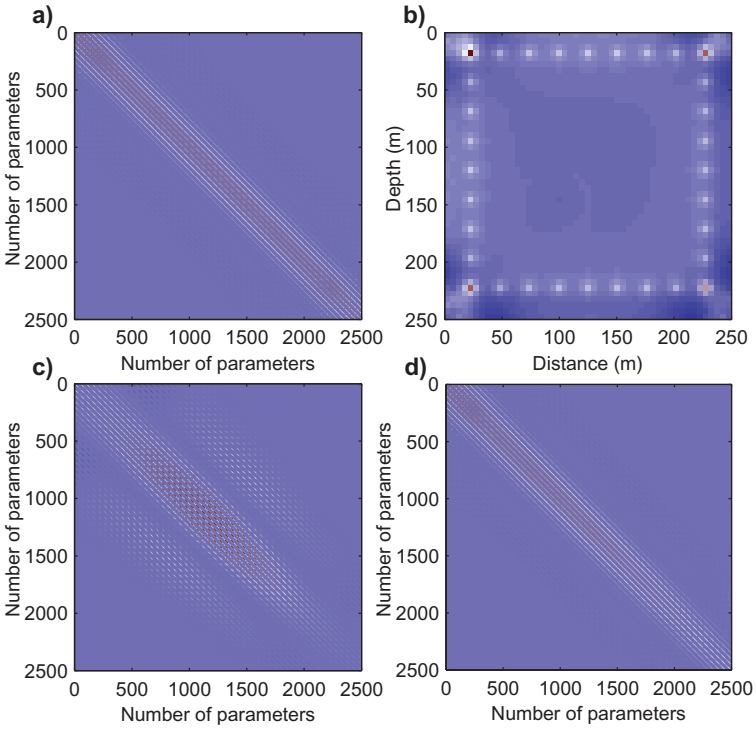

Figure 2.2: (a) shows the non-encoded approximate Hessian (equation (2.30)) corresponding to the full survey; (b) shows the diagonal part of the non-encoded approximate Hessian (equation (2.31)); (c) shows the receiver-side linear phase-encoded Hessian contaminated by crosstalk artifacts (equation (2.33)); The receiver-side Green's functions are constructed with $p = 0$ s/km. (d) shows the receiver-side linear phase-encoded approximate Hessian (equation (2.33)).

2.4.2 Applying τ-p domain FWI on modified Marmousi model

In this numerical section, I consider inversion for one portion of the Marmousi model using τ-p domain FWI. The Marmousi model is modified by introducing a water layer with a

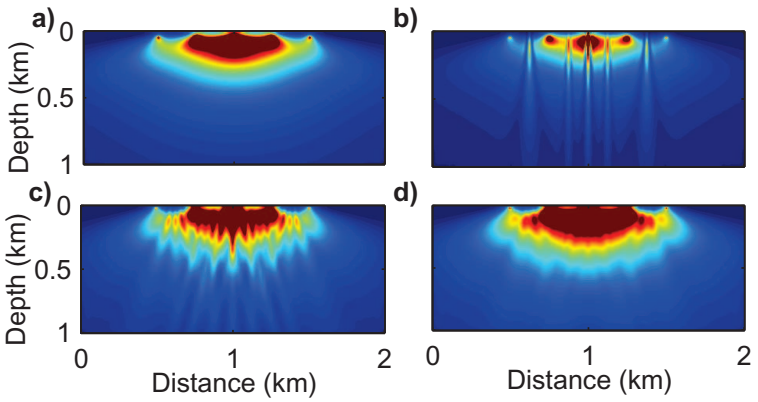

Figure 2.3: Phase-encoded diagonal Hessian. (a) is the non-encoded diagonal Hessian (equation (2.31)); (b) is the phase-encoded diagonal Hessian contaminated by crosstalk noise; (c) is the phase-encoded diagonal Hessian (equation (2.33)) using 7 simulations; (d) is the phase-encoded diagonal Hessian (equation (2.33)) using 13 simulations. The values have been normalized and amplitudes decrease from red to white and blue.

thickness of 70 m and a velocity of 1500 m/s. The model subset has 180×767 grid cells with a grid interval of 5 m in horizontal and vertical dimensions. A number of 380 sources from 10 m to 3800 m on the surface are distributed with a source interval of 10 m. A total of 765 receivers are deployed from 5 m to 3825 m on the surface, with a receiver interval of 5 m. The source function is a Ricker wavelet with a dominant frequency of 30 Hz. The ray parameter range used for linear phase-encoding method is [−0.3 s/km, 0.3 s/km] with $\Delta p = 0.1$ s/km. The multi-scale approach is implemented starting with a frequency band of [1 Hz, 5 Hz] and the frequency band increases by 2 Hz every 10 iterations. The parameter ε for chirp phase-encoding in the following examples is 0.05. Figure 2.4a and Figure 2.4b show the exact P-wave velocity model and initial velocity model, which is obtained by smoothing the true model using a Gaussian function.

2.4.2.1 Variation of the ray parameter

To verify the effectiveness of the slant update strategy, the FWI results are illustrated with PGN method after 50 iterations, calculated using fixed ray parameter and varied ray parameter during each iteration in Figure 2.5. Figures 2.5a, b and c show the inversion results when updating the model using slant gradient (equation (2.19)) with the ray parameter p fixed at 0 s/km, −0.2 s/km and 0.2 s/km respectively. The source spacing Δx_s used for these experiments is 10 m.

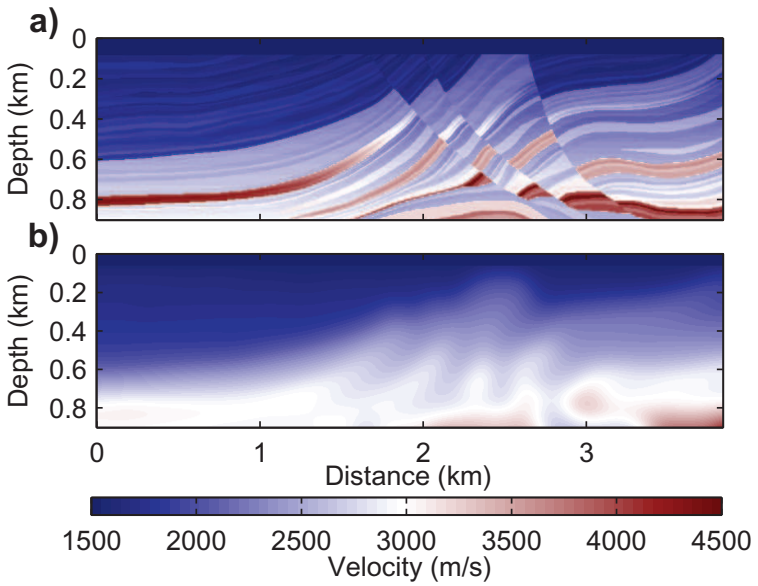

Figure 2.4: The modified Marmousi model. (a) is true P-wave velocity model; (b) is the initial P-wave velocity model.

For the extremely sparse p sampling, the inversion results suffer from aliasing due to limited wavenumber coverage in the model space and crosstalk noise according to equation (2.20). As seen that, the inversion results are biased and contaminated with noise or anomalies. For dense source arrangement ($\Delta x_s = 10$ m), the encoded wavefields approach plane wavefields and limited amount of crosstalk noise shows up (Liu et al., 2006). So, the artifacts in Figures 2.5a, b and c mainly result from inappropriate sampling of the wavenumber spectrum of subsurface. In Figure 2.5d, the inversion result is given when varying p values from -0.3 s/km to 0.3 s/km with an interval of 0.1 s/km for every 7 iterations. It can be seen that the inversion result becomes much better for denser p sampling and the noise has been mitigated as iteration proceeds. Figure 2.6 shows the RLSE calculated using equation (2.42) for different ray parameter arrangements. The values have been normalized by the RLSE of the initial model. The slant update strategy can provide a better quality of the inversion result, as indicated by the black-solid line in Figure 2.6.

Next it is shown that an appropriate ray parameter range for PGN method should be determined to balance the updates and guarantee the convergence rate. Because the ray parameter is controlled by the take-off angle and top surface velocity, different ray parameters are responsible for illuminating or updating subsurface layers with different dip angles (Wang

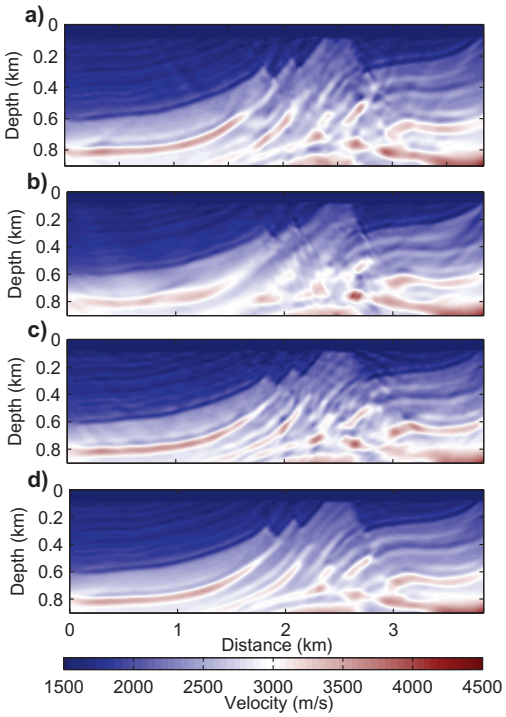

Figure 2.5: PGN FWI results for different ray parameter settings. (a) the ray parameter is fixed at $p = 0$ s/km; (b) the ray parameter is fixed at $p = -0.2$ s/km; (c) the ray parameter is fixed at $p = 0.2$ s/km; (d) the ray parameter varies from -0.3 s/km to 0.3 s/km with an interval of 0.1 s/km for every 7 iterations.

et al., 2006). Hence, the maximum p values should be determined to update steep dip layers properly (Vigh and Starr, 2008). Figure 2.7 shows the inversion result when the ray parameter range is $[-0.1$ s/km, 0.1 s/km$]$ with $\Delta p = 0.05$ s/km. The preconditioning and multi-scale strategies applied are the same as those of the experiment for the inversion result in Figure 2.5d. Comparing Figure 2.7 with Figure 2.5d, it is seen that if the ray parameter range becomes smaller, the bandwidths of the scattering angles and subsurface wavenumber are also narrowed and the subsurface layers with steep dip angles cannot be recovered very well, as indicated by the dashed cycle in Figure 2.7 and the model updating is imbalanced. The normalized RLSE $\tilde{\epsilon}$ for the inverted model is 0.8623.

2.4.2.2 Sensitivity to source spacing

Next the influence of source spacing on τ-p domain FWI with slant update strategy is considered. From equation (2.27), it is known that source spacing Δx_s should be small enough to avoid aliasing. The ray parameter range used in this experiment is 2×0.3 s/km

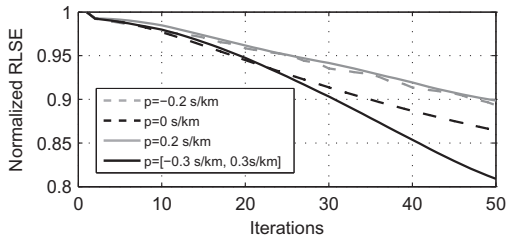

Figure 2.6: Relative least-squares errors comparison for different ray parameter settings. The gray-dash line, black-dash line and gray-solid line indicate the normalized least-squares errors when the ray parameter p is fixed at $p = -0.2$ s/km, $p = 0$ s/km and $p = 0.2$ s/km respectively. The black-solid line indicates the least-squares errors with varied ray parameters.

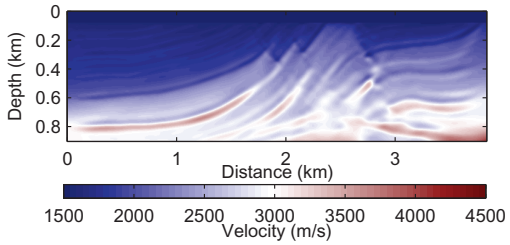

Figure 2.7: The inversion result when the ray parameter range is [−0.1 s/km, 0.1 s/km] with a spacing of 0.05 s/km.

= 0.6 s/km and the maximum frequency used for 50 iterations is 13 Hz. To avoid aliasing, the source spacing should $\Delta x_s < 111$ m. So, the inversion results are compared when source spacing $\Delta x_s = 100$ m, 50 m and 10 m respectively.

The inversion results for the source spacings Δx_s of 100 m, 50 m and 10 m are illustrated in Figures 2.8a, c and Figure 2.5d respectively. For the inversion results in Figures 2.8a and c, the ray parameter range is [−0.3 s/km, −0.3 s/km] with $\Delta p = 0.1$ s/km. In Figure 2.8a, when the sources are sparsely distributed ($\Delta x_s = 100$ m), the inversion result is seen to be seriously contaminated by crosstalk artifacts, especially in its reconstruction of shallow layers. Its normalized RLSE $\tilde{\epsilon} > 1$, which indicates non-valuable inversion result. By decreasing source spacing, as shown in Figure 2.8c (normalized $\tilde{\epsilon} = 0.8803$) and Figure 2.5d (normalized $\tilde{\epsilon} = 0.8107$), the noise becomes weaker and the convergence rate also increases. The well logs at 0.5 km and 3.0 km for different source spacings are compared in Figures 2.9a and b. By decreasing the source spacing, the inversion result more closely approximates the true velocity model, especially for the deep layers. According to equation (2.28), to mitigate the

crosstalk noise shown in Figure 2.8a, more ray parameters $\tilde{N}_p$ for stacking at each iteration are needed. Figure 2.8b is the inversion result (normalized $\tilde{\epsilon} = 0.9016$) when the source spacing is also 100 m, but wherein a stacking over ray parameter from -0.2 s/km to 0.2 s/km with a step of 0.1 s/km in each iteration ($\tilde{N}_p = 5$) is applied. Compared to Figure 2.8a, the artifacts are suppressed and inversion result is de-blurred to some extent. It is concluded that τ-p domain FWI with slant update strategy is sensitive to source spacing. For sparsely sampled experiments, stacking densely sampled ray parameters should be performed at each iteration to create a high quality inversion result. This will incur greater computational cost, and should be considered part of the cost/inversion quality trade-off.

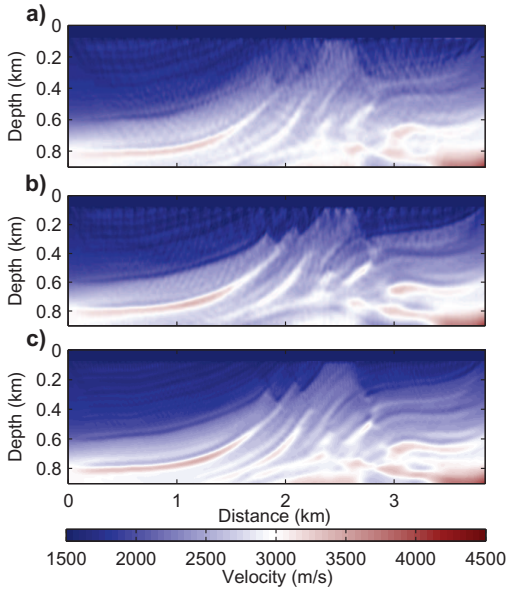

Figure 2.8: Sensitivity of the FWI results to the source spacing. (a) $\Delta x_s = 100$ m; (b) $\Delta x_s = 100$ m with stacking over ray parameters from -0.2 s/km to 0.2 s/km with a spacing of 0.1 s/km at each FWI iteration; (c) $\Delta x_s = 50$ m.

2.4.2.3 Comparison of different preconditioning methods

Next the effects of several preconditioning methods are examined. In Figure 2.10, diagonal Hessian approximations constructed using the initial velocity model are illustrated. Figures 2.10a and b illustrate the diagonal approximate Hessian and diagonal pseudo-Hessian with shot-profile method using equations (2.31) and (2.32), respectively. The diagonal pseudo-Hessian overestimates the illumination energy. Figures 2.10c and d illustrate the diagonal parts of receiver-side linear and chirp phase-encoded Hessian, constructed with 13 simula-

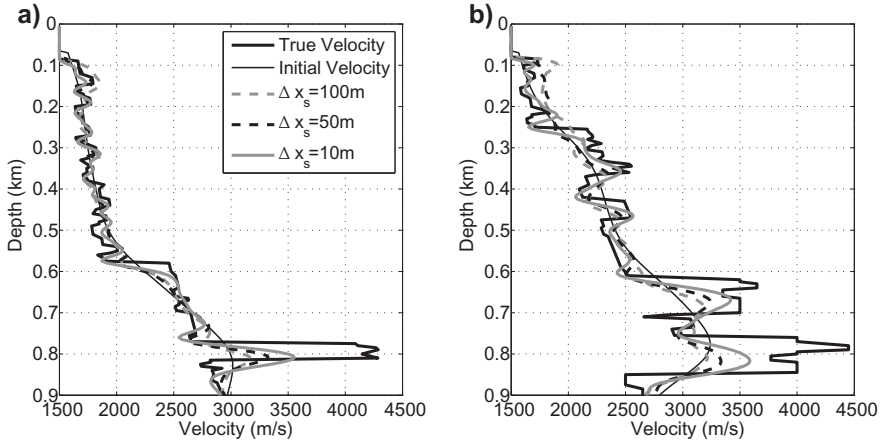

Figure 2.9: Well data comparison at 0.5 km (a) and 3 km (b) for different source spacings. The black-bold-solid and black-thin-solid lines indicate the true velocity and initial velocity respectively. The gray-bold-dash lines, black-bold-dash lines and gray-bold-solid lines indicate the inversion results when source spacing Δx_s is 100 m, 50 m and 10 m respectively.

tions. The diagonal receiver-side linear and chirp phase-encoded Hessian can be obtained from equations (2.33) and (2.37) ($\varepsilon = 0.05$) when $\mathbf{x} = \mathbf{x}'$. The receiver-side ray parameter change is [−0.3 s/km, 0.3 s/km] with $\Delta p = 0.05$ s/km. In Figure 2.10c, the vertical stripes are still be observed slightly and moreover, the effects caused by the heterogeneity of the initial velocity model also remain. While in Figure 2.10d, these effects have been reduced and the vertical stripes are dispersed. So, the chirp phase-encoding method reduces the crosstalk artifacts more efficiently than the linear phase-encoding method, at the same computational cost. In Figure 2.10e, the diagonal part of the receiver-side chirp phase-encoded Hessian with 61 simulations, which approaches the non-encoded diagonal Hessian, is illustrated. The receiver-side ray parameter range is [−0.3 s/km, 0.3 s/km] with $\Delta p = 0.01$ s/km. A τ-p domain full-waveform inversion is then carried out for 200 iterations using the steepest-descent method, the diagonal pseudo-Hessian preconditioning method, and the diagonal phase-encoded approximate Hessian preconditioning method (PGN). The results are illustrated in Figure 2.11a, b and c respectively. The frequency band increases from [1 Hz, 5 Hz] to [1 Hz, 43 Hz] by 2 Hz every 10 iterations for 200 iterations. The deeper parts of the steepest-descent inversion result are not recovered well. In comparison, the inversion results generated using the diagonal pseudo-Hessian, and the diagonal phase-encoded Hessian preconditioning are both significantly improved. The well logs from these reconstructed

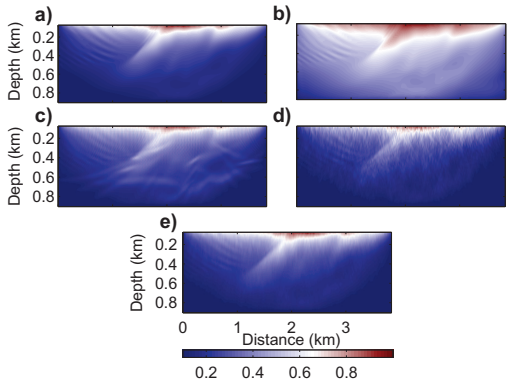

Figure 2.10: Phase-encoded diagonal Hessian comparison using the initial velocity model. (a) is the non-encoded diagonal Hessian (equation (2.31)); (b) is the non-encoded diagonal pseudo-Hessian (equation (2.32)); (c) and (d) are the receiver-side linear and chirp phase-encoded diagonal Hessian with 13 simulations; (e) shows the receiver-side chirp phase-encoded diagonal Hessian with 61 simulations.

models at 0.5 km and 3 km are extracted for comparison, as illustrated in Figures 2.12a and b respectively. The inversion result obtained using the PGN method (the gray-bold-solid lines) matches the true velocity model more accurately than those generating using other two methods. This is particularly noticeable in the deeper parts of the inversion results.

2.4.3 Examining phase-encoding strategies in frequency-ray parameter domain

In this section, numerical examples of FWI with proposed phase-encoding strategies in frequency-ray parameter domain are given. The effectiveness of proposed strategies in reconstructing the velocity model and the importance of partial overlap-frequency strategy are examined. Furthermore, inverted models with random slant update (SU) and sequential slant update (SU) strategies are given in comparison with traditional shot-profile method.

The Marmousi-II model has 244×681 grid cells with a grid interval of 10 m in both horizontal and vertical directions. I deploy 67 sources from 100 m to 6700 m with a source interval of 100 m and a depth of 20 m. A total of 681 receivers are distributed from 10 m to 6810 m with a receiver interval of 10 m and a depth of 20 m. A Ricker wavelet with a 30 Hz dominant frequency is used as the source function. Figures 2.13a and 2.13b show the true Marmousi-II P-wave velocity model and initial P-wave velocity model respectively. The initial velocity model is obtained by smoothing the true model with a Gaussian function.

Two different frequency selection strategies are used for comparison. In the mono-

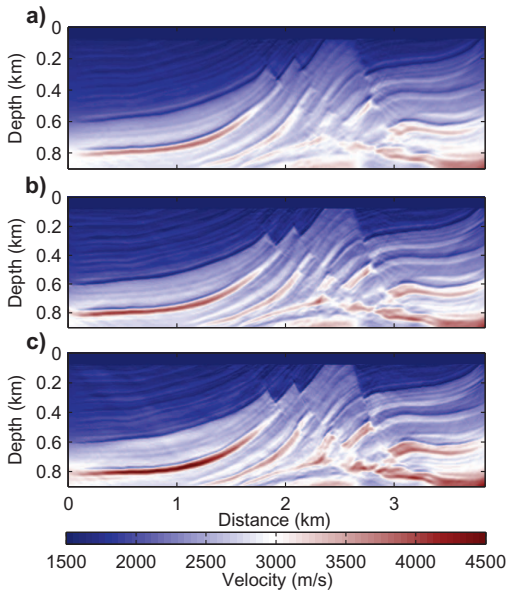

Figure 2.11: FWI results after 200 iterations for different preconditioning methods. (a) steepest-descent method; (b) Phase-encoded diagonal pseudo-Hessian preconditioning; (c) Phase-encoded diagonal approximate Hessian preconditioning.

frequency strategy, the frequency is increased sequentially from 1 Hz to 36 Hz by 1 Hz every 11 iterations. In the partial overlap-frequency selection strategy, a group of 3 frequencies are used for inversion simultaneously and the frequency band expands every 11 iterations with overlapping 2 frequencies. For the slant update strategy, two different ray parameter selections are compared. The random slant update is implemented by randomly selecting the ray parameter in the range of [-0.5 s/km, 0.5 s/km] at each iteration. While for the sequential slant update, the ray parameter is sequentially changed from -0.5 s/km to 0.5 s/km with an interval $\Delta p = 0.1$ s/km for each frequency group.

First, the shot-profile gradient is compared with the slant gradient and phase-encoded gradient. Figure 2.14a shows the shot-profile gradient with $f = 10$ Hz. Figures 2.14b, 2.14c, and 2.14d show the slant gradients with ray parameter $p = 0$ s/km, -0.1 s/km and -0.3 s/km respectively. Figures 2.14e and 2.14f show the phase-encoded gradients with $N_p = 7$ and $N_p = 11$. Varying the ray parameters for different iterations with the slant update strategy can provide balanced model updates iteratively. Phase-encoded gradients give better model updates but need higher computational cost. Furthermore, it is noticed that with stacking more ray parameters, the phase-encoded gradient (Figure 2.14f) approaches the shot-profile

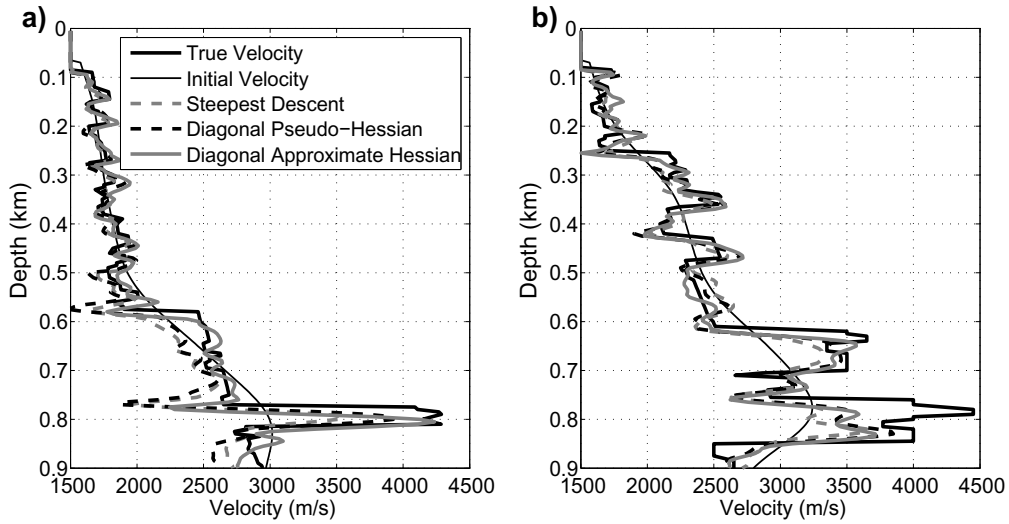

Figure 2.12: Well data comparison at 0.5 km (a) and 3 km (b) for different preconditioning methods. The black-bold-solid and black-thin-solid lines indicate the true velocity and initial velocity respectively. The gray-bold-dash lines, black-bold-dash lines and gray-bold–solid lines indicate the inversion results of steepest-descent method, phase-encoded diagonal pseudo-Hessian preconditioning method and phase-encoded diagonal approximate Hessian preconditioning method.

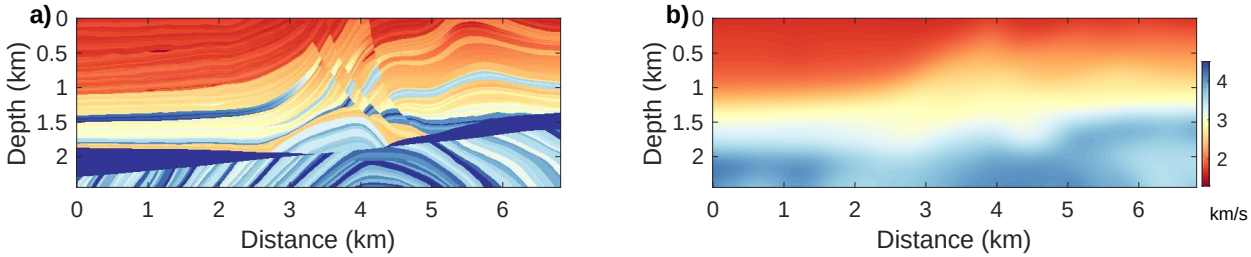

Figure 2.13: (a) the true Marmousi-II P-wave velocity model; (b) the initial P-wave velocity model.

gradient (Figure 2.14e) better.

The slant update strategy with fixed ray parameters $p = -0.1$ s/km and $p = -0.3$ s/km is then practiced, as shown in Figures 2.15a and 2.15b respectively. The inversion results with a single fixed ray parameter for all iterations are contaminated by artifacts seriously. The main geological structures of the model are obscured. This is because one ray parameter provides insufficient illumination given subsurface layers with varying dip angles. Figures 2.15c and 2.15e show the inversion results obtained using the random and sequential slant update mono-frequency strategies. The inversion results become much better in comparison with Figures 2.15a and 2.15b, however, they are still contaminated by strong crosstalk artifacts.

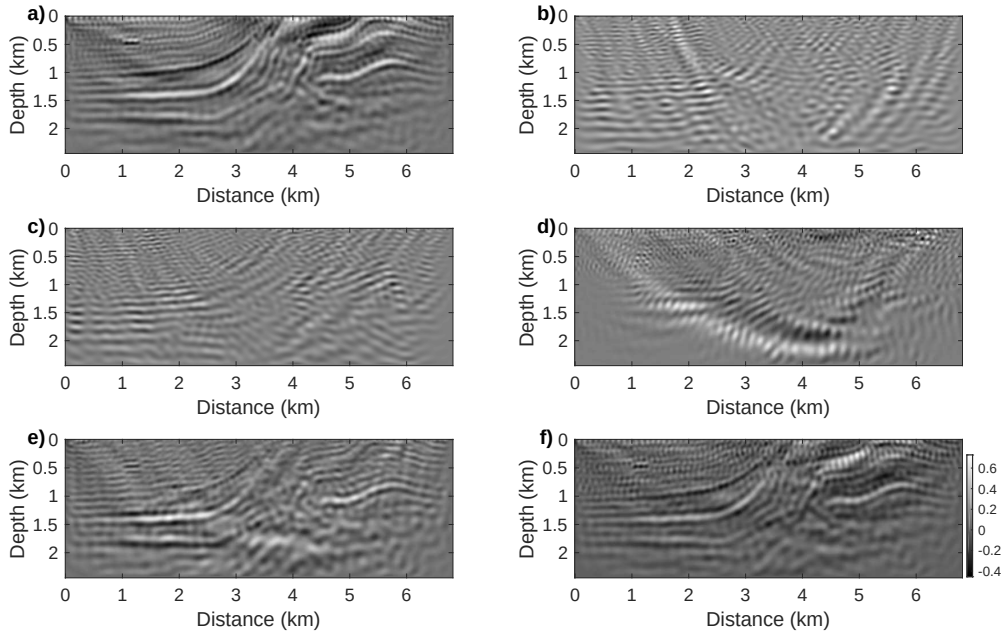

Figure 2.14: (a) Shot-profile gradient; (b) Slant gradient with $p = 0$ s/km; (c) Slant gradient with $p = -0.1$ s/km; (d) Slant gradient with $p = -0.3$ s/km; (e) Phase-encoded gradient with $p = [-0.3$ s/km, 0.3 s/km$]$ and $\Delta p = 0.1$ s/km ($N_p = 7$); (f) Phase-encoded gradient with $p = [-0.5$ s/km, 0.5 s/km$]$ and $\Delta p = 0.1$ s/km ($N_p = 11$).

Next the partial overlap-frequency strategy is considered for inversion. Figures 2.15d and 2.15f show the inversion results by random and sequential SU with partial overlap-frequency strategy. Compared to Figures 2.15c and 2.15e, the artifacts are effectively suppressed. This demonstrates the importance of inverting multiple frequencies simultaneously when applying a slant update strategy. Furthermore, the sequential SU strategy appears to provide a better inversion result than random SU strategy.

Next the inversion results obtained using traditional linear phase-encoding (TLPE) and shot-profile (SP) methods are compared, considering in particular the computational efficiency of the slant update strategy. The TLPE method is examined with two ray parameter settings. First, the ray parameter is varied from -0.3 s/km to 0.3 s/km with an interval of 0.1 s/km ($N_p = 7$). Second, the ray parameter is varied from -0.5 s/km to 0.5 s/km with the same ray parameter interval ($N_p = 11$).

Figures 2.16a and 2.16c show the inverted velocity models obtained using the mono-frequency TLPE method with $N_p = 7$ and $N_p = 11$ respectively. Comparing with Figures 2.15c and 2.15e, the stacking of ray parameters has led to improved results at each iter-

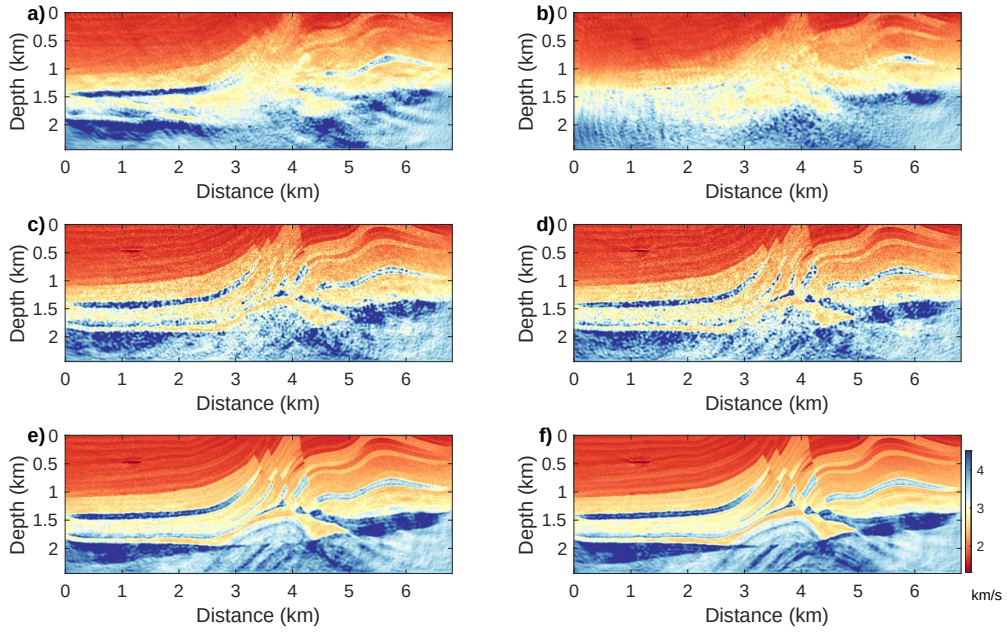

Figure 2.15: (a) Mono-frequency slant update with $p = -0.1$ s/km; (b) Mono-frequency slant update with $p = -0.3$ s/km; (c) Mono-frequency random slant update; (d) Partial overlap-frequency random slant update ($\tilde{\epsilon} = 0.4862$); (e) Mono-frequency sequential slant update; (f) Partial overlap-frequency sequential slant update ($\tilde{\epsilon} = 0.4596$).

ation with mono-frequency strategy. The SP method provides the best result, as shown in Figure 2.16e, but at the cost of extensive computation. Figures 2.16b, 2.16d and 2.16f are the inverted models using TLPE ($N_p = 7$), TLPE ($N_p = 11$) and SP methods with partial overlap-frequency strategy. An improvement in those obtained with mono-frequency strategy. It is noticed that with partial overlap-frequency strategy, SU methods can obtain results (Figures 2.15d and 2.15f) comparable to those by TLPE and SP methods (Figures 2.16b, 2.16d and 2.16f).

The RLSE vs. iterations for different encoding methods are plotted, using the mono-frequency strategy and partial overlap-frequency strategy in Figures 2.17a and 2.17b respectively. In Figure 2.17a, it is seen that the random and sequential SU methods fail to converge. The SU methods with partial overlap-frequency strategy converge efficiently, as shown in Figure 2.17b. Figure 2.18 show the RLSE vs. number of forward modelling problems solved for different encoding methods with partial overlap-frequency strategy. Figure 2.18b is the enlarged view of Figure 2.18a. In obtaining comparable inversion results, sequential SU strategy reconstructs the velocity model most efficiently, compared to TLPE

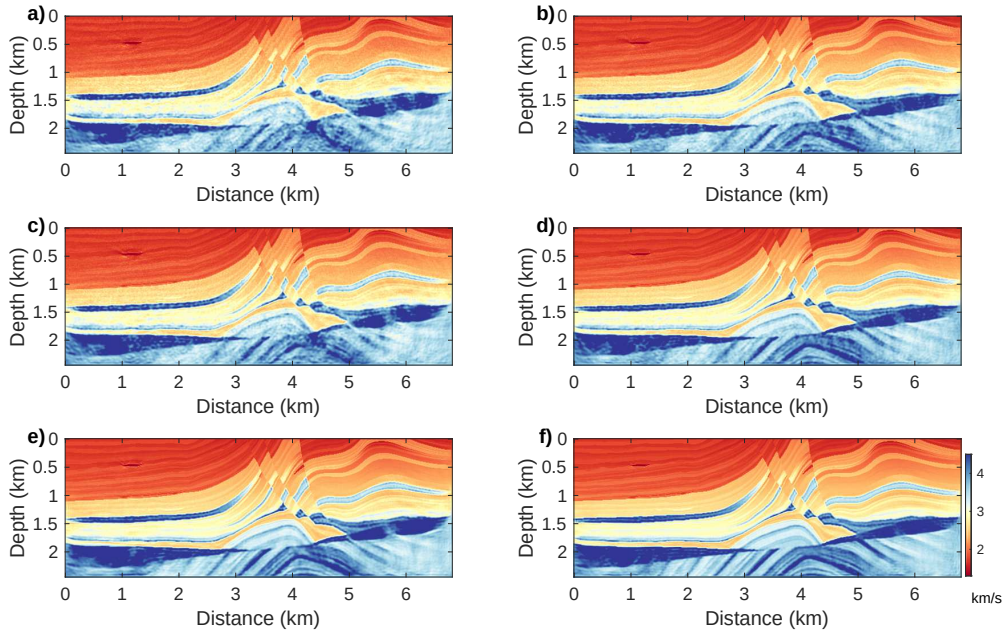

Figure 2.16: (a) Mono-frequency TLPE with $N_p = 7$ ($\tilde{\epsilon} = 0.5305$); (b) Partial overlap-frequency TLPE with $N_p = 7$ ($\tilde{\epsilon} = 0.4251$); (c) Mono-frequency TLPE with $N_p = 11$ ($\tilde{\epsilon} = 0.4472$); (d) Partial overlap-frequency TLPE with $N_p = 11$ ($\tilde{\epsilon} = 0.3598$); (e) Mono-frequency SP ($\tilde{\epsilon} = 0.3777$); (f) Partial overlap-frequency SP ($\tilde{\epsilon} = 0.3064$).

and SP methods.

2.4.3.1 Sensitivity to random noise

Another major challenge of FWI with phase-encoding technique is the sensitivity of the method to noisy data (van Leeuwen et al., 2011). To test the robustness of the proposed strategies to random noise, Gaussian noise is added to the seismic data set with SNR = 3 and SNR = 5. SNR means signal to noise ratio and smaller SNR indicates stronger random noise.

The models reconstructed using the partial overlap-frequency SP with SNR = 3 and SNR = 5 are presented in Figures 2.19a and 2.19b respectively. Figure 2.20a shows the RLSE vs. iterations for the SP method with noisy data. Noise effects are not obvious in the inverted models for SP method, however the final inversion results have been affected, compared to Figure 2.16f. Figures 2.19c and 2.19d are the inverted models by TLPE ($N_p = 7$) method with SNR = 3 and SNR = 5. The models reconstructed using the SU method with SNR = 3 and SNR = 5 are shown in Figures 2.19e and 2.19f. Figure 2.20b and 2.20c show the

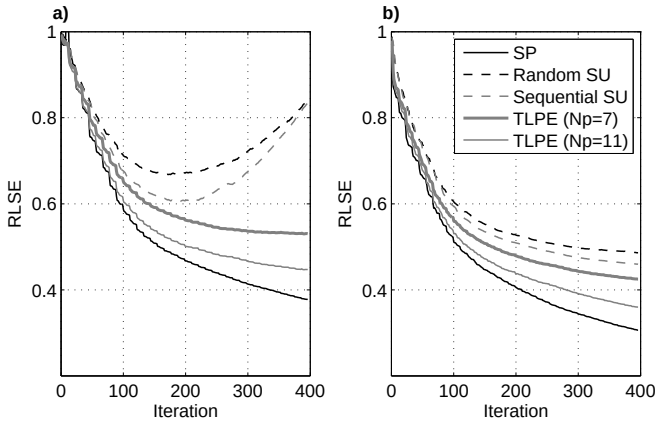

Figure 2.17: RLSE vs. Iteration for different encoding methods with mono-frequency strategy (a) and partial overlap-frequency strategy (b). The black-solid, black-dash, grey-dash, grey-bold and grey-thin lines indicate the RLSE as iteration proceeds for SP, random SU, sequential SU, TLPE ($N_p = 7$) and TLPE ($N_p = 11$) methods.

corresponding RLSE as iteration proceeds for TLPE and SU methods respectively. For TLPE and SU methods, the artifacts caused by random noise become more obvious, compared to the inversion results with noise-free data (Figure 2.16b and Figure 2.15f). At early iterations when using low frequencies for inversion, the inversion results are less sensitive to random noise, compared to iterations involving high frequencies. Furthermore, it is noticed that SU method is relatively insensitive to noisy data, especially when using low frequencies. It is concluded that the proposed slant update strategy is more robust than TLPE method ($N_p = 7$).

2.5 Discussion

The numerical experiments verify that the velocity model can be reconstructed very well by slant update strategy with single p value at each iteration. Maximum p values can be determined by the dip angles of the subsurface layers, as discussed in delayed-shot migration (Wang et al., 2006). Narrowing the ray parameter range means narrowing the bandwidths of the scattering angles and subsurface wavenumber, which results in the subsurface layers with steep dip angles not being recovered very well. The inversion result is influenced by the ray parameter interval, which can be examined in further research. This slant update strategy is also very sensitive to source spacing. For linear phase-encoding, sparse source

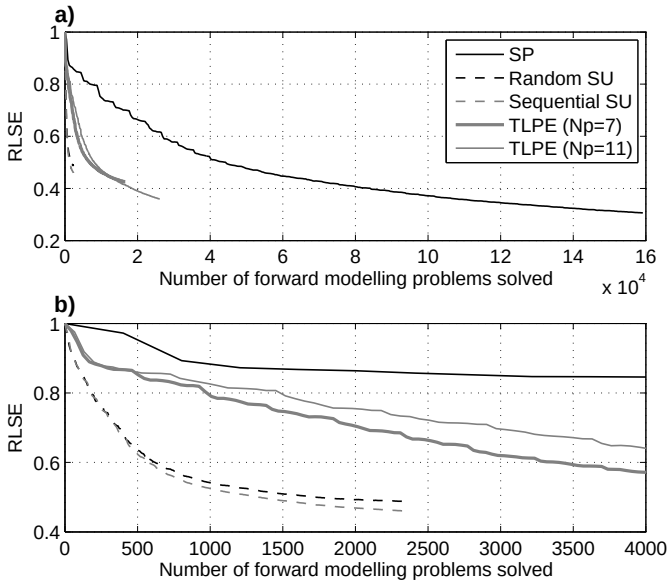

Figure 2.18: RLSE vs. Number of forward modelling problems solved. The black-solid, black-dash, grey-dash, grey-bold and grey-thin lines indicate the RLSE for SP, random SU, sequential SU, TLPE ($N_p = 7$) and TLPE ($N_p = 11$) methods. (b) Enlarged view of (a).

arrangement will result in strong crosstalk artifacts and then non-valuable inversion result, as indicated in Figure 2.8a. This is different from random phase-encoding FWI, in which crosstalk noise becomes stronger with increasing the number of encoded sources.

For diagonal Hessian construction, the proposed chirp phase-encoding can disperse the crosstalk artifacts more efficiently, compared to the linear phase-encoding. It can also be employed for constructing the second-order term in full Hessian. No regularization technique is applied for gradient and diagonal Hessian calculation and the smoothing regularization technique can be used to remove crosstalk noise further. For the numerical experiments in this research, the diagonal Hessian is calculated at each iteration and for reducing the computation further, the diagonal Hessian can be reused for a group of iterations. The advantage of diagonal approximate Hessian preconditioning method will be more obvious for transmission survey.

The initial model presented in Figure 2.4b is obtained by smoothing the true model using a Gaussian function. Hence, discussing the sensitivities of the strategies to initial model is also necessary. A 2D model is used to practice the proposed strategies. The PGN method used in this research is expected to reduce the computational burden greatly for 3D

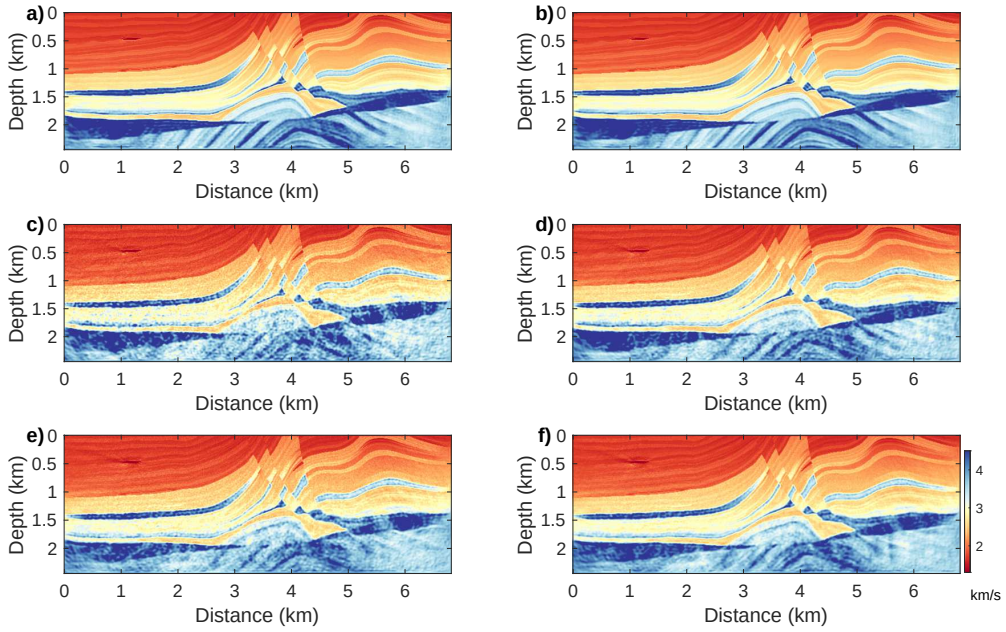

Figure 2.19: Inversion results obtained by different encoding methods with partial overlap-frequency strategy and noisy data. (a) SP method with SNR = 3 ($\tilde{\epsilon} = 0.3064$); (b) SP method with SNR = 5 ($\tilde{\epsilon} = 0.3251$); (c) TLPE ($N_p = 7$) method with SNR = 3 ($\tilde{\epsilon} = 0.5503$); (d) TLPE ($N_p = 7$) method with SNR = 5 ($\tilde{\epsilon} = 0.4675$); (e) SU method with SNR = 3 ($\tilde{\epsilon} = 0.5078$); (f) SU method with SNR = 5 ($\tilde{\epsilon} = 0.4705$).

FWI. The traditional simultaneous-source technique and the proposed slant update strategy in this research are applicable for fixed-spread acquisition (land or ocean bottom survey). Incorporating the water layer in the modified Marmousi model makes the model update operation easier for numerical modelling. For dealing with land data in real application, more complex near surface effects, such as topography and weathering layer, should be taken into consideration. For nonfixed spread acquisition (marine-streamer survey), the inconsistent acquisition geometries between the observed data and modelled data can cause strong artifacts in the inversion results. To apply the proposed strategies for marine-streamer acquisition, a correlation-based method or missing data completion method should be performed to remove the unwanted information in the data residuals (Routh et al., 2011; Choi and Alkhalifah, 2012; Son et al., 2014).

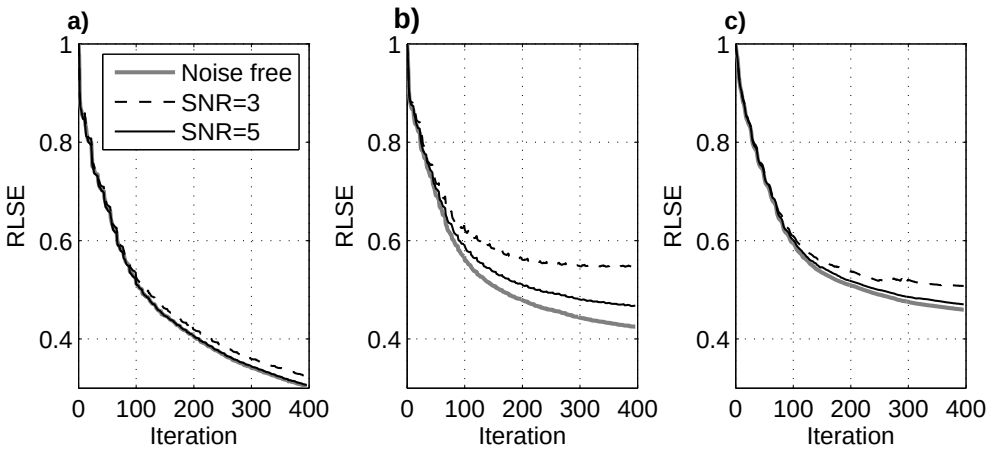

Figure 2.20: RLSE vs. Iterations for partial overlap-frequency SP (a), TLPE ($N_p = 7$) (b) and SU (c) methods with noisy data. The black, red, and blue lines indicate the RLSE for noise-free data, noisy data with SNR = 3 and SNR = 5 respectively.

2.6 CONCLUSION

In this chapter, which is aimed at addressing obstacles to the practical implementation of full-waveform inversion, I have described a range of strategies, which, assembled, I refer to as efficient τ-p domain pseudo-Gauss-Newton FWI. In this approach, the gradient is constructed using a linear phase-encoding technique, which reduces the computational expense significantly. To reduce computation further, I introduce a slant update strategy, in which the model is updated using the slant gradient with single ray parameter. By varying the ray parameter regularly, the model update can be balanced through a set of iterations. A similar quality inversion result can be obtained using this strategy with reducing the computational burden considerably. An appropriate ray parameter range and schedule must be determined to obtain an optimal convergence rate; geological information such as prevalent dips can be used to guide this choice. Our strategy is sensitive to source spacing; for sparsely sampled sources, a stack over a set of ray parameters, should be carried out at each iteration to suppress crosstalk artifacts. Finally, this chapter introduces an additional chirp phase-encoding strategy, which appears to attenuate crosstalk artifacts more efficiently than other approaches to diagonal Hessian construction. The proposed pseudo-Gauss-Newton approach appears to converge to the true velocity model faster than the steepest-descent and diagonal pseudo-Hessian preconditioning methods, which is consistent with the observations by ans

S. Treitel (1984). In frequency domain, the proposed phase-encoding approach is also able to obtained high quality inverted models with reducing the computation cost.

Appendix A

Proof of the equivalence between source illumination and diagonal pseudo-Hessian

Equivalence between source illumination and diagonal pseudo-Hessian can be proved using a non-perturbation approach. Taking partial derivative with respect to the model parameter $\mathbf{m}(\mathbf{x})$ on both sides of equation (2.2) yields:

$$\mathbf{L}\left(\mathbf{x},\omega\right)\frac{\partial\mathbf{u}\left(\mathbf{x},\mathbf{x}_s,\omega\right)}{\partial\mathbf{m}(\mathbf{x})}=-\frac{\partial\mathbf{L}\left(\mathbf{x},\omega\right)}{\partial\mathbf{m}(\mathbf{x})}\mathbf{u}\left(\mathbf{x},\mathbf{x}_s,\omega\right). \tag{A.1}$$

Equation (A.1) underlines the fact that the scattered wavefields due to the perturbation in the model parameter at position $\mathbf{x}$ can be interpreted as the wavefields propagating in the reference medium generated by a secondary body force. The right hand side of equation (A.1) is always considered as "scattered source" or "virtual source" due to the model perturbation at $\mathbf{x}$ (Shin et al., 2001a,b; Virieux and Operto, 2009):

$$\begin{aligned}\tilde{f}_s(\mathbf{x},\mathbf{x}_s,\omega)&=-\frac{\partial\mathbf{L}\left(\mathbf{x},\omega\right)}{\partial\mathbf{m}(\mathbf{x})}\mathbf{u}\left(\mathbf{x},\mathbf{x}_s,\omega\right)\\&=-\omega^2\mathbf{u}\left(\mathbf{x},\mathbf{x}_s,\omega\right).\end{aligned} \tag{A.2}$$

Recalling the wavefields solution with Green's function (equation (F.1)), the virtual source becomes:

$$\tilde{f}_s(\mathbf{x},\mathbf{x}_s,\omega)=-\omega^2 f_s\left(\omega\right)G\left(\mathbf{x},\mathbf{x}_s,\omega\right). \tag{A.3}$$

The pseudo-Hessian is defined as the correlation of two virtual sources:

$$\begin{aligned}H_{pseudo}(\mathbf{x},\mathbf{x}')&=\sum_{\mathbf{x}_s}\sum_{\omega}\tilde{f}_s(\mathbf{x},\mathbf{x}_s,\omega)\tilde{f}_s^*(\mathbf{x}',\mathbf{x}_s,\omega)\\&=\sum_{\mathbf{x}_s}\sum_{\omega}\Re\left(\omega^4|f_s(\omega)|^2G(\mathbf{x},\mathbf{x}_s,\omega)G^*(\mathbf{x}',\mathbf{x}_s,\omega)\right),\end{aligned} \tag{A.4}$$

where $\mathbf{x}'$ is the neighboring point around the imaging point $\mathbf{x}$ in subsurface (Valenciano, 2008). When $\mathbf{x}'\neq\mathbf{x}$, the off-diagonal elements of the pseudo-Hessian, the crosscorrelation of

the two virtual sources, can be obtained. When $\mathbf{x} = \mathbf{x}'$, diagonal part of the pseudo-Hessian, the autocorrelation of the two virtual sources, can be obtained:

$$H^{\text{diag}}_{pseudo}(\mathbf{x}) = \sum_{\mathbf{x}_s} \sum_{\omega} \Re \left(\omega^4 |f_s(\omega)|^2 G(\mathbf{x}, \mathbf{x}_s, \omega) G^*(\mathbf{x}, \mathbf{x}_s, \omega) \right), \tag{A.5}$$

where the factor ω^4 is equivalent to double second-order time-derivative operations in time domain and for single frequency inversion, it is of no consequence, as discussed by Sirgue and Pratt (2004). So, equation (A.5) is equivalent to the source illumination, the autocorrelation of down-going wavefields (Pan et al., 2014b). It can also be obtained by removing the two receiver-side Green's functions in equation (2.31) with the assumption of infinite receiver coverage. Tarantola (1984) found that the gradient calculation in least-squares wave equation inversion resembles the migration process with a crosscorrelation imaging condition. Here, it is seen that preconditioning the gradient using one portion of the Hessian resembles a deconvolution imaging condition (Pan et al., 2014b).

Chapter 3

Non-linear optimization methods for full-waveform inversion

3.1 Summary

The Hessian-free (HF) optimization method represents an attractive alternative to Newton-based and gradient-based optimization methods. At each iteration, the HF approach obtains the search direction by approximately solving the Newton linear system using a matrix-free conjugate-gradient (CG) algorithm. The main drawback with HF optimization is that the CG algorithm requires many iterations. In this chapter, I develop and compare different preconditioning schemes for the CG algorithm to accelerate the HF Gauss-Newton (GN) method. Traditionally, preconditioners are designed as diagonal Hessian approximations. I additionally use a new pseudo diagonal GN Hessian as a preconditioner, making use of the reciprocal property of Greens function. Furthermore, an l-BFGS inverse Hessian preconditioning strategy with the diagonal Hessian approximations as an initial guess is developed. Several numerical examples are carried out. It is determined that the quasi-Newton l-BFGS preconditioning scheme with the pseudo diagonal GN Hessian as the initial guess is most effective in speeding up the HF GN FWI. Finally, in the case of multiparameter acoustic FWI, it is proved that the l-BFGS preconditioned HF GN method can reconstruct velocity and density models better and more efficiently in comparison with the nonpreconditioned method.

3.2 Introduction

Traditional optimization methods for FWI in exploration geophysics are gradient-based (i.e., steepest-descent (SD) and non-linear conjugate-gradient (NCG) methods). In SD method, the search direction is simply the negative of the gradient and in NCG method, the search direction is the linear combination of the current gradient and previous search direction

(Fletcher and Reeves, 1964; Nocedal and Wright, 2006; Hu et al., 2011). Within the adjoint-state method, the gradient of the misfit function can be calculated efficiently by applying a zero-lag crosscorrelation between the forward modelled wavefield and back-propagated data residual wavefield (Pratt et al., 1998b; Tromp et al., 2005). Thus, gradient-based methods are computationally attractive for large-scale inverse problems. However, they suffer from slow convergence rates.

The second-order partial derivative of the misfit function (namely Hessian operator) carries crucial information in the reconstruction process (Santosa and Symes, 1988; Pan et al., 2016). The search direction can be significantly enhanced by multiplying the gradient with the inverse Hessian matrix (Pratt et al., 1998b). Furthermore, the second-order term in the Hessian matrix which accounts for non-linear scattering effects, can help to remove the second-order scattering artifacts in the gradient (Pratt et al., 1998b; Métivier et al., 2013; Pan et al., 2015b, 2016). However, explicit calculation, storage and inversion of the Hessian at each iteration is computationally impractical for large-scale inverse problems. Hence, various approaches have been proposed for approximating the Hessian (Shin et al., 2001a; Plessix and Mulder, 2004; Tang, 2009) or inverse Hessian (Nocedal and Wright, 2006; Nammour and Symes, 2009; Demanet et al., 2012). In Gauss-Newton method, an approximate Hessian is introduced by involving only the first-order term and ignoring the second-order contributions (Pratt et al., 1998b). Diagonal Gauss-Newton Hessian and diagonal pseudo-Hessian also serve as good preconditioners, which have been discussed in chapter 2.

Instead of constructing the Hessian explicitly, quasi-Newton methods approximate the inverse Hessian iteratively by storing the model and gradient changes from previous iterations (Nocedal and Wright, 2006). One popular quasi-Newton method is the BFGS method (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970). However, the storage requirement of the inverse Hessian approximation and computation cost of preconditioning for large-scale inverse problems is still very high. To mitigate this difficulty, a limited-memory BFGS (l-BFGS) method is developed by storing information from a limited number l ($l < 10$) of previous iterations (Nocedal, 1980; Byrd et al., 1995; Nocedal and Wright, 2006). Compared to gradient-based methods, l-BFGS methods provide faster convergence rates for large-scale inverse problems (Brossier et al., 2010; Ma and Hale, 2012). The convergence performance

of l-BFGS method is closely related to the initial guess of inverse Hessian approximation (Brossier et al., 2009; Guitton and Díaz, 2012).

Hessian-free optimization methods (truncated-Newton or inexact-Newton method) represent attractive alternatives to the above-described optimization methods (Nash, 1985; Santosa and Symes, 1988; Nash, 2000; Akcelik et al., 2002; Epanomeritakis et al., 2008; Métivier et al., 2012, 2014). At each iteration, the search direction is computed by approximately solving the Newton equations using a matrix-free scheme of the conjugate-gradient (CG) algorithm, which is an optimal method for solving a positive definite system (Nash, 1985; Hu et al., 2009). This linear iterative solver only requires the Hessian-vector products instead of forming the Hessian operator explicitly (Métivier et al., 2014). In this chapter, the full Hessian is replaced with the Gauss-Newton Hessian, which is always symmetric and positive semi-definite. One issue of the HF optimization method is that obtaining the search direction approximately requires a large number of CG iterations, which is still very expensive. Our main goal in this chapter is to precondition the CG algorithm that accelerates the HF Gauss-Newton full-waveform inversion (Nash, 2000; Sainath et al., 2013).

Preconditioning makes the CG problem well-conditioned, hence easier to solve, and it reduces the number of CG iterations. The preconditioner for the CG algorithm is designed by approximating the Hessian or its inverse (Nash, 2000). Different preconditioning schemes are developed for comparison in this chapter. Traditional Hessian approximations, diagonal pseudo-Hessian and diagonal Gauss-Newton Hessian, are first considered as preconditioners for the CG algorithm. Based on the reciprocal property of the Green's function, a pseudo diagonal Gauss-Newton Hessian approximation is used as the preconditioner for the CG solver (Plessix and Mulder, 2004). Quasi-Newton l-BFGS inverse Hessian approximations also serve as effective preconditioners for CG iterative solver (Nash, 1985; Métivier et al., 2013). The initial guess of the inverse Hessian approximation is important to the performance of the l-BFGS method. In this chapter, I propose to construct the l-BFGS inverse Hessian preconditioners using the diagonal Hessian approximations as initial guess. It is demonstrated that the proposed l-BFGS inverse Hessian preconditioning strategy with pseudo diagonal Gauss-Newton Hessian as initial guess works best in accelerating the Hessian-free Gauss-Newton FWI among these preconditioning strategies. Numerical examples are given to examine the

sensitivity of this proposed preconditioning strategy to noisy data.

Inverting for multiple physical parameters simultaneously is a more challenging task compared to monoparameter FWI. The coupling effects between different physical parameters (parameter crosstalk) make the inverse problem less well-determined (Operto et al., 2013; Innanen, 2014b; Métivier et al., 2015). HF Gauss-Newton FWI is expected to mitigate parameter crosstalk for large-scale multiparameter FWI. In this chapter, the velocity and density in acoustic media are reconstructed simultaneously with the preconditioned HF Gauss-Newton FWI. It is observed that with the proposed l-BFGS preconditioning strategy, HF Gauss-Newton method inverts the velocity and density models better than other non-linear optimization methods.

The chapter is organized as follows. First, different non-linear optimization methods for full-waveform inversion are presented and the CG-based Hessian-free optimization method is introduced in detail. Then, different preconditioning schemes for the inner CG iteration in HF method are discussed. In the numerical modelling section, the performances of different non-linear optimizations for FWI are compared and different preconditioning strategies for HF Gauss-Newton method with simple Gaussian-anomaly model and more complex Marmousi model are examined. Then, the performances of preconditioned HF Gauss-Newton method with noisy data are investigated. Finally, different non-linear optimization methods are applied to reconstruct velocity and density parameters simultaneously. The numerical modelling experiments demonstrate that the l-BFGS preconditioned Hessian-free Gauss-Newton method works best to reconstruct the model parameters.

3.3 Methodology

In chapter 2, the expressions for gradient (equation (2.26)) and Hessian (equation (2.30)) with Green's functions were presented. In this chapter, the non-linear least-squares inverse problem is first reviewed. The gradient is derived with adjoint-state method. Expressions of Hessian are given with matrix formulations. Then, various non-linear optimization methods for FWI are described. The preconditioning strategies for accelerating Hessian-free Gauss-Newton FWI are introduced.

3.3.1 The non-linear least-squares inverse problem: review with matrix formulation

The common l-2 norm misfit function Φ is given by:

$$\Phi\left(\mathbf{m}\right) = \frac{1}{2}\|\mathcal{P}\mathbf{u} - \mathbf{d}_{\text{obs}}\|^2, \tag{3.1}$$

which $\mathcal{P}$ is the receiver sampling operator $\mathbf{d}_{\text{syn}} = \mathcal{P}\mathbf{u}$. The gradient of the misfit function Φ is given by correlating the Fréchet derivative wavefields with the complex conjugate of the data residuals (Pratt et al., 1998b; Virieux and Operto, 2009):

$$\mathbf{g}\left(\mathbf{x}\right) = \nabla_{\mathbf{m}(\mathbf{x})}\Phi\left(\mathbf{m}\right) = \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\sum_{\omega}\Re\left(\nabla_{\mathbf{m}(\mathbf{x})}\mathbf{d}_{\text{syn}}^{\dagger}\left(\mathbf{x}_g, \mathbf{x}_s, \omega; \mathbf{m}\right)\Delta\mathbf{d}^{*}\left(\mathbf{x}_s, \mathbf{x}_g, \omega\right)\right), \tag{3.2}$$

where the Jacobian matrix $\nabla_{\mathbf{m}(\mathbf{x})}\mathbf{d}_{\text{syn}}$ can be expressed with matrices:

$$\nabla_{\mathbf{m}}\mathbf{d}_{\text{syn}} = -\mathcal{P}\mathbf{L}\left(\mathbf{m}, \omega\right)^{-1}\nabla_{\mathbf{m}}\mathbf{L}\left(\mathbf{m}, \omega\right)\mathbf{u}, \tag{3.3}$$

where $\partial\mathbf{L}\left(\mathbf{m}, \omega\right)/\partial\mathbf{m}$ is written as $\nabla_{\mathbf{m}}\mathbf{L}\left(\mathbf{m}, \omega\right)$ for sake of compactness. Inserting equation (3.3) into equation (3.2) gives the gradient as:

$$\mathbf{g} = \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\sum_{\omega}\mathbf{u}^{\dagger}\left(\mathbf{x}_s, \mathbf{x}_g, \omega\right)\nabla_{\mathbf{m}}\mathbf{L}^{\dagger}\left(\mathbf{m}, \omega\right)\mathbf{L}\left(\mathbf{m}, \omega\right)^{-1}\mathcal{P}^{\dagger}\Delta\mathbf{d}^{*}\left(\mathbf{x}_s, \mathbf{x}_g, \omega\right), \tag{3.4}$$

where $\mathbf{L}^{\dagger}\left(\mathbf{m}, \omega\right)^{-1}$ is replaced with $\mathbf{L}\left(\mathbf{m}, \omega\right)^{-1}$ when considering reciprocity of Green's function and $\mathbf{L}\left(\mathbf{m}, \omega\right)^{-1}\mathcal{P}^{\dagger}\Delta\mathbf{d}^{*}\left(\mathbf{x}_s, \mathbf{x}_g, \omega\right)$ serves as backward propagated data residual wavefields.

The gradient expressions can also be derived based on adjoint-state method. Considering that misfit function Φ (equation (4.1)) subjects to $\mathbf{L}\left(\mathbf{m}, \omega\right)\mathbf{u} = \mathbf{f}_s$, FWI can thus be formulated as a PDE-constrained inverse problem. The augmented Lagrangian functional $\boldsymbol{\chi}$ associated with this problem can be written as:

$$\boldsymbol{\chi}\left(\mathbf{m}, \mathbf{u}, \Lambda\right) = \frac{1}{2}\|\mathcal{P}\mathbf{u} - \mathbf{d}_{\text{obs}}\|^2 + \Lambda^{\dagger}\left(\mathbf{L}\left(\mathbf{m}, \omega\right)\mathbf{u} - \mathbf{f}_s\right), \tag{3.5}$$

where Λ is the Lagrangian multiplier. Perturbation of the Lagrangian functional $\boldsymbol{\chi}$ due to the perturbations of the quantities $\mathbf{m}$ and $\mathbf{u}$ is (Liu and Tromp, 2006; Métivier et al., 2013):

$$\Delta\boldsymbol{\chi}\left(\mathbf{m}, \mathbf{u}, \Lambda\right) = \mathbf{u}^{\dagger}\nabla_{\mathbf{m}}\mathbf{L}^{\dagger}\left(\mathbf{m}, \omega\right)\Lambda\Delta\mathbf{m} + \mathcal{P}^{\dagger}\left(\mathcal{P}\mathbf{u} - \mathbf{d}_{\text{obs}}\right)^{*}\Delta\mathbf{u} - \mathbf{L}^{\dagger}\left(\mathbf{m}, \omega\right)\Lambda\Delta\mathbf{u}. \tag{3.6}$$

Equation (3.6) is stationary, when the coefficient of the perturbations of wavefields is zero, which gives the following equation:

$$\mathbf{L}\left(\mathbf{m},\omega\right)^{\dagger}\Lambda=\mathcal{P}^{\dagger}\left(\mathcal{P}\mathbf{u}-\mathbf{d}_{\text{obs}}\right)^{*}. \tag{3.7}$$

Solution of equation (3.7) Λ is defined as the adjoint wavefield: $\Lambda=\left(\mathbf{L}\left(\mathbf{m},\omega\right)^{\dagger}\right)^{-1}\mathcal{P}^{\dagger}\left(\mathcal{P}\mathbf{u}-\mathbf{d}_{\text{obs}}\right)^{*}$. Taking partial derivative of the misfit function $\boldsymbol{\chi}$ with respect to model parameter $\mathbf{m}$:

$$\nabla_{\mathbf{m}}\boldsymbol{\chi}=\mathbf{u}^{\dagger}\nabla_{\mathbf{m}}\mathbf{L}^{\dagger}\left(\mathbf{m},\omega\right)\Lambda. \tag{3.8}$$

Inserting adjoint wavefield Λ into equation (3.8) gives the gradient of misfit function $\boldsymbol{\chi}$:

$$\nabla_{\mathbf{m}}\boldsymbol{\chi}=\mathbf{u}^{\dagger}\nabla_{\mathbf{m}}\mathbf{L}^{\dagger}\left(\mathbf{m},\omega\right)\left(\mathbf{L}\left(\mathbf{m},\omega\right)^{\dagger}\right)^{-1}\mathcal{P}^{\dagger}\Delta\mathbf{d}^{*}. \tag{3.9}$$

Equation (3.9) is equivalent to equation (3.4). The full Hessian represents the second derivative of misfit function:

$$H\left(\mathbf{x},\mathbf{x}'\right)=\nabla_{\mathbf{m}(\mathbf{x})}\nabla_{\mathbf{m}(\mathbf{x}')}\Phi\left(\mathbf{m}\right)=\nabla_{\mathbf{m}(\mathbf{x})}\mathbf{d}_{\text{syn}}^{\dagger}\nabla_{\mathbf{m}(\mathbf{x})}\mathbf{d}_{\text{syn}}^{*}+\nabla_{\mathbf{m}(\mathbf{x})}\nabla_{\mathbf{m}(\mathbf{x}')}\mathbf{d}_{\text{syn}}^{\dagger}\Delta\mathbf{d}^{*}, \tag{3.10}$$

where the first term is the Gauss-Newton Hessian approximation $\mathbf{H}_a$:

$$\begin{aligned}\mathbf{H}_a&=\nabla_{\mathbf{m}}\mathbf{d}_{\text{syn}}^{\dagger}\nabla_{\mathbf{m}}\mathbf{d}_{\text{syn}}^{*},\\&=\mathbf{u}^{\dagger}\nabla_{\mathbf{m}}\mathbf{L}^{\dagger}\left(\mathbf{m},\omega\right)\left(\mathbf{L}^{\dagger}\left(\mathbf{m},\omega\right)\right)^{-1}\mathcal{P}^{\dagger}\mathcal{P}\mathbf{L}^{*}\left(\mathbf{m},\omega\right)^{-1}\nabla_{\mathbf{m}}\mathbf{L}^{*}\left(\mathbf{m},\omega\right)\mathbf{u}^{*}.\end{aligned} \tag{3.11}$$

The second term of the full Hessian indicating the second-order scattering effects is formed by correlating the data residual vector $\Delta\mathbf{d}$ with the second-order partial derivative wavefields:

$$\nabla_{\mathbf{m}}\nabla_{\mathbf{m}}\mathbf{d}_{\text{syn}}=-\mathbf{L}\left(\mathbf{m},\omega\right)^{-1}\mathcal{P}\left(2\nabla_{\mathbf{m}}\mathbf{L}\left(\mathbf{m},\omega\right)\nabla_{\mathbf{m}}\mathbf{u}+\nabla_{\mathbf{m}}\nabla_{\mathbf{m}}\mathbf{L}\left(\mathbf{m},\omega\right)\mathbf{u}\right). \tag{3.12}$$

Inserting equation (3.3) into equation (3.12) gives:

$$\nabla_{\mathbf{m}}\nabla_{\mathbf{m}}\mathbf{d}_{\text{syn}}=-\mathbf{L}\left(\mathbf{m},\omega\right)^{-1}\mathcal{P}\left(2\nabla_{\mathbf{m}}\mathbf{L}\left(\mathbf{m},\omega\right)\mathbf{L}\left(\mathbf{m},\omega\right)^{-1}\nabla_{\mathbf{m}}\mathbf{L}\left(\mathbf{m},\omega\right)\mathbf{u}+\nabla_{\mathbf{m}}\nabla_{\mathbf{m}}\mathbf{L}\left(\mathbf{m},\omega\right)\mathbf{u}\right). \tag{3.13}$$

Then substituting equation (3.13) into the second term of the full Hessian (equation (3.10)):

$$\bar{\mathbf{H}}=-2\mathbf{u}^{\dagger}\nabla_{\mathbf{m}}\mathbf{L}^{\dagger}\left(\mathbf{m},\omega\right)\left(\mathbf{L}^{\dagger}\left(\mathbf{m},\omega\right)\right)^{-1}\nabla_{\mathbf{m}}\mathbf{L}^{\dagger}\left(\mathbf{m},\omega\right)\mathfrak{v}-\mathbf{u}^{\dagger}\nabla_{\mathbf{m}}\nabla_{\mathbf{m}}\mathbf{L}^{\dagger}\left(\mathbf{m},\omega\right)\mathfrak{v}, \tag{3.14}$$

where $\mathfrak{v}$ is the adjoint wavefield obtained by $\mathfrak{v}=\left(\mathbf{L}^{\dagger}\left(\mathbf{m},\omega\right)\right)^{-1}\mathcal{P}^{\dagger}\Delta\mathbf{d}^{*}$. When considering linearized inverse problem, the second-order scattered wavefields is very small and the second-order term of full Hessian is always ignored.

3.3.2 Full-Newton and Gauss-Newton methods

Newton-based optimization methods (e.g., full-Newton (FN) and Gauss-Newton (GN) methods) use the quadratic search direction and exhibit fast convergence given a limited number of unknown parameters. The FN search direction is formed by preconditioning the gradient with the full Hessian $\mathbf{H}$ (equation (3.10)):

$$\Delta \mathbf{m}_k = -\mathbf{H}_k^{-1}\mathbf{g}_k, \tag{3.15}$$

where the full Hessian $\mathbf{H}$ is actually inverted. The Gauss-Newton approximate Hessian $\mathbf{H}_a$ only accounts for the first-order scattering effects, as indicated by the first term of equation (3.10). For these Newton-based methods, explicit evaluation and inversion the Hessian matrix $\mathbf{H}$ and Gauss-Newton Hessian $\mathbf{H}_a$ at each iteration are required. Though Newton-based methods benefit from fast convergence rate, the computation, storage and inversion of Hessian at each iteration are prohibitively expensive, which limits their applications for large-scale inverse problems in exploration geophysics.

3.3.3 Gradient-based methods

Gradient-based methods (e.g., steepest-descent (SD) and non-linear conjugate-gradient (NCG) methods) approximate the Hessian matrix $\mathbf{H}$ as an identity matrix $\mathbf{I}$ and they are computationally more attractive than the Newton-based ones when inverting a large number of unknown model parameters. The SD method simply determines the search direction to be the negative of the gradient:

$$\Delta \mathbf{m}_k = -\mathbf{g}_k. \tag{3.16}$$

In mathematics, conjugate-gradient (CG) method seeks the solution of a linear system. The non-linear conjugate-gradient (NCG) method generalizes the conjugate-gradient method to non-linear optimization and obtain the local minimum of a non-linear function using its gradient alone (Nocedal and Wright, 2006). The search direction in NCG method is just a linear combination of current gradient and previous search direction:

$$\Delta \mathbf{m}_k = -\mathbf{g}_k + \beta_k \Delta \mathbf{m}_{k-1}, \tag{3.17}$$

where β_k is a scalar selected such that $\Delta\mathbf{m}_k$ and $\Delta\mathbf{m}_{k-1}$ are conjugate. There are many approaches for determining parameter β_k including "Fletcher-Reeves (FR)", "Polak-Ribière (PR)", "Hestenes-Stiefel (HS)" and "Dai-Yuan (DY)" method:

$$\beta_k^{FR} = \frac{\mathbf{g}_k^\dagger \mathbf{g}_k}{\mathbf{g}_{k-1}^\dagger \mathbf{g}_{k-1}}, \tag{3.18}$$

$$\beta_k^{PR} = \frac{\mathbf{g}_k^\dagger \left(\mathbf{g}_k - \mathbf{g}_{k-1}\right)}{\mathbf{g}_{k-1}^\dagger \mathbf{g}_{k-1}}, \tag{3.19}$$

$$\beta_k^{HS} = -\frac{\mathbf{g}_k^\dagger \left(\mathbf{g}_k - \mathbf{g}_{k-1}\right)}{\Delta\mathbf{m}_{k-1}^\dagger \left(\mathbf{g}_k - \mathbf{g}_{k-1}\right)}, \tag{3.20}$$

$$\beta_k^{DY} = -\frac{\mathbf{g}_k^\dagger \mathbf{g}_k}{\Delta\mathbf{m}_{k-1}^\dagger \left(\mathbf{g}_k - \mathbf{g}_{k-1}\right)}. \tag{3.21}$$

In this research, the "Fletcher-Reeves" method (Fletcher and Reeves, 1964) is used to obtain the parameter β_k. The gradient-based methods are known to converge globally, but possibly very slowly. In most cases, preconditioning is necessary to ensure the fast convergence of the NCG method (Hu et al., 2011).

3.3.4 Quasi-Newton methods

Quasi-Newton methods provide an attractive alternative to Newton-based and gradient-based methods by approximating the inverse Hessian iteratively instead of constructing the Hessian matrix explicitly (Nocedal and Wright, 2006; Brossier et al., 2009; Ma and Hale, 2012). BFGS method, named after Broyden (1970), Fletcher (1970), Goldfarb (1970) and Shanno (1970), is one popular quasi-Newton strategy to approximate the inverse Hessian iteratively using the changes of the model and gradient (Nocedal and Wright, 2006).

In the BFGS updating formula, a symmetric and positive definite matrix $\mathcal{H}_k$ that approximates the inverse of the Hessian, and a pair of vectors $\mathbf{s}_k = \mathbf{m}_{k+1} - \mathbf{m}_k$, and $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$ that indicates the model and gradient changes and satisfies the condition $\mathbf{s}_k^\dagger \mathbf{y}_k > 0$ are given. Using these vectors, the inverse Hessian approximation $\mathcal{H}_{k+1}$ can be computed with the following formula:

$$\mathcal{H}_{k+1} = \mathbf{v}_k^\dagger \mathcal{H}_k \mathbf{v}_k + \mathbf{w}_k \mathbf{s}_k \mathbf{s}_k^\dagger, \tag{3.22}$$

where $\mathbf{w}_k = 1/\mathbf{y}_k^\dagger \mathbf{s}_k$, $\mathbf{v}_k = \mathbf{I} - \mathbf{w}_k \mathbf{y}_k \mathbf{s}_k^\dagger$ and $\mathbf{I}$ is the identity matrix. The initial inverse Hessian approximation $\mathcal{H}_0$ is important to BFGS method and it is usually set as an identity

1. Given initial inverse Hessian approximation $\mathcal{H}_0 = \mathbf{I}$;
2. $\Delta\mathbf{m} = \mathbf{g}$;
3. **For** $i = n-1, n-2, ..., n-M$
4. $\quad \mathbf{x}_i$=$\mathbf{s}_i^\dagger \Delta\mathbf{m} / \mathbf{y}_i^\dagger \mathbf{s}_i$;
5. $\quad \Delta\mathbf{m}$=$\Delta\mathbf{m}$-$\mathbf{x}_i\mathbf{y}_i$;
6. $\quad \mathbf{v} = \mathcal{H}_0 \Delta\mathbf{m}$;
7. **End**
8. **For** $i = n-M, n-M+1, ..., n-1$
9. $\quad \mathbf{z}$=$\mathbf{y}_i^\dagger \mathbf{x} / \mathbf{y}_i^\dagger \mathbf{s}_i$;
10. $\quad \mathbf{v}$=$\mathbf{v} - \mathbf{s}_i (\mathbf{x}_i - \mathbf{z})$;
11. **End**
12. $\mathcal{H}_n \mathbf{g}_n = \mathbf{v}$;
13. $\Delta\mathbf{m}_n$=$\mathbf{g}_n$.

Table 3.1: Two-loop recursion scheme for l-BFGS FWI.

matrix to make sure that the updated matrix maintains positive definiteness (Wu et al., 2015). A limited-memory BFGS (l-BFGS) method was developed by storing the model and gradient changes from a limited number l of previous iterations (typically $l < 10$) (Nocedal, 1980). The stored information is then used to construct an approximated inverse Hessian. A "two-loop recursion" scheme (Table 3.1) is implemented in this research to obtain the search direction using the information of previous updates (Nocedal and Wright, 2006; Métivier et al., 2016).

3.3.5 Hessian-free optimization method

Instead of constructing Hessian or inverse Hessian approximations, the Hessian-free (HF) optimization method, also known as truncated-Newton or inexact-Newton method, obtains the search direction by solving the Newton linear system (equation (2.12)) approximately using a conjugate-gradient (CG) method with matrix-free scheme (Saad, 2003; Anagaw and Sacchi, 2012; Métivier et al., 2014). The CG method is an optimal algorithm for solving a symmetric positive definite system **Wx=b** and it only requires computing the Hessian-vector products $\mathbf{H}v$ instead of forming the Hessian matrix explicitly, where v is an arbitrary vector in model space. The Hessian-vector products can be calculated via finite-difference method (Nocedal and Wright, 2006) or the adjoint-state method (Métivier et al., 2014, 2016).

Notations: $\mathbf{W}$ is a symmetric and positive definite matrix;
$\mathbf{q}$ is the residual; $\tilde{k}$ is the iteration index; $\mathcal{M}$ is the preconditioner;
γ_{min} is the relative residual tolerance; $\tilde{k}_{max}$ is the maximum iteration.

Input: $\mathbf{W}$, $\mathbf{b}$, $\tilde{k}_{max}$, $\mathcal{M}$, γ_{min}

Output: $\mathbf{x}$, γ

Initialization: $\mathbf{q}_0 = \mathbf{b} - \mathbf{W}\mathbf{x}_0$, $\mathbf{z}_0 = \mathcal{M}^{-1}\mathbf{q}_0$, $\mathbf{p}_0 = \mathbf{z}_0$, $\tilde{k} = 0$;

While $\gamma_{\tilde{k}} > \gamma_{min}$ & $\tilde{k} < \tilde{k}_{max}$

$\alpha_{\tilde{k}} = \frac{\mathbf{q}_{\tilde{k}}^{\dagger}\mathbf{z}_{\tilde{k}}}{\mathbf{p}_{\tilde{k}}^{\dagger}\mathbf{W}\mathbf{p}_{\tilde{k}}}$;

$\mathbf{x}_{\tilde{k}+1} = \mathbf{x}_{\tilde{k}} + \alpha_{\tilde{k}}\mathbf{p}_{\tilde{k}}$;

$\mathbf{q}_{\tilde{k}+1} = \mathbf{q}_{\tilde{k}} - \alpha_{\tilde{k}}\mathbf{W}\mathbf{p}_{\tilde{k}}$;

$\mathbf{z}_{\tilde{k}+1} = \mathcal{M}^{-1}\mathbf{q}_{\tilde{k}+1}$;

$\beta_{\tilde{k}+1} = \frac{\mathbf{z}_{\tilde{k}+1}^{\dagger}\mathbf{q}_{\tilde{k}+1}}{\mathbf{z}_{\tilde{k}}^{\dagger}\mathbf{q}_{\tilde{k}}}$;

$\mathbf{p}_{\tilde{k}+1} = \mathbf{z}_{\tilde{k}+1} + \beta_{\tilde{k}}\mathbf{p}_{\tilde{k}}$;

$\gamma_{\tilde{k}} = \frac{\|\mathbf{b}-\mathbf{W}\mathbf{x}_{\tilde{k}+1}\|}{\|\mathbf{b}\|}$;

$\tilde{k} = \tilde{k} + 1$;

End

Table 3.2: Pseudo-code of the preconditioned conjugate-gradient (PCG) method.

The formulas of computing Hessian-vector products have been given by Pratt et al. (1998a). Fichtner and Trampert (2011a) calculated the products of full Hessian with Gaussian shape perturbation vectors with second-order adjoint-state approach in time domain for resolution analysis. Métivier et al. (2013) gave the general framework of the first-order adjoint-state method for gradient calculation and the second-order adjoint-state method for Hessian-vector product calculation. The full Hessian $\mathbf{H}$ (equation (3.10)) arising from the second-order partial derivative is not guaranteed to be positive definite (Nash, 2000). Thus, CG method is no longer appropriate for solving an indefinite linear system. In this research, instead of using the Hessian $\mathbf{H}$, Gauss-Newton Hessian $\mathbf{H}_a$ is used, which is always symmetric and positive semi-definite:

$$\left(\mathbf{H}_{a,k} + \hat{\epsilon}\hat{\mathbf{A}}_k\right) \Delta\mathbf{m}_k = -\mathbf{g}_k, \tag{3.23}$$

where $\hat{\epsilon}\hat{\mathbf{A}}_k$ is the damping term ensuring that $\mathbf{H}_{a,k} + \hat{\epsilon}\hat{\mathbf{A}}$ is positive definite, $\hat{\epsilon}$ is a small constant value and $\hat{\mathbf{A}}_k$ indicates a diagonal matrix consisting of the diagonal elements of the Gauss-Newton Hessian. The resulting algorithm becomes a Levenberg-Marquardt method (Levenberg, 1944; Marquardt, 1963). In this chapter, the HF Gauss-Newton FWI is implemented in a double-iterative scheme: the outer loop is to iteratively update the model parameters for the non-linear optimization problem, and the inner loop is to solve the linear system (equation(3.23)) iteratively with the CG algorithm. The inner iteration is typically stopped or "truncated" before the solution of the Newton equation is obtained.

Now, a review of the adjoint-state method for calculating the products of an arbitrary vector υ with Gauss-Newton Hessian $\mathbf{H}_a$ in frequency domain following Métivier et al. (2013) is given. I start from minimizing the following Lagrangian function:

$$\tilde{\boldsymbol{\chi}}\left(\mathbf{m}\right) = \mathbf{d}_{\text{syn}}^{\dagger}\boldsymbol{\nu} + \tilde{\Lambda}^{\dagger}\left(\mathbf{L}\left(\mathbf{m},\omega\right)\mathbf{u} - \mathbf{f}_s\right), \tag{3.24}$$

where $\boldsymbol{\nu}$ is an arbitrary vector and $\tilde{\Lambda}$ is a new Lagrangian multiplier. Following the process of deriving gradient from equation (3.5) to equation (3.9), the first derivative of the misfit function with respect to model parameter is given by:

$$\nabla_{\mathbf{m}}\tilde{\boldsymbol{\chi}}\left(\mathbf{m}\right) = \mathbf{u}^{\dagger}\nabla_{\mathbf{m}}\mathbf{L}^{\dagger}\left(\mathbf{m},\omega\right)\tilde{\Lambda}, \tag{3.25}$$

where the new Lagrangian multiplier satisfies $\mathbf{L}^{\dagger}\left(\mathbf{m},\omega\right)\tilde{\Lambda} = -\mathcal{P}^{\dagger}\boldsymbol{\nu}$. Replacing $\boldsymbol{\nu}$ with

Jacobian-vector product $\nabla_{\mathbf{m}}\mathbf{d}^*_{\text{syn}}\upsilon$ gives the new Lagrangian multiplier as:

$$\tilde{\Lambda} = -\left(\mathbf{L}^{\dagger}\left(\mathbf{m},\omega\right)\right)^{-1}\nabla_{\mathbf{m}}\mathbf{d}^*_{\text{syn}}\upsilon. \tag{3.26}$$

Inserting equation (3.3) into equation (3.26) and then equation (3.25) gives the Gauss-Newton Hessian-vector product as:

$$\mathbf{H}_a\upsilon = \mathbf{u}^{\dagger}\nabla_{\mathbf{m}}\mathbf{L}^{\dagger}\left(\mathbf{m},\omega\right)\left(\mathbf{L}^{\dagger}\left(\mathbf{m},\omega\right)\right)^{-1}\mathcal{P}^{\dagger}\mathcal{P}\left(\mathbf{L}^{*}\left(\mathbf{m},\omega\right)\right)^{-1}\nabla_{\mathbf{m}}\mathbf{L}^{*}\left(\mathbf{m},\omega\right)\mathbf{u}^{*}\upsilon. \tag{3.27}$$

To obtain the Gauss-Newton Hessian-vector product, one needs to solve two forward modelling problems for getting $\mathbf{u}$ and $\boldsymbol{\nu}$ and another adjoint modelling for getting Λ, which is 1.5 times more expensive than calculating gradient.

A Hessian-free optimization method can be made more competitive with further enhancements, such as, an effective preconditioner for the linear system and appropriate stopping criteria for the inner iterative algorithm (Métivier et al., 2013). With these enhancements, Hessian-free optimization method is a powerful tool for large-scale inverse problems.

3.3.5.1 Preconditioning

The CG iterative algorithm requires many iterations to obtain the approximate solution of a linear system $\mathbf{W}\mathbf{x} = \mathbf{b}$. The convergence rate of the CG method depends on the spectral properties (e.g., its eigenvalues) of the coefficient matrix $\mathbf{W}$ (Nash, 2000). It is often convenient to transform the equation system into one which has the same solution but more favorable spectral properties. This can be achieved by applying a suitable preconditioner $\mathcal{M}$ on the linear system: $\mathcal{M}^{-1}\mathbf{W}\mathbf{x} = \mathcal{M}^{-1}\mathbf{b}$. Thus, the preconditioned Newton system for the HF Gauss-Newton FWI is given by:

$$\mathcal{M}_k^{-1}\left(\mathbf{H}_{a,k} + \hat{\epsilon}\hat{\mathbf{A}}_k\right)\Delta\mathbf{m}_k = -\mathcal{M}_k^{-1}\mathbf{g}_k. \tag{3.28}$$

The solution of equation (3.28) can be obtained by the preconditioned conjugate-gradient (PCG) method. Pseudo-code of the PCG method is given in Table 3.2. The PCG method is expected to reduce the number of inner iterations, improve the convergence rate and accelerate the HF Gauss-Newton FWI. This chapter focuses on developing efficient and stable preconditioning schemes for inner iterative algorithm.

The preconditioner for the CG method is always devised to approximate the Hessian or the inverse Hessian. The traditional Hessian approximations (e.g., diagonal pseudo-Hessian and diagonal Gauss-Newton Hessian) are first considered as the preconditioners for the CG inner iteration. In chapter 2, the pseudo-Hessian is constructed by replacing the Fréchet derivative wavefield with the virtual source in the correlation process (Shin et al., 2001b). Pseudo-Hessian can also be expressed in terms of matrices:

$$\mathbf{H}_{pseudo} = \mathbf{u}^{\dagger} \nabla_{\mathbf{m}} \mathbf{L}^{\dagger} \left(\mathbf{m}, \omega\right) \nabla_{\mathbf{m}} \mathbf{L}^{*} \left(\mathbf{m}, \omega\right) \mathbf{u}^{*}. \tag{3.29}$$

Diagonal pseudo-Hessian $\mathbf{H}_{pseudo}^{\text{diag}}$ can be used as effective preconditioner for inner CG algorithm in HF Gauss-Newton FWI. Calculating diagonal pseudo-Hessian at each iteration does not involve any additional cost. When employing the diagonal Gauss-Newton Hessian as a preconditioner, more computation cost is required for constructing the receiver-side Green's functions (Tao and Sen, 2013).

In this chapter, a pseudo diagonal Gauss-Newton Hessian approximation is used as the preconditioner for the CG algorithm in the inner loop. Furthermore, considering reflection survey and assuming that the sources and receivers are co-located, the receiver-side Green's function can be replaced by the source-side Green's function, leading to a new diagonal Hessian approximation. This diagonal Hessian approximation was employed by Plessix and Mulder (2004) and Choi et al. (2008) to precondition the gradient. In this chapter, it is used to precondition the Newton system and referred to as pseudo diagonal Gauss-Newton Hessian $\tilde{\mathbf{H}}_{a}^{\text{diag}}$. It can be constructed at no additional cost. This approximation is valid for reflection survey but invalid for transmission or VSP (Vertical Seismic Profile) survey. Summarily, the diagonal pseudo-Hessian, diagonal Gauss-Newton Hessian and pseudo diagonal Gauss-Newton Hessian preconditioners are given by:

$$\mathcal{M}_{k}^{DPH} = \mathbf{H}_{pseudo,k}^{\text{diag}} + \hat{\lambda}\mathcal{B}_{k}, \mathcal{M}_{k}^{DGH} = \mathbf{H}_{a,k}^{\text{diag}} + \hat{\lambda}\mathcal{C}_{k}, \mathcal{M}_{k}^{PDGH} = \tilde{\mathbf{H}}_{a,k}^{\text{diag}} + \hat{\lambda}\mathcal{D}_{k}, \tag{3.30}$$

where $\hat{\lambda}\mathcal{B}_{k}$, $\hat{\lambda}\mathcal{C}_{k}$ and $\hat{\lambda}\mathcal{D}_{k}$ are the stabilization terms, $\hat{\lambda}$ is a small constant value, $\mathcal{B}_{k}$, $\mathcal{C}_{k}$ and $\mathcal{D}_{k}$ indicate diagonal matrices consisting of the maximum values of the diagonal preconditioners. These three different preconditioning strategies are referred to as DPH-GN, DGH-GN and PDGH-GN here. When the parameter $\hat{\lambda}$ is very large, these preconditioning methods approach the non-preconditioned HF Gauss-Newton (CG-GN) method.

Furthermore, this thesis develops an l-BFGS preconditioning scheme for the HF optimization method (Nash, 1985, 2000; Métivier et al., 2012; Sainath et al., 2013), namely l-BFGS-GN method for $\mathcal{H}_0 = \mathbf{I}$. The l-BFGS approximated inverse Hessian $\mathcal{H}$ can also be used as a preconditioner for the CG iterative method:

$$\mathcal{H}_k \left(\mathbf{H}_{a,k} + \hat{\epsilon} \hat{\mathbf{A}}_k \right) \Delta \mathbf{m}_k = -\mathcal{H}_k \mathbf{g}_k. \tag{3.31}$$

Traditionally, an identity matrix $\mathbf{I}$ is usually set as the initial guess $\mathcal{H}_0$. In this chapter, I consider using the stabilized diagonal pseudo-Hessian, diagonal Gauss-Newton Hessian and pseudo diagonal Gauss-Newton Hessian (equation (3.30)) as the initial guess for constructing the l-BFGS preconditioners. These methods are referred to as l-BFGS-GN-DPH, l-BFGS-GN-DGH and l-BFGS-GN-PDGH methods respectively. Similarly, when parameter $\hat{\lambda}$ is very large, these three methods approach l-BFGS-GN method ($\mathcal{H}_0 = \mathbf{I}$). Allowing the diagonal approximations to vary throughout the iterations can also enhance the l-BFGS inverse Hessian approximation, which has been used in preconditioned l-BFGS methods by Métivier et al. (2014). The use of this strategy in constructing l-BFGS preconditioner is a match of ongoing study.

3.3.5.2 Stopping criteria

Newton's method is based on the Taylor series approximation (equation (2.10)). If this approximation is inaccurate then it may not be suitable to solve the Newton equations accurately and "over-solving" the Newton equation will not produce a better search direction (Nash, 2000). The CG algorithm should be terminated with an appropriate stopping criteria. The maximum inner iteration number $\tilde{k}_{max}$ is given and the relative residual $\gamma_{\tilde{k}}$ is defined as:

$$\gamma_{\tilde{k}} = \frac{\|\tilde{\mathbf{H}}_{\tilde{k}} \Delta \mathbf{m}_{\tilde{k}} + \mathbf{g}_{\tilde{k}}\|}{\|\mathbf{g}_{\tilde{k}}\|}, \tag{3.32}$$

where $\tilde{k}$ indicates the CG inner iteration index. The inner iteration is stopped when $\gamma_{\tilde{k}} < \gamma_{min}$, where γ_{min} indicates the relative residual tolerance. The stopping criterion for the CG algorithm used in this research is relative basic. Eisenstat and Walker (1996) and Métivier et al. (2014) used a more advanced adaptive stopping criteria, within which the CG iterations are stopped, when a negative curvature direction is calculated during solving the Newton system.

Notations: k_{max} is the maximum outer iteration; ϕ is the normalized misfit; ϕ_{min} is the minimum normalized misfit; $\mathbf{N}_\omega$ is the multi-frequency group.

Input: $\mathbf{m}_0$, ϕ_{min}, $\hat{\epsilon}$, k_{max}, M, $\mathbf{N}_\omega$, $\hat{\lambda}$, $\mathbf{d}_{\text{obs}}$, $\tilde{k}_{max}$, γ_{min}

Output: $\mathbf{m}$, ϕ

Initialization: $k = 0$;

For $k < k_{max}$ & $\phi_k > \phi_{min}$

1. Calculate $\mathbf{d}_{\text{syn}}\left(\mathbf{m}_k, \mathbf{N}_\omega^k\right)$ and $\Delta\mathbf{d}\left(\mathbf{m}_k, \mathbf{N}_\omega^k\right) = \mathbf{d}_{\text{obs}}\left(\mathbf{N}_\omega^k\right) - \mathbf{d}_{\text{syn}}\left(\mathbf{m}_k, \mathbf{N}_\omega^k\right)$;
2. Construct $\mathbf{g}_k\left(\mathbf{m}_k, \mathbf{N}_\omega^k\right) \leftarrow \Delta\mathbf{d}\left(\mathbf{m}_k, \mathbf{N}_\omega^k\right)$ with the adjoint-state method;
3. Construct $\mathcal{M}_k = \mathcal{M}_k^{DPH}, \mathcal{M}_k^{DGH}, \mathcal{M}_k^{PDGH}$ or $\mathcal{M}_k^{-1} = \mathcal{H}_k$;
4. Call PCG method $\leftarrow \tilde{k}_{max}, \gamma_{min}, \mathcal{M}_k, \mathbf{g}_k, \mathbf{H}_{a,k}\upsilon$ and calculate $\Delta\mathbf{m}_k$:
   $$\mathcal{M}_k^{-1}\left(\mathbf{H}_{a,k} + \hat{\epsilon}\hat{\mathbf{A}}_k\right)\Delta\mathbf{m}_k = -\mathcal{M}_k^{-1}\mathbf{g}_k;$$
5. Get μ_k with a line search method satisfying weak Wolfe condition;
6. $\mathbf{m}_{k+1} = \mathbf{m}_k + \mu_k\Delta\mathbf{m}_k$;
7. Calculate the normalized misfit ϕ_k;
8. $k = k + 1$;

End

Table 3.3: Pseudo-code of the HF Gauss-Newton FWI with PCG method.

3.3.6 Line search with Wolfe condition

A common inexact line search condition stipulates that step length μ_n should lead to sufficient decrease in the objective function, as measured by the following inequality (Nocedal and Wright, 2006):

$$\Phi\left(\mathbf{m}_k + \mu_k\Delta\mathbf{m}_k\right) \leq \Phi\left(\mathbf{m}_k\right) + \mu_k c_1 \Delta\mathbf{m}_k^\dagger \nabla\Phi\left(\mathbf{m}_k\right), \tag{3.33}$$

where c_1 is one constant parameter satisfying $0 < c_1 < 1$. Equation (3.33) is the Armijo condition, which ensures that the step length μ_k decreases the misfit function sufficiently (Armijo, 1966). While, this sufficient decrease condition is not enough to ensure convergence since this condition is satisfied for all small enough μ. To rule out unacceptably small steps, a weak Wolfe condition (or curvature condition) is introduced:

$$\nabla\Phi\left(\mathbf{m}_k + \mu_k\Delta\mathbf{m}_k\right)^\dagger \Delta\mathbf{m}_k \geq c_2\nabla\Phi\left(\mathbf{m}_n\right)^\dagger \Delta\mathbf{m}_k, \tag{3.34}$$

where c_2 is constant parameter satisfying $c_1 < c_2 < 1$. Equation (3.34) ensures that the slope has been reduced sufficiently (Nocedal and Wright, 2006). In practice, parameter c_1 can be chosen as $c_1 = 10^{-4}$ and parameter c_2 should be much larger $c_2 = 0.9$ (Nocedal and Wright, 2006). The initial step length μ_0 is always chosen as 1 and then after a set of trial

step lengths, the optimal one will be accepted to satisfy the above conditions. Pseudo-code of the HF Gauss-Newton method with a PCG algorithm is provided in Table 3.3.

3.4 Numerical examples

In this section, first the preconditioned Hessian-free Gauss-Newton FWI is applied on a Gaussian-anomaly model in comparison with steepest-descent (SD), non-linear conjugate-gradient (NCG) and l-BFGS methods. Different non-linear optimization methods are also applied on a modified Marmousi model. The preconditioning schemes for the HF Gauss-Newton FWI described in the previous section are also examined. The inversion results verify that the preconditioning strategies accelerate the HF Gauss-Newton FWI, improve its convergence rate and reduce computation cost. At the end, different optimization methods are applied to reconstruct velocity and density simultaneously.

3.4.1 The Gaussian-anomaly model

In this numerical example, the SD, NCG, l-BFGS and preconditioned HF Gauss-Newton methods are applied to the reconstruction of a Gaussian-anomaly model. A Gaussian-anomaly velocity model is constructed consisting of 50×100 grid cells with a grid interval of 10 m in both horizontal and vertical directions. A total of 49 sources are deployed from 20 m to 980 m with a source interval of 20 m and a depth of 20 m. A total of 100 receivers are distributed on the surface from 10 m to 1000 m with a receiver interval of 10 m and a depth of 20 m. A Ricker wavelet with a 30 Hz dominant frequency is used as the source function. In Figure 3.1, the true velocity model is illustrated. The initial velocity model is homogeneous with a constant velocity of 2 km/s. Most of the energy in the data residuals are scattered waves, which are used for inversion in this numerical example.

For comparison of different methods, the stopping criteria for inversion is set as: the maximum outer iteration $k_{max} = 101$ and the minimum normalized misfit $\phi_{min} = 2.0e\text{-}5$. The normalized misfit is defined to be the misfit of the inverted model divided by the misfit of the initial model. To illustrate the performances of different preconditioning strategies, the stopping criteria for the inner CG algorithm is also defined: the maximum inner iteration

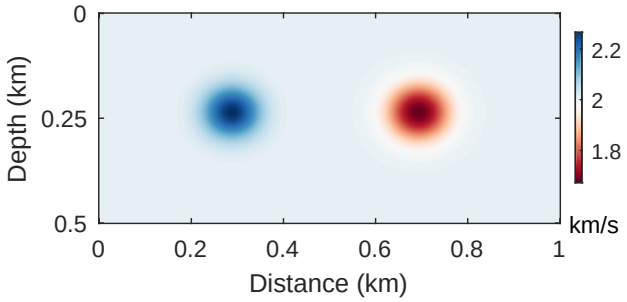

Figure 3.1: The true Gaussian-anomaly model.

$\tilde{k}_{max} = 100$ and the minimum relative residual $\gamma_{min} = 1.0e\text{-}2$. The parameters used for stabilization are $\hat{\epsilon} = 2.0e\text{-}2$ (equation (3.23)) and $\hat{\lambda} = 5.0e\text{-}2$ (equations (3.30)). From our experience, overly small $\hat{\epsilon}$ and $\hat{\lambda}$ values may result in artifacts in the inverted models and the two stabilization factors should not be smaller than $1.0e\text{-}2$.

Figure 3.2a shows the gradient update in SD method and Figure 3.2b shows the update in Hessian-free Gauss-Newton method with 30 inner conjugate-gradient iterations. By applying Hessian iteratively to the gradient update, the gradient has been obviously de-blurred. Next, the inverted models using different non-linear optimization methods are illustrated. Figures 3.3a, 3.3b, 3.3c and 3.3d show the inversion results obtained with SD, NCG, l-BFGS ($\mathcal{H}_0 = \mathbf{I}$) and non-preconditioned CG-GN methods. Figure 3.4 shows the convergence history and RLSE (equation (2.42)) of the these different optimization methods. The SD, NCG and l-BFGS methods undergo $k_{max} = 101$ iterations without reaching the minimum misfit requirement of $\phi_{min} = 2.0e\text{-}5$. The non-preconditioned CG-GN method converges towards the minimum misfit requirement quadratically with 8 outer iterations and in this sense has clear advantages with respect to convergence rate. However, the CG-GN method is still very expensive for solving the Newton equations iteratively. Effective preconditioning strategies can be applied to improve its performance.

The diagonal Hessian approximations preconditioned HF Gauss-Newton methods: DPH-GN, DGH-GN and PDGH-GN methods, are then applied. The diagonal pseudo-Hessian, diagonal Gauss-Newton Hessian and pseudo diagonal Gauss-Newton Hessian are then used as initial guesses to construct the l-BFGS preconditioners, which form the above-mentioned l-BFGS-GN-DPH, l-BFGS-GN-DGH and l-BFGS-GN-PDGH methods. Because the inversion results produced by these preconditioned HF Gauss-Newton methods are very close to the

non-preconditioned one (Figure 3.3d), only the maximum iterations, normalized minimum misfit (ϕ), number of forward problems solved and computation time for these different preconditioning strategies are illustrated in Table 3.4. From Table 3.4, it can be seen that the preconditioning strategies improve the convergence rate, reduce the computation cost and accelerate HF Gauss-Newton method. Of the three diagonal Hessian approximations preconditioning methods, the DGH-GN method provides the best inversion result with the minimum normalized misfit of 1.58e-5. However, this method needs to solve more forward modelling problems for constructing the receiver-side Green's functions, which increases the computational burden. Considering both inversion accuracy and computation efficiency simultaneously, PDGH-GN method outperforms DPH-GN and DGH-GN methods. The three l-BFGS preconditioning methods show better performances than the corresponding diagonal Hessian approximations preconditioning methods. Furthermore, the proposed l-BFGS-GN-PDGH method gives the best inversion result and is most efficient among these preconditioning strategies. In comparison with the non-preconditioned CG-GN method, l-BFGS-GN-PDGH method decreases the normalized misfit value by 12% and reduces the computation cost by 28% approximately.

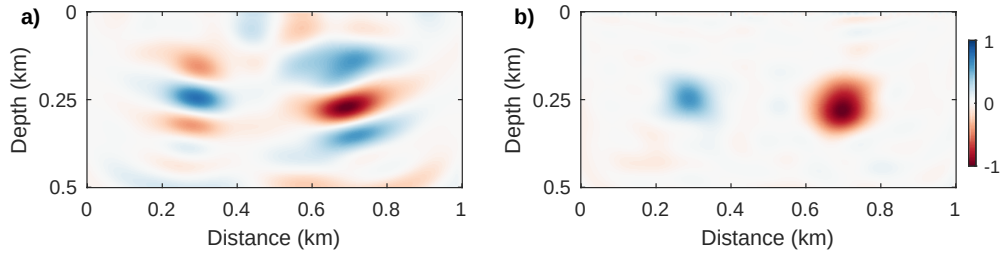

Figure 3.2: (a) gradient update (b) Hessian-free Gauss-Newton update with 30 inner conjugate-gradient iterations.

3.4.2 The modified Marmousi model

A more complex modified Marmousi model is used to compare the performances of different optimization methods and examine the efficiency of different preconditioning schemes for the HF Gauss-Newton full-waveform inversion. The truncated Marmousi model has 100×100 grid cells with a grid interval of 10 m in both horizontal and vertical directions. A number of 49 sources from 20 m to 980 m at depth of 20 m with a regular source spacing of 20

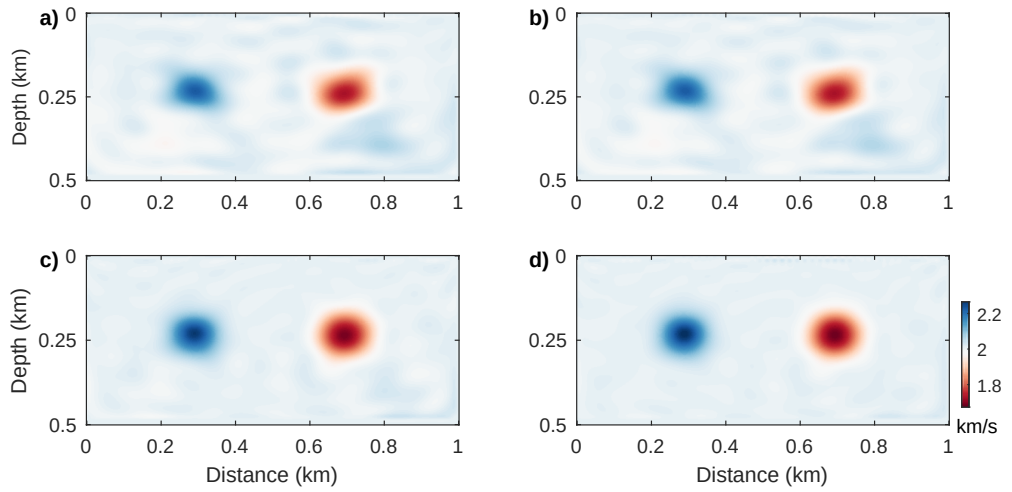

Figure 3.3: Inversion results by (a) SD method; (b) NCG method; (c) l-BFGS method; (d) CG-GN method.

Methods	Iteration	Minimum Misfit	Problems Solved	Time (s)
CG-GN	8	1.77e-5	619	683.46
DPH-GN	7	1.73e-5	600	672.63
DGH-GN	6	1.58e-5	631	700.23
PDGH-GN	6	1.59e-5	552	643.78
l-BFGS-GN-DPH	6	1.71e-5	482	532.34
l-BFGS-GN-DGH	6	1.56e-5	617	679.54
l-BFGS-GN-PDGH	6	1.56e-5	454	521.20

Table 3.4: Efficiency comparison for preconditioning strategies with the Gaussian-anomaly model.

m are distributed. Fifty receivers are arranged from 10 m to 1000 m every 20 m at the depth of 20 m. The source function is a Ricker wavelet with a dominant frequency of 30 Hz. Figures 3.5a and 3.5b show the true P-wave velocity model and initial P-wave velocity model. The initial velocity model is obtained by smoothing the true model with a Gaussian function. The inversion process is carried out with a multi-scale approach for mitigating the cycle-skipping problem (Pratt and Chapman, 1992; Bunks et al., 1995; Sirgue and Pratt, 2004). The frequencies used for inversion are increased from 5 Hz to 40 Hz with a partial overlap-frequency selection strategy, in which a group of 3 frequencies are used for inversion simultaneously. The frequency group increases from low to high with 2 frequencies overlapped and for each frequency band, a number of 5 outer iterations are performed. The

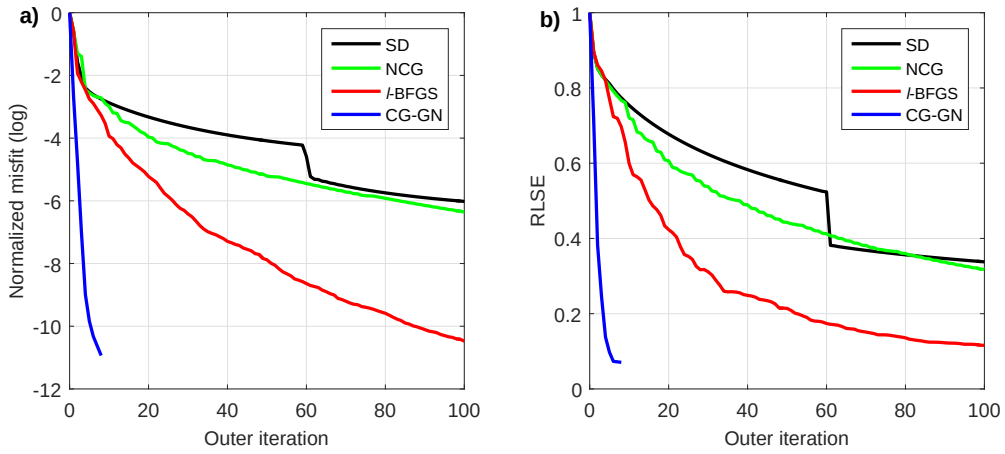

Figure 3.4: Comparison of the convergence history (a) and RLSE (b) for different optimization methods. The black, green, red and blue lines SD, NCG, l-BFGS and non-preconditioned CG-GN methods respectively.

stopping criteria for the inner iteration are $\tilde{k}_{max} = 10$ and/or $\gamma_{min} = 2.0e\text{-}1$. The stabilization parameter $\hat{\epsilon}$ is $1.0e\text{-}2$.

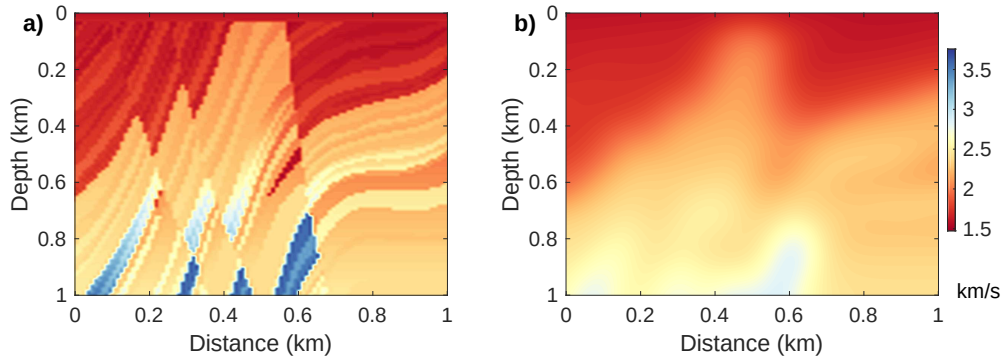

Figure 3.5: (a) True P-wave velocity model; (b) Initial P-wave velocity model.

Figures 3.6a and 3.6b illustrate the inversion results obtained by SD and NCG methods. Figures 3.7a and 3.7b illustrate the inversion results obtained by l-BFGS ($\mathcal{H}_0 = \mathbf{I}$) and non-preconditioned CG-GN methods. Figures 3.6c, 3.6d, 3.7c and 3.7d show the comparison of well log data at 0.1 km and 0.6 km. The SD and NCG methods are limited in recovering the deep parts of the model. The l-BFGS method ($\mathcal{H}_0 = \mathbf{I}$) provides better inversion result but the deep parts of the inversion result are still not satisfactory. Compared to SD, NCG, and l-BFGS methods, the deep parts of the reconstructed model by non-preconditioned CG-GN method have been enhanced obviously.

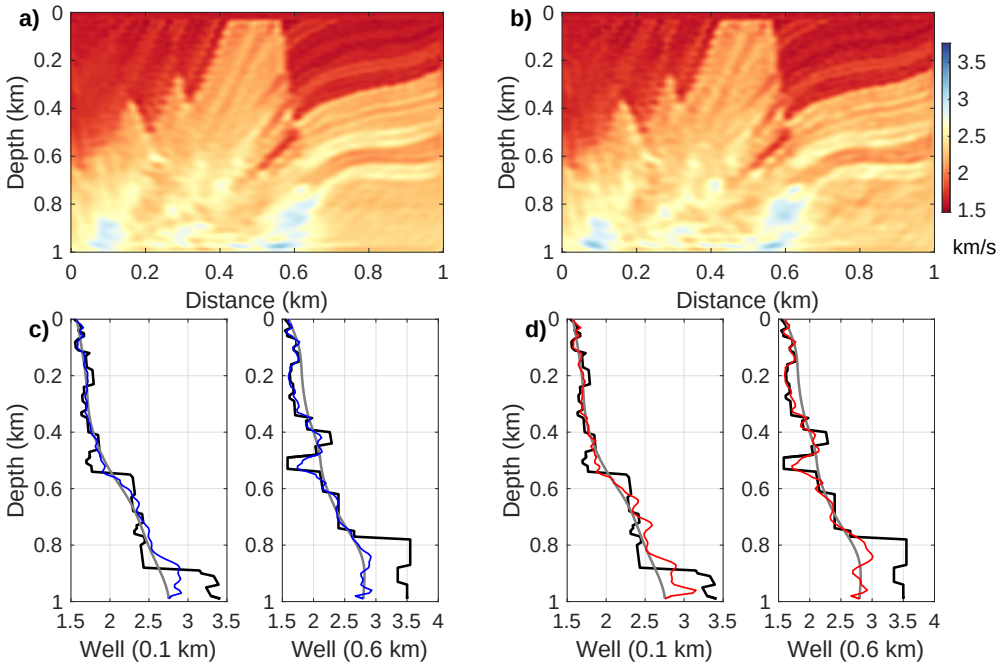

Figure 3.6: (a) SD method ($\phi = 0.31$); (b) NCG method ($\phi = 0.11$). (c) and (d) are the comparison of well log data at 0.1 km and 0.6 km. The black, gray, blue and red lines indicate the true model, initial model, inverted models by SD and NCG methods.

The inversion results produced by diagonal Hessian approximations preconditioned HF Gauss-Newton methods are then illustrated. Figures 3.8a, 3.8b and 3.8c are the inverted models obtained by DPH-GN, DGH-GN and PDGH-GN methods respectively. Figures 3.8d, 3.8e, and 3.8f show the well log data comparison at 0.1 km and 0.6 km. For DPH-GN method, the stabilization parameter $\hat{\lambda}$ is 1.0e-2. Incorporating receiver-side Green's functions increases the instability of preconditioning. The stabilization parameter $\hat{\lambda}$ for DGH-GN and PDGH-GN methods is 5.0e-2. Compared to the model inverted by the non-preconditioned CG-GN method (Figure 3.7b), the inverted model (Figure 3.8a) by diagonal pseudo-Hessian preconditioned HF Gauss-Newton method (DPH-GN) is improved without any additional cost. When employing diagonal Gauss-Newton Hessian for preconditioning, the inversion result is improved further but at the cost of a greater computational burden for constructing the receiver-side Green's functions. In pseudo diagonal Gauss-Newton Hessian preconditioned HF Gauss-Newton method, the receiver-side Green's functions are approximated by source-side Green's functions. This preconditioning strategy provides comparable inversion result with the diagonal Gauss-Newton Hessian preconditioning one but without involving

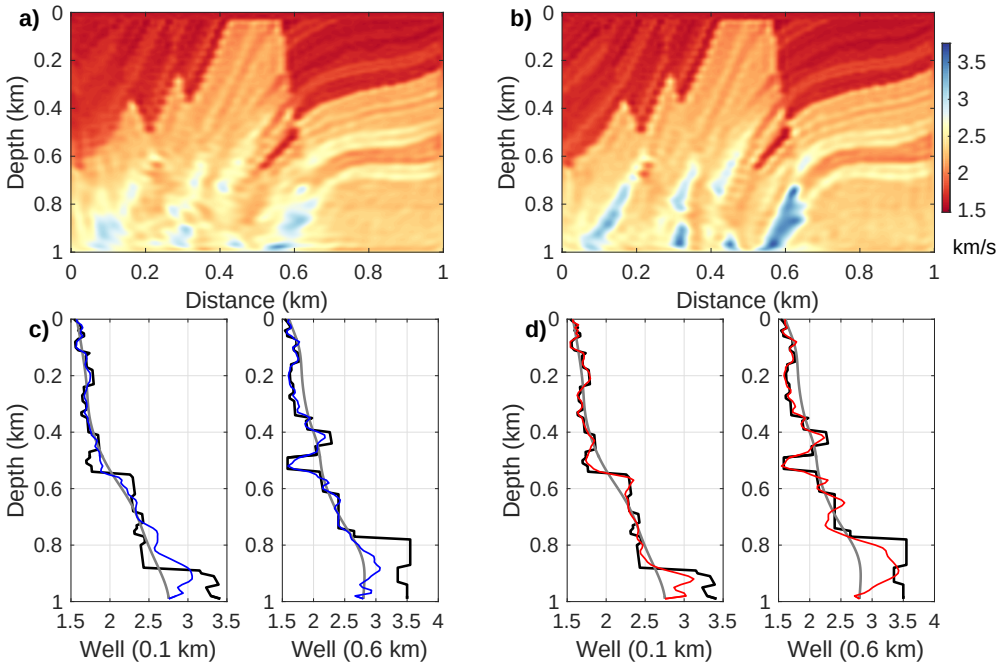

Figure 3.7: (a) l-BFGS method ($\phi = 0.07$); (b) CG-GN method ($\phi = 4.1e$-3). (c) and (d) are the comparison of well log data at 0.1 km and 0.6 km. The black, gray, blue and red lines indicate the true model, initial model, inverted models by l-BFGS ($\mathcal{H}_0 = \mathbf{I}$) and non-preconditioned CG-GN methods.

additional cost. Hence, it is concluded that PDGH-GN method outperforms DPH-GN and DGH-GN methods.

The second class of preconditioning scheme is l-BFGS inverse Hessian preconditioning strategy with diagonal Hessian approximations as initial guesses. Figures 3.9a, 3.9b and 3.9c show the inverted models by l-BFGS-GN-DPH ($\hat{\lambda} = 1.0e$-2), l-BFGS-GN-DGH ($\hat{\lambda} = 5.0e$-2) and l-BFGS-GN-PDGH ($\hat{\lambda} = 5.0e$-2) methods. Figures 3.9d, 3.9e and 3.9f show the well log data comparison at 0.1 km and 0.6 km. Compared to Figures 3.8a, 3.8b and 3.8c, it is observed that the corresponding inverted models are further enhanced with the l-BFGS preconditioning strategies as indicated by the normalized misfit values ϕ in figure captions. Furthermore, l-BFGS-GN-DGH and l-BFGS-GN-PDGH methods obtain better inversion results than l-BFGS-GN-DPH method for considering the contributions from receiver-side Green's functions. l-BFGS-GN-DGH method is more expensive for calculating the receiver-side Green's functions at each outer iteration. Thus, l-BFGS-GN-PDGH method becomes the best choice among these preconditioning methods.

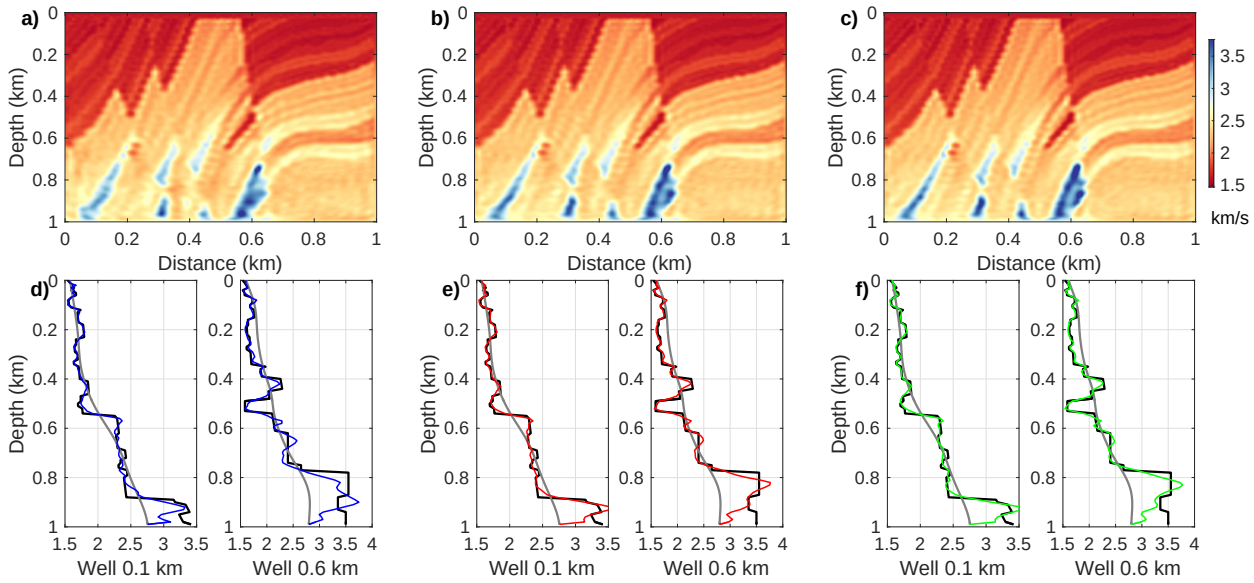

Figure 3.8: Comparison of inversion results by diagonal Hessian approximations preconditioned Hessian-free Gauss-Newton methods. (a) DPH-GN method (ϕ = 4.4e-3); (b) DGH-GN method (ϕ = 2.5e-3); (c) PDGH-GN method (ϕ = 2.1e-3); (d), (e) and (f) show the comparison of well log data at 0.1 km and 0.6 km. The blue, red and green lines indicate the inverted models by DPH-GN, DGH-GN and PDGH-GN methods.

To further examine the proposed l-BFGS-GN-PDGH method, convergence history of non-preconditioned CG-GN, PDGH-GN and l-BFGS-GN-PDGH methods are plotted in Figure 3.10. With preconditioning, the convergence rate of Hessian-free Gauss-Newton method is significantly improved. Furthermore, the l-BFGS preconditioning strategy converges faster than diagonal Hessian approximation preconditioning one. In Figure 3.11a, the Normalized misfit vs. Number of forward problems solved is illustrated. Figure 3.11b shows the Normalized misfit vs. Computation time (s). The proposed l-BFGS-GN-PDGH method improves the convergence rate, reduces the computation cost and accelerates the HF Gauss-Newton FWI. In this example, compared to non-preconditioned CG-GN method, the normalized misfit value of the final inverted model is reduced by 27% approximately with the proposed preconditioning strategy. To achieve the same normalized misfit valve (e.g., $\log(\phi)$=$-$6), the computation cost is nearly reduced by 53%.

Figures 3.12a and 3.12b illustrate the inversion results generated using non-preconditioned CG-GN method with noisy data when SNR (signal to noise ratio) are 10 and 6. Figures 3.12c and 3.12d are the inversion results obtained by l-BFGS-GN-PDGH method when SNR

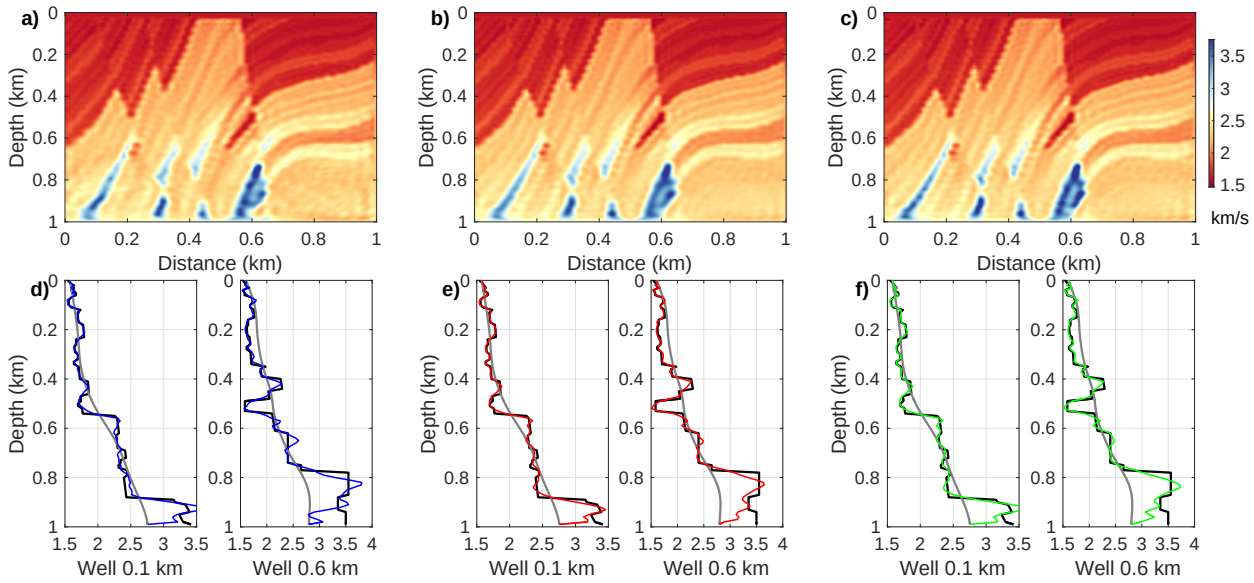

Figure 3.9: Comparison of inversion results by l-BFGS preconditioned Hessian-free Gauss-Newton methods. (a) l-BFGS-GN-DPH method ($\phi = 1.1e$-3); (b) l-BFGS-GN-DGH method ($\phi = 8.4e$-4); (c) l-BFGS-GN-PDGH method ($\phi = 8.3e$-4); (d), (e) and (f) show the comparison of well log data at 0.1 km and 0.6 km. The blue, red and green lines indicate the inverted models by l-BFGS-GN-DPH, l-BFGS-GN-DGH and l-BFGS-GN-PDGH methods.

are 10 and 6. Figures 3.12e and 3.12f show the well log data comparison at 0.1 km and 0.6 km. When SNR is 10, the structures of the inverted models can still be recognized and l-BFGS-GN-PDGH method gives a slightly better inversion result in comparison with non-preconditioned CG-GN method. As increasing random noise when SNR=6, the deep parts of the inverted models are contaminated by artifacts more seriously and the l-BFGS-GN-PDGH inverted model appears to involve stronger artifacts compared to CG-GN inverted model. This is because incorporating diagonal Hessian approximations for preconditioning increases instability of the inverse problem. The preconditioned HF Gauss-Newton FWI is more sensitive to random noise. Enlarging the stabilization parameters $\hat{\epsilon}$ and $\hat{\lambda}$ is expected to reduce the involved artifacts. The performances of different non-linear optimization methods for this Marmousi model example are compared by examining the RLSE vs. Iterations, as indicated in Figure 3.13. Compared to SD, NCG, l-BFGS methods, non-preconditioned CG-GN method will provide inverted model with higher quality. Furthermore, the l-BFGS preconditioned HF Gauss-Newton method outperforms the non-preconditioned one.

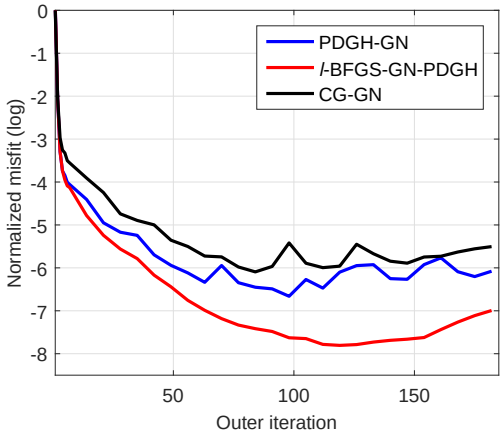

Figure 3.10: Comparison of the convergence history for the HF Gauss-Newton methods with different preconditioning schemes. The black-solid, blue-dash and red-dash-dot lines indicate the convergence history of CG-GN, PDGH-GN and l-BFGS-GN-PDGH methods respectively. Note: here I plot the minimum normalized misfit of each frequency band.

3.4.3 Applying preconditioned Hessian-free Gauss-Newton method on multiparameter acoustic FWI

In this numerical example, different non-linear optimization methods are applied to reconstruct velocity and density simultaneously in multiparameter acoustic FWI. Figures 3.14a, 3.14b, 3.14c and 3.14d show the true velocity, true density, initial velocity and initial density models respectively. The model and acquisition parameters are the same with previous example. The frequencies used for inversion are increased from 8 Hz to 26 Hz. At each frequency band, 2 frequencies are used for inversion simultaneously and 10 outer iterations are performed. The maximum number of inner iteration $\tilde{k}_{max}$ is 5. Other inversion parameters are the same with previous example.

Figure 3.15 shows the inverted velocity and density models using SD, NCG and l-BFGS methods. It can be observed that the velocity models are better recovered than density models. Deep parts in the inverted velocity models are very weak. Additionally, it shows that the density models are either overestimated or underestimated, which may be caused by the parameter crosstalk artifacts when inverting velocity and density simultaneously (Operto et al., 2013). Figure 3.16 show the inverted velocity and density models by non-preconditioned CG-GN and l-BFGS preconditioned HF Gauss-Newton (l-BFGS-GN-PDGH) methods. Deep parts of the velocity model (Figure 3.16a) and density model (Figure 3.16b)

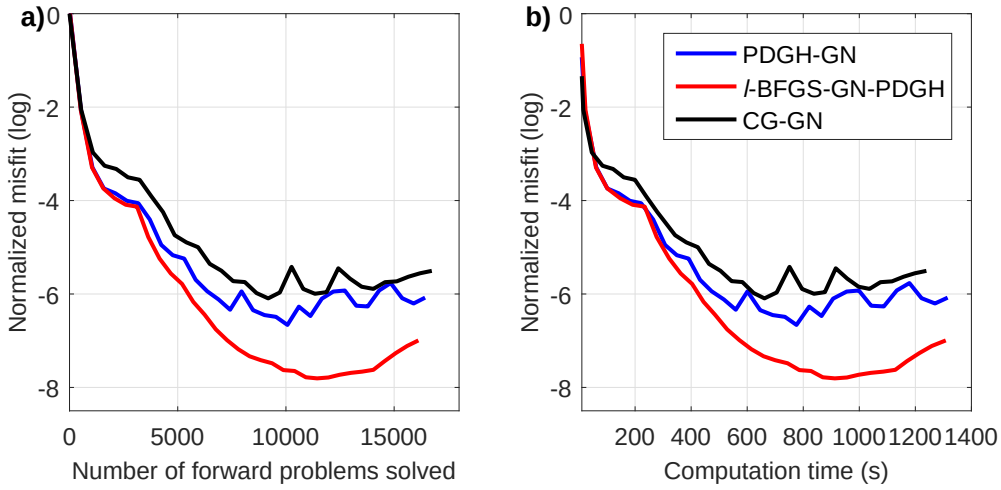

Figure 3.11: (a) Normalized misfit (log) vs. Number of forward problems solved; (b) Normalized misfit (log) vs. Computation time (s). The black-solid, blue-dash and red-dash-dot lines indicate non-preconditioned CG-GN, PDGH-GN and l-BFGS-GN-PDGH methods respectively. Note: here I plot the minimum normalized misfit of each frequency band.

by non-preconditioned CG-GN method are obscure and blurred. In Figures 3.16c and 3.16d, l-BFGS-GN-PDGH method enhances and resolves the deep parts of the models further. This means that the proposed l-BFGS preconditioner constructed using pseudo diagonal Gauss-Newton Hessian can improve the capability of inverse Hessian in the inversion process. In Figures 3.17a and 3.17b, the convergency history and Normalized misfit vs. Number of forward problems solved for non-preconditioned CG-GN and l-BFGS-GN-PDGH methods are plotted. With the proposed preconditioning strategy, the convergence rate is improved by 17% and to achieve the same normalized misfit value (e.g., $\log(\phi)$=-4), the computation cost is reduced by 20% approximately. Figure 3.18 plots the RLSE vs Number of Outer Iterations for reconstructing velocity and density simultaneously, from which it is seen that the l-BFGS preconditioned HF Gauss-Newton methods gives the best inverted models.

Newton-based methods are expected to suppress the parameter crosstalk artifacts in multiparameter FWI. However, for Hessian-free optimization methods, only a few number iterations in the inner loop is affordable for large-scale inverse problems (generally, $\tilde{k}_{max} \leq 10$) (Eisenstat and Walker, 1996; Baumstein, 2014; Métivier et al., 2016). For complex models, if the maximum number of inner iteration is very large (e.g., $\tilde{k}_{max}$=30), the HF Gauss-Newton FWI becomes unstable and the inverted models will be contaminated by artifacts. The

maximum number of inner iteration $\tilde{k}_{max}$ is set to be 5 in this multiparameter example. In this condition, only partial Hessian at each outer iteration is applied to precondition the gradient. The strong parameter crosstalk between velocity and density is difficult to be eliminated completely by the approximate inverse Hessian. In this example, the initial density model is pretty poor. At earlier iterations, when using low frequencies for inversion, there are residual parameter crosstalk artifacts in the inverted density model. As the iteration proceeds, the artifacts accumulate, which makes the density model quite difficult to be reconstructed.

3.5 Discussion

It is known that multiparameter FWI suffers from parameter crosstalk because of the coupling effects between different physical parameters (Operto et al., 2013). The Gauss-Newton Hessian has the ability of suppressing these parameter crosstalk (Innanen, 2014b; Métivier et al., 2015). Explicit construction of the Hessian matrix is extremely expensive. The preconditioned Hessian-free Gauss-Newton method developed in this chapter represents an attractive strategy for large-scale multiparameter FWI. However, Hessian-free optimization methods only apply partial Hessian to precondition the gradient, which is controlled by the stopping criteria of the inner loop. If $\tilde{k}_{max}$ is set to be very small (e.g., $\tilde{k}_{max}$=5), Hessian-free optimization methods may still be limited in mitigating parameter crosstalk especially when the long wavelength components of the multiparameter models are deficient. The pseudo diagonal Gauss-Newton Hessian used in this chapter only incorporates the diagonal elements of diagonal blocks in multiparameter Gauss-Newton Hessian. Diagonal elements of the off-diagonal blocks combined with the l-BFGS preconditioning strategy should be employed for preconditioning in multiparameter FWI (Métivier et al., 2015). For better reconstructing density model, the empirical relationship (e.g., Gardner's rule) can be used to create the initial density model or considered as constraints in the inversion process. Furthermore, parameterization is considered to be important for managing the parameter crosstalk in multiparameter FWI (Tarantola, 1986). In the numerical section, a simple multiparameter acoustic FWI with velocity-density parameterization is given using the preconditioned HF

Gauss-Newton method. For further studies, multiparameter FWI in more complex media (e.g., elastic and anisotropic media) with different parameter classes will be examined with the preconditioned HF Gauss-Newton method.

Inappropriate $\hat{\epsilon}$ (equation (3.23)) and $\hat{\lambda}$ (equations (3.30)) selections may make the inverse problems unstable and involve artifacts in the inversion results especially in the presence of noisy data. In this research, these two damping coefficients are chosen according to our personal experience and kept unchanged during the inversion process. A better strategy for determining $\hat{\epsilon}$ and $\hat{\lambda}$ values is to adapt them iteratively (Marquardt, 1963), which can ensure the stability and quadratic convergence rate (Pratt et al., 1998b).

3.6 Conclusions

In this chapter, different non-linear optimization methods for full-waveform inversion are described and their performances in reconstructing velocity and density parameters are evaluated. To accelerate the Hessian-free Gauss-Newton method, this research develops different preconditioning schemes for the inner CG algorithm at each outer iteration. A pseudo diagonal Gauss-Newton Hessian is also used as preconditioner based on the reciprocal property of the Green's function. Furthermore, I propose a l-BFGS preconditioning strategy by employing the diagonal Hessian approximations as initial guess. Some numerical examples are presented to compare the inverted models by different optimization methods and show that the preconditioning schemes can improve the convergence rate of HF Gauss-Newton FWI and reduce the computation cost. It is demonstrated that the l-BFGS preconditioning method with pseudo diagonal Gauss-Newton Hessian as initial guess can speed up the HF Gauss-Newton FWI most efficiently but is more sensitive to noisy data compared to non-preconditioned one. The l-BFGS preconditioned HF Gauss-Newton method can also reconstruct velocity and density better and more efficiently compared to non-preconditioned HF Gauss-Newton method. Different optimization methods are applied to reconstruct velocity and density parameters simultaneously. Preconditioned HF Gauss-Newton method provides the best inverted models.

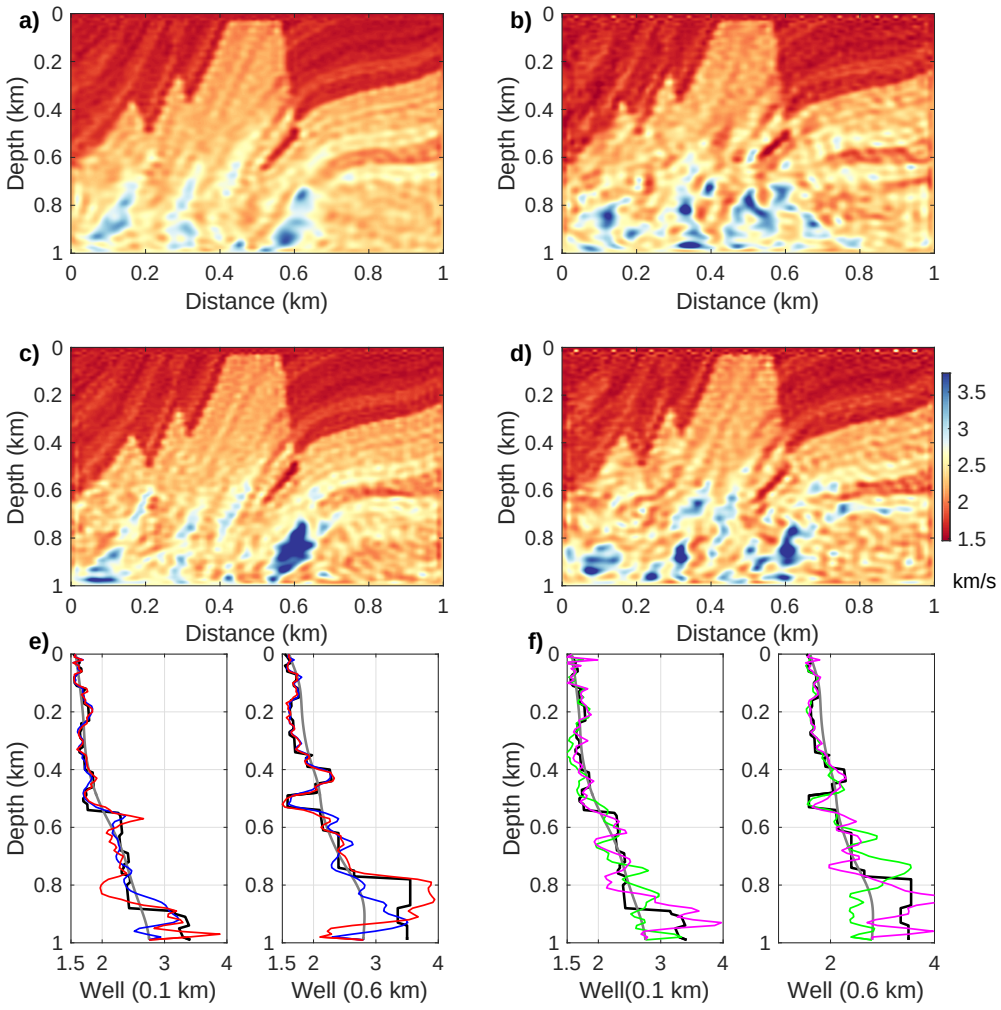

Figure 3.12: (a) and (b) show the non-preconditioned CG-GN inversion results with noisy data when SNR=10 (ϕ=0.20) and SNR=6 (ϕ=0.22); (c) and (d) show the l-BFGS-GN-PDGH inversion results when SNR=10 (ϕ=0.19) and SNR=6 (ϕ=0.26); (e) and (f) show the well log data comparison at 0.1 km and 0.6 km. The blue-dash-dot and green–dash-dot lines indicate the well log data by CG-GN method when SNR=10 and SNR=6. The red-dot and magenta-dot lines indicate the well log data by l-BFGS-GN-PDGH method when SNR=10 and SNR=6.

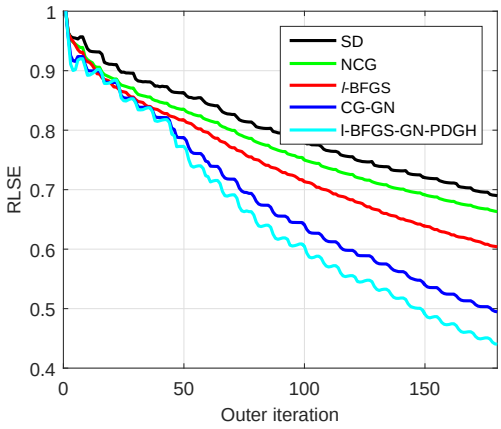

Figure 3.13: RLSE (equation (2.42)) vs. Iterations for the monoparameter example. The black, green, red, blue and cyan lines indicate the model errors for SD, NCG, l-BFGS, non-preconditioned CG-GN and l-BFGS-GN-PDGH methods respectively.

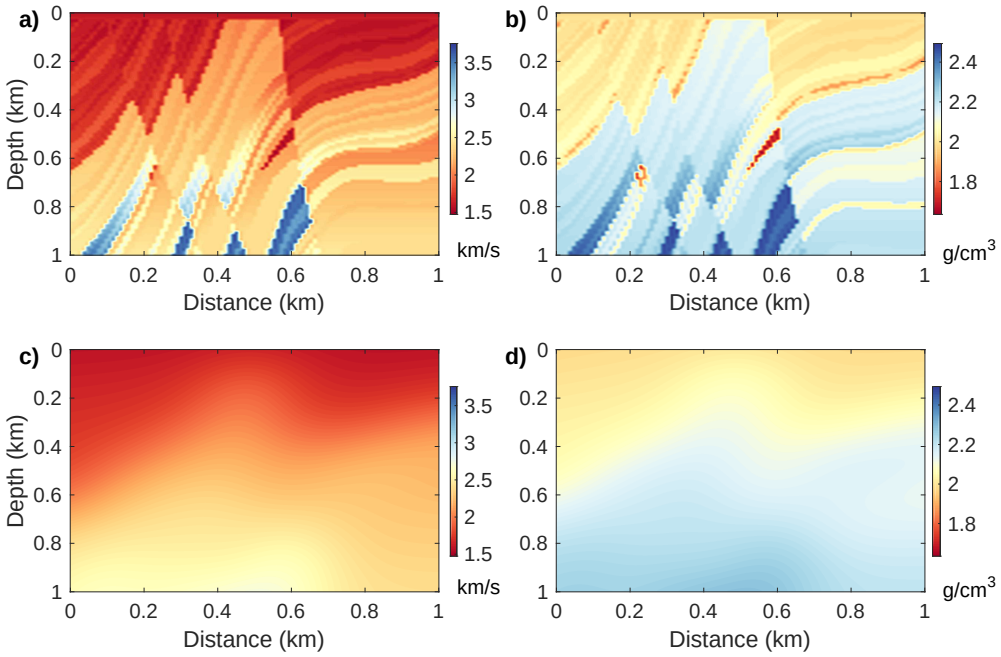

Figure 3.14: (a) and (b) show the true velocity and true density models; (c) and (d) show the initial velocity and initial density models.

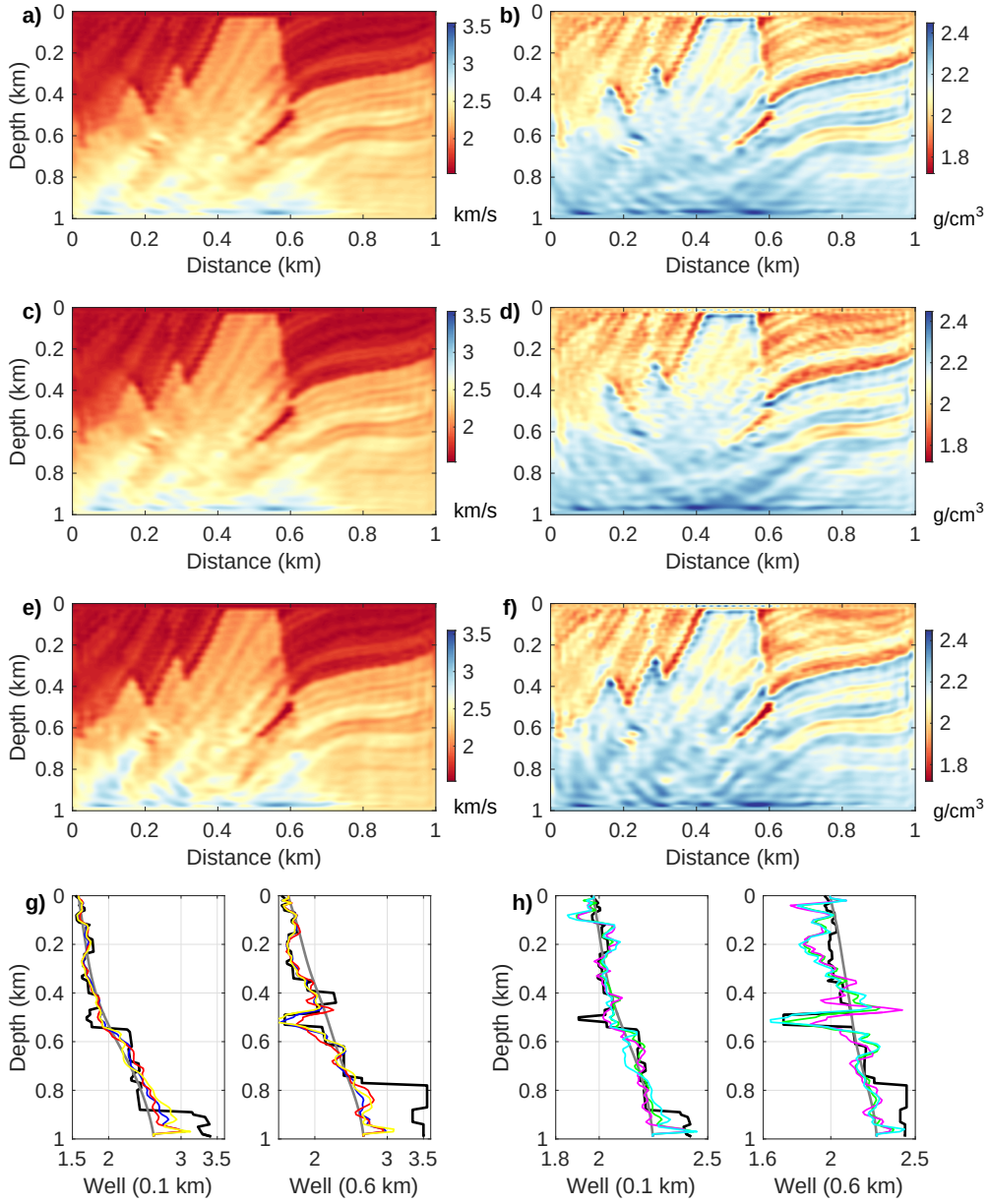

Figure 3.15: (a) and (b) show the inverted velocity and density models using SD method; (c) and (d) show inverted velocity and density models using NCG method; (e) and (f) show inverted velocity and density models using l-BFGS method; (g) and (h) show the well log comparison at 0.1 km and 0.6 km. The blue, red and yellow lines indicate the inverted velocity models by SD, NCG, and l-BFGS methods. The green, magenta and cyan lines indicate the inverted density models using SD, NCG and l-BFGS methods respectively.

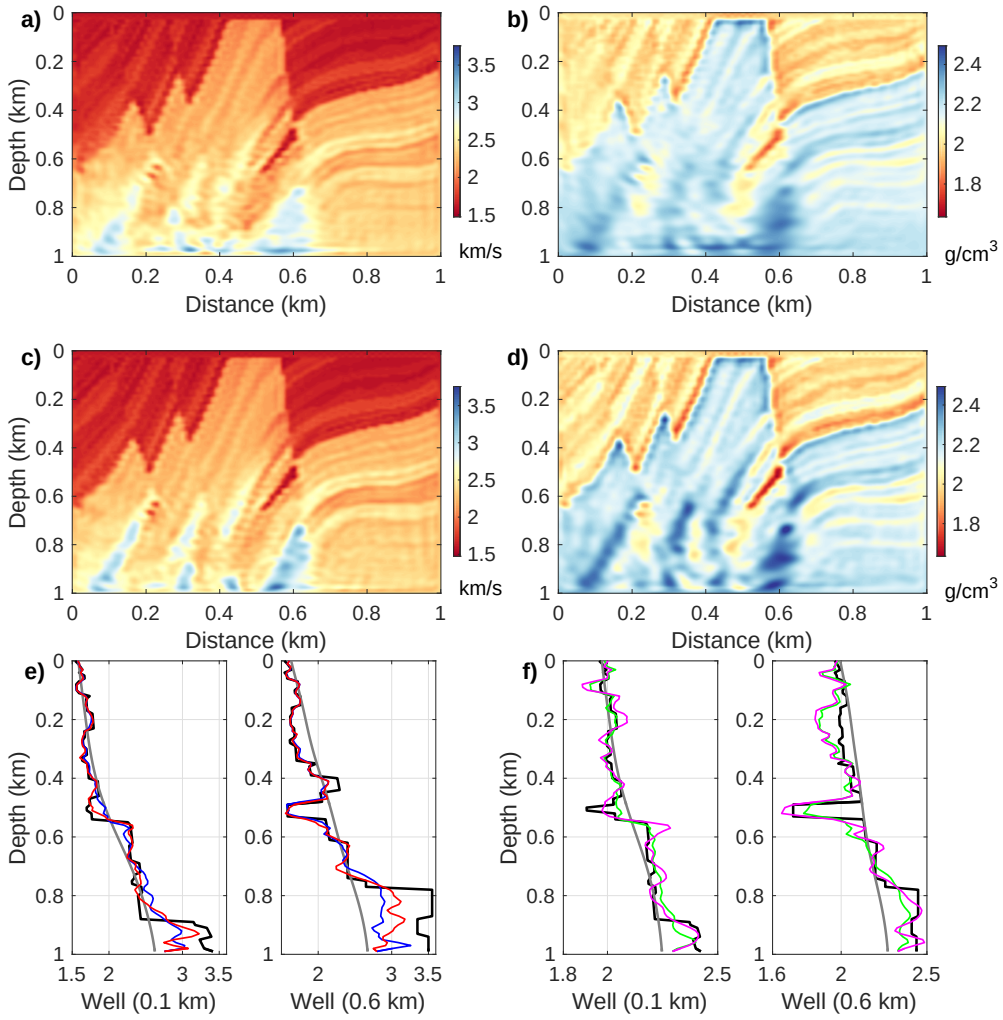

Figure 3.16: (a) and (b) show the inverted velocity and density models using non-preconditioned CG-GN method; (c) and (d) show inverted velocity and density models using l-BFGS-GN-PDGH method; (e) and (f) show the well log comparison at 0.1 km and 0.6 km. The blue and green lines indicate the inverted velocity and density by non-preconditioned CG-GN method. The red and magenta lines indicate the inverted velocity and density by l-BFGS-GN-PDGH method.

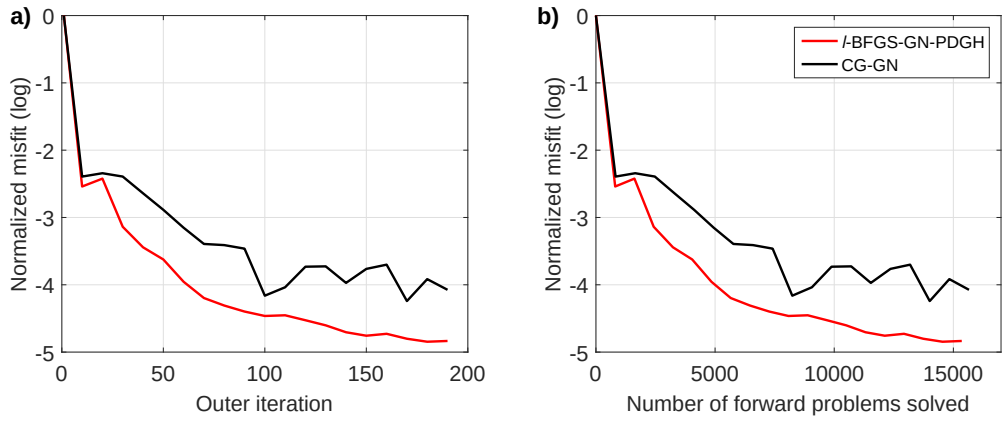

Figure 3.17: (a) Normalized misfit (log) vs. Iterations; (b) Normalized misfit (log) vs. Number of forward problems solved. The black and red lines indicate non-preconditioned CG-GN, and l-BFGS-GN-PDGH methods respectively. Note: here I plot the minimum normalized misfit of each frequency band.

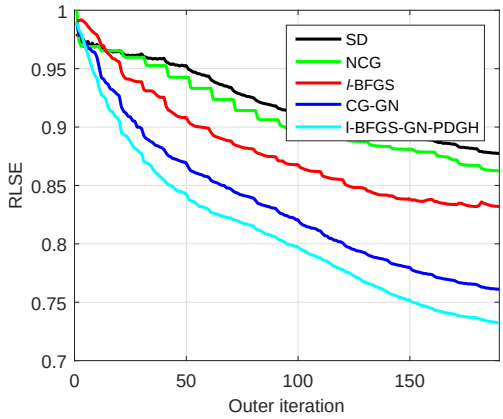

Figure 3.18: RLSE (equation (2.42)) vs. Iterations for the multiparameter example. The black, green, red, blue and cyan lines indicate the model errors for SD, NCG, l-BFGS, non-preconditioned CG-GN and l-BFGS-GN-PDGH methods respectively.

Chapter 4

Interparameter tradeoff quantification and reduction in isotropic-elastic FWI: synthetic experiments and Hussar dataset application

4.1 Summary

This chapter aims at evaluating and reducing the interparameter tradeoffs in isotropic-elastic FWI with multiparameter Hessian matrix-vector products. It is revealed that products of multiparameter Hessian off-diagonal blocks with model perturbation vectors, defined as *interparameter contamination kernels*, mainly account for interparameter tradeoff. The multiparameter Hessian is applied to various vectors designed to provide information regarding the strengths and characteristics of interparameter contaminations locally or within the whole volume. Based on these findings, a novel strategy is developed to mitigate the influence of interparameter tradeoffs with approximate contamination kernels. Furthermore, I propose to quantify resolution of the inverted models with approximate eigenvalue volume and extended multiparameter point spread functions (EMPSFs) by preconditioned conjugate-gradient algorithm. Finally, the proposed inversion strategies are applied to invert isotropic-elastic parameters with synthetic data and Hussar practical seismic dataset. Resolution of the inverted models are also evaluated.

4.2 Introduction

Elastic parameters are important for reservoir characterization. Simultaneously reconstructing multiple physical parameters suffers from interparameter tradeoffs arising from the inherent ambiguities among these parameters, which increases the nonlinearity and uncertainty of the inverse problems significantly (Tarantola, 1986; Köhn et al., 2012; Innanen, 2013; Operto et al., 2013; Alkhalifa and Plessix, 2014; Innanen, 2014b). This chapter aims at: (1) creating

more complete tools for quantifying the interparameter tradeoffs (or parameter crosstalk) than currently exist; (2) evaluating the strengths and characteristics of the interparameter contaminations in isotropic-elastic FWI by applying the multiparameter Hessian to various types of test vectors; (3) developing an effective way to reduce the influence of interparameter tradeoffs based on approximate contamination kernels; (4) quantifying local spatial and interparameter tradeoff of the inverted models with extended multiparameter point spread functions (EMPSFs).

Researchers have devoted intensive efforts to the study of parameter resolution based on analytic solutions of Fréchet derivative wavefields ("scattering" or "radiation" patterns) for different parameter classes (Tarantola, 1986; Gholami et al., 2013b; Alkhalifa and Plessix, 2014; Kamath and Tsvankin, 2014; Podgornova et al., 2015; Oh and Alkhalifah, 2016). Coupling effects appear between two different physical parameters, if the scattered wavefields due to the model perturbations overlap at certain range of scattering angles (Tarantola, 1986). A high-resolution parameterization should have scattering patterns as different as possible (Tarantola, 1986). Gholami et al. (2013a) investigated the scattering patterns of parameters resulting from various parameterizations of multiparameter acoustic FWI. Alkhalifa and Plessix (2014) emphasized the power of horizontal P-wave velocity in reducing the number of parameters for VTI FWI.

Amplitude variations of scattering patterns provide invaluable information for understanding the interparameter coupling effects but also ignore some important aspects due to a series of assumptions including incident plane-wave, homogeneous and isotropic-elastic background, high-frequency approximation, etc (Podgornova et al., 2015). These assumptions are regularly violated in seismic data sets, i.e., finite-frequency effects and traveltime information are not negligible in their influence on parameter resolution; heterogeneities should be considered; spatial correlations of different physical parameters are neglected (Alkhalifa and Plessix, 2014). Overlapping the scattering patterns due to different physical parameters in fact represents only an asymptotic approximation of the crosstalk quantification intrinsic to the Gauss-Newton Hessian (Operto et al., 2013). These limitations of the scattering patterns may result in misunderstandings concerning the interparameter tradeoffs. The problem of isotropic-elastic FWI has been investigated by many researchers (Mora, 1987; Brossier et al.,

2009; Köhn et al., 2012; Yuan and Simons, 2014; Borisov and Singh, 2015; Raknes and Arntsen, 2015; Modrak et al., 2016; Pan and Innanen, 2016a,c,b), but many challenges and open questions remain. Density structures are still poorly constrained, which may be caused by the weak sensitivity of traveltime to density variations and strong contaminations from velocity parameters. Some issues associated with the interparameter tradeoffs of isotropic-elastic parameters are actually not explained completely and clearly. Further unanswered questions include:

1. how do the interparameter tradeoffs affect the inversion process ?
2. how to evaluate the strengths and characteristics of the interparameter contaminations quantitatively ?
3. how to assess the uncertainties of the inverted models due to the interparameter tradeoffs ?

The first objective of this chapter is to evaluate the relative strengths and characteristics of interparameter contamination in isotropic-elastic FWI with multiparameter Hessian, which describes geometry of the objective function in terms of curvature or convexity (Fichtner and Trampert, 2011c; Fichtner and van Leeuwen, 2015). The diagonal blocks in the multiparameter Hessian characterize spatial correlations of the same physical parameter. Off-diagonal blocks measure correlations between different physical parameters (Fichtner and Trampert, 2011a; Operto et al., 2013). Rows in the multiparameter Hessian are averaging kernels (Backus and Gilbert, 1968) and columns are defined as multiparameter point spread functions (MPSFs) (Valenciano et al., 2006; Fichtner and Trampert, 2011c; Trampert et al., 2013; Tang and Lee, 2015; Zhu and Fomel, 2016). This chapter reveals that products of multiparameter Hessian off-diagonal blocks with the model perturbation vectors, which I will refer to as *interparameter contamination kernels*, account for the interparameter tradeoffs. For most large-scale inverse problems, explicitly constructing the Hessian matrix is considered to be computationally unaffordable. However, characteristics of the Hessian can be inferred via matrix probing techniques, which are useful when explicit representation of a matrix are too expensive to be constructed (Trampert et al., 2013). A low rank approximation of the Hessian can be efficiently computed by applying it to various types of vectors

(Halko et al., 2011; Demanet et al., 2012; An, 2012; Zhu et al., 2016; Rawlinson and Spakman, 2016). This chapter also examines the adjoint-state and finite-difference approaches for multiparameter Hessian matrix-vector product calculation. The product of the Hessian with a point-localized model perturbation vector preserves one Hessian column (Spakman, 1991). The MPSFs measure the relative strengths and finite-frequency features of the local interparameter tradeoffs. Furthermore, it is also shown (see below in this chapter) that S-wave velocity perturbations tend generally to produce strong contaminations into density update and phase-reversed contaminations within the P-wave velocity update, which may make density highly under- or overestimated and cancel the update for P-wave velocity.

To assess the interparameter tradeoffs within the whole volume of interest, MPSFs should be computed for each type of model parameter at every spatial position, which also results in prohibitive computation expense (Fichtner and Trampert, 2011c; Chen and Xie, 2015). Assuming that the multiparameter Hessian matrix is diagonally dominant, I have adopted a stochastic probing strategy by applying multiparameter Hessian to random vectors. Expectation values of the correlations between the random vector with its Hessian-vector products approximate Hessian diagonals (Sacchi et al., 2007; MacCarthy et al., 2011; Trampert et al., 2013). Arranging different random probes, the diagonals of multiparameter Hessian off-diagonal blocks, which measure the coupling strengths of different physical parameters in the whole volume, can be estimated stochastically. Stochastic estimations of the Hessian diagonals can also be used as preconditioners for acceleration (Modrak and Tromp, 2016).

Reducing the uncertainties introduced by the interparameter tradeoff is becoming essential for multiparameter FWI. Newton-based optimization methods are promising because they incorporate the inverse multiparameter Hessian with its ability to suppress the unwanted parameter crosstalk artifacts (Innanen, 2014b; Métivier et al., 2015; Wang et al., 2016; Yang et al., 2016). As stated in previous chapters, explicitly constructing and inverting multiparameter Hessian for large-scale inverse problems is, however, impracticably expensive as I have mentioned. Truncated-Newton (or Hessian-free) optimization methods represent affordable strategies for multiparameter FWI, in which the Newton equation is solved iteratively with matrix-free scheme of conjugate-gradient algorithm (Métivier et al., 2013; Boehm and Ulbrich, 2014; Métivier et al., 2015; Liu et al., 2015). However, itera-

tively solving the Newton equation is also expensive. Furthermore, by increasing savings through use of a small number of inner iterations, the effectiveness of removing interparameter mappings is reduced (Baumstein, 2014). Mode decomposition is a potential strategy for mitigating interparameter tradeoffs by isolating P and S wavefields but may also be limited in reducing the contaminations in density updates and multiparameter acoustic FWI (Wang and Cheng, 2017). Subspace optimization methods mitigate interparameter tradeoffs by scaling different physical parameters but do not prevent their occurrence (Kennett et al., 1988; Bernauer et al., 2014). In this chapter, based on a set of observations made on synthetic examples of interparameter tradeoff, a novel strategy is developed to reduce the interparameter tradeoff by approximating quantities I will refer to as *interparameter contamination kernels*. This strategy approximates the parameter contamination in model space by applying multiparameter Hessian off-diagonal blocks to estimated model vectors. The result is a model estimate which is approximately free of parameter crosstalk, and which has been created without iteratively solving large Newton systems, and which is in principle applicable to any tomographic or FWI misfit function. Numerical examples are given to illustrate that this new strategy is able to remove the contaminations from S-wave velocity partially and provide more reliable density estimations in isotropic-elastic FWI.

In addition to suppressing interparameter contaminations, parameter resolution quantification is key to a well-posed inverse scheme; it has been investigated by many researchers (Backus and Gilbert, 1968; Spakman, 1991; Fichtner and Trampert, 2011c; Rawlinson et al., 2014; Rawlinson and Spakman, 2016; Zhu et al., 2016). Within a Bayesian inference framework, uncertainties of the maximum a posterior model are evaluated based on posterior covariance operator which has a direct relationship with the (inverse) Hessian (Gouveia and Scales, 1998; Tarantola, 2005; Dettmer et al., 2007; Fichtner and Trampert, 2011c; Flath et al., 2011). In recent years, researchers have evaluated the local resolution of the inverted model with point spread functions by applying multiparameter Hessian to Gaussian-shape model perturbations (Fichtner and Trampert, 2011c; Rickers et al., 2013; Zhu et al., 2015; Bozdağ et al., 2016). However, the point spread functions actually represent conservative estimations of columns in resolution matrix by approximating inverse Hessian as an identity matrix (Oldenborger and Routh, 2009; Fichtner and van Leeuwen, 2015). The similarities

and differences of the Hessian matrix and resolution matrix in resolution analysis are investigated. Approximate eigenvalue volumes are used to evaluate resolution of inverted models within the whole volume. Local spatial and interparameter tradeoffs of the inverted models are quantified with extended multiparameter point spread functions (EMPSFs) by applying the approximate inverse Hessian to the traditional MPSFs iteratively with preconditioned conjugate-gradient algorithm. The approximate inverse Hessian will de-blur the MPSFs further, balance the relative magnitudes by compensating geometrical spreading and mitigate interparameter contaminations, which represent more accurate local measurements of spatial and interparameter tradeoffs.

This chapter first reviews the basic principles of isotropic-elastic FWI. Benefits and limitations of the parameter resolution studies based upon scattering patterns are explored. *Interparameter contamination kernels* are then defined and how to evaluate the interparameter tradeoffs with multiparameter Hessian-vector products is explained. An explanation of the novel inversion strategy in reducing the interparameter tradeoffs with approximate contamination kernels is then given. Strategies for quantifying the resolution of the inverted models with approximate eigenvalue volumes and extended multiparameter point spread functions (EMPSFs) are explained. In the numerical modelling section, the strengths and characteristics of local interparameter tradeoffs are first examined with multiparameter point spread functions. Proposed matrix probing techniques are applied on a complex Marmousi model to assess the interparameter tradeoffs within the whole volume. The new inversion strategy is also applied to invert isotropic-elastic parameters with synthetic data and the low-frequency Hussar field seismic data set acquired by the CREWES Project and collaborators in 2011 (Margrave et al., 2012). The approximate eigenvalue volumes and EMPSFs are used to quantify resolution of the inverted models. The performances of various parameterizations for reconstructing subsurface isotropic-elastic properties are also examined with synthetic examples. Note: in this chapter, the expressions of gradients, Hessian-vector products, etc are given in time domain with integral formulas, whereas in previous chapters for convenience discrete formulas were used.

4.3 Methodology

4.3.1 Isotropic-elastic full-waveform inversion

The common l-2 norm misfit function in time domain can be expressed as:

$$\Phi(\mathbf{m}) = \sum_{\mathbf{x}_s} \sum_{\mathbf{x}_g} \int_0^T \| \Delta \mathbf{d}(\mathbf{x}_s, \mathbf{x}_g, t; \mathbf{m}) \|^2 dt, \tag{4.1}$$

where $\Delta \mathbf{d}(\mathbf{x}_s, \mathbf{x}_g, t; \mathbf{m}) = \mathbf{d}_{\text{syn}}(\mathbf{x}_s, \mathbf{x}_g, t) - \mathbf{d}_{\text{obs}}(\mathbf{x}_s, \mathbf{x}_g, t; \mathbf{m})$ is the data residual, $\mathbf{x}_s$ ($s = 1, ..., S$) and $\mathbf{x}_g$ ($g = 1, ..., R$) indicate source and receiver locations, S and R are the maximum source and receiver indexes, and T represents maximum recording time. In order to solve the inverse problem and find the model which minimizes the adopted cost function, the model is updated iteratively. The gradient of the misfit function can be obtained by correlating the Fréchet derivative wavefield with data residual:

$$\nabla_{\mathbf{m}} \Phi(\mathbf{m}) = \sum_{\mathbf{x}_s} \sum_{\mathbf{x}_g} \int_0^T \int_{\Omega(\mathbf{x})} \nabla_{\mathbf{m}(\mathbf{x})} \mathbf{u}^{\star}(\mathbf{x}_s, \mathbf{x}_g, t; \mathbf{m}) \, \Delta \mathbf{d}(\mathbf{x}_s, \mathbf{x}_g, t; \mathbf{m}) \, d\mathbf{x} dt, \tag{4.2}$$

where $\nabla_{\mathbf{m}(\mathbf{x})} \mathbf{u}(\mathbf{x}_s, \mathbf{x}_g, t; \mathbf{m})$ indicates the Fréchet derivative wavefield, Ω indicates the whole volume, and the symbol $\star$ means complex conjugate transpose. Considering general anisotropic-elastic media, based on Born approximation the perturbed nth displacement field due to model perturbation $\Delta \mathbf{m}_{\rho}$ and $\Delta \mathbf{m}_{c_{ijkl}}$ is expressed as:

$$\begin{aligned} \Delta u_n(\mathbf{x}_s, \mathbf{x}_g, t; \Delta \mathbf{m}) = - \int_{\Omega(\mathbf{x})} \int_0^t [& \Delta m_{\rho}(\mathbf{x}) \, G_{ni}(\mathbf{x}, \mathbf{x}_r, t - t') \, \partial_t^2 u_i(\mathbf{x}, \mathbf{x}_s, t') \\ & + \Delta m_{c_{ijkl}}(\mathbf{x}) \, \partial_j G_{ni}(\mathbf{x}, \mathbf{x}_r, t - t') \, \partial_k u_l(\mathbf{x}, \mathbf{x}_s, t')] dt' d\mathbf{x}, \end{aligned} \tag{4.3}$$

where ρ and c_{ijkl} (i, j, k, l take on the values of x, y, z) denote density and elastic constant tensor with Einstein summation convention, G_{ni} is the Green's tensor, the nth displacement response due to impulse source at the ith direction. For isotropic-elastic media, the perturbation of elastic constants can be expressed in terms of the perturbations of bulk modulus $\Delta \mathbf{m}_{\kappa}$ and shear modulus $\Delta \mathbf{m}_{\mu}$: $\Delta m_{c_{ijkl}}(\mathbf{x}) = (\Delta m_{\kappa}(\mathbf{x}) - 2/3 \Delta m_{\mu}(\mathbf{x})) \, \delta_{ij} \delta_{kl} + \Delta m_{\mu}(\mathbf{x}) (\delta_{ik} \delta_{jl} + \delta_{jk} \delta_{il})$. Wavefield perturbation due to the perturbations of isotropic-elastic

parameters are expressed as:

$$\begin{aligned}\Delta u_n\left(\mathbf{x}_s,\mathbf{x}_g,t;\Delta\mathbf{m}\right)=&-\int_{\Omega(\mathbf{x})}\int_0^t[\Delta m_\rho\left(\mathbf{x}\right)G_{ni}\left(\mathbf{x},\mathbf{x}_r,t-t'\right)\partial_t^2u_i\left(\mathbf{x},\mathbf{x}_s,t'\right)\\&+\left(\Delta m_\kappa\left(\mathbf{x}\right)-\frac{2}{3}\Delta m_\mu\left(\mathbf{x}\right)\right)\delta_{ij}\delta_{kl}\partial_jG_{ni}\left(\mathbf{x},\mathbf{x}_r,t-t'\right)\partial_ku_l\left(\mathbf{x},\mathbf{x}_s,t'\right)\\&+\Delta m_\mu\left(\mathbf{x}\right)\left(\delta_{ik}\delta_{jl}+\delta_{jk}\delta_{il}\right)\partial_jG_{ni}\left(\mathbf{x},\mathbf{x}_r,t-t'\right)\partial_ku_l\left(\mathbf{x},\mathbf{x}_s,t'\right)]dt'd\mathbf{x}.\end{aligned}\tag{4.4}$$

Substituting equation (4.4) into equation (4.5) gives the Fréchet derivative of the misfit function:

$$\nabla_{\mathbf{m}}\Phi\left(\mathbf{m}\right)=\int_{\Omega(\mathbf{x})}[K_\kappa\left(\mathbf{x}\right)a_\kappa\left(\mathbf{x}\right)+K_\mu\left(\mathbf{x}\right)a_\mu\left(\mathbf{x}\right)+K_\rho\left(\mathbf{x}\right)a_\rho\left(\mathbf{x}\right)]d\mathbf{x},\tag{4.5}$$

where K_κ, K_μ and K_ρ represent sensitivity kernels with respect to bulk modulus κ, shear modulus μ and density ρ, $a_\kappa=\Delta\mathbf{m}_\kappa/\mathbf{m}_\kappa$, $a_\mu=\Delta\mathbf{m}_\mu/\mathbf{m}_\mu$ and $a_\rho=\Delta\mathbf{m}_\rho/\mathbf{m}_\rho$ are relative model perturbations. Explicit expressions of the sensitivity kernels for these isotropic-elastic parameters can be written as (Tromp et al., 2005; Liu et al., 2006; Zhu et al., 2009; Luo et al., 2013; Yuan and Simons, 2014):

$$K_\kappa\left(\mathbf{x}\right)=-\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_0^T\kappa\left(\mathbf{x}\right)\partial_i\tilde{u}_i\left(\mathbf{x}_g,\mathbf{x},T-t\right)\partial_ku_k\left(\mathbf{x},\mathbf{x}_s,t\right)dt,\tag{4.6}$$

$$\begin{aligned}K_\mu\left(\mathbf{x}\right)=&-\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_0^T\mu\left(\mathbf{x}\right)[\partial_j\tilde{u}_i\left(\mathbf{x}_g,\mathbf{x},T-t\right)\left(\partial_iu_j\left(\mathbf{x},\mathbf{x}_s,t\right)+\partial_ju_i\left(\mathbf{x},\mathbf{x}_s,t\right)\right)\\&-\frac{2}{3}\partial_i\tilde{u}_i\left(\mathbf{x}_g,\mathbf{x},T-t\right)\partial_ku_k\left(\mathbf{x},\mathbf{x}_s,t\right)]dt,\end{aligned}\tag{4.7}$$

$$K_\rho\left(\mathbf{x}\right)=-\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_0^T\rho\left(\mathbf{x}\right)\tilde{u}_i\left(\mathbf{x}_g,\mathbf{x},T-t\right)\partial_t^2u_i\left(\mathbf{x},\mathbf{x}_s,t\right)dt,\tag{4.8}$$

where $\tilde{u}_i\left(\mathbf{x}_g,\mathbf{x},T-t\right)$ represents the ith component of the adjoint wavefield:

$$\tilde{u}_i\left(\mathbf{x}_g,\mathbf{x},T-t\right)=\int_0^{T-t}G_{in}\left(\mathbf{x}_g,\mathbf{x},T-t-t'\right)\tilde{f}_n\left(\mathbf{x},t'\right)dt',\tag{4.9}$$

where $\tilde{f}_n\left(\mathbf{x},t'\right)$ is the adjoint source (Tromp et al., 2005; Bozdag et al., 2011):

$$\tilde{f}_n\left(\mathbf{x},t'\right)=\sum_{\mathbf{x}_g}\Delta d_n\left(\mathbf{x}_g,T-t'\right)\delta\left(\mathbf{x}-\mathbf{x}_g\right).\tag{4.10}$$

With velocity-density parameterization, the corresponding sensitivity kernels for P-velocity α, S-wave velocity β and density ρ' are given by (Tromp et al., 2005; Köhn et al., 2012; Yuan

et al., 2015):

$$K_\alpha = 2\left(1+\frac{4}{3}\frac{\mu}{\kappa}\right)K_\kappa, K_\beta = 2\left(K_\mu - \frac{4}{3}\frac{\mu}{\kappa}K_\kappa\right), K_{\rho'} = K_\rho + K_\kappa + K_\mu. \tag{4.11}$$

For impedance-density parameterization, sensitivity kernels of P-wave impedance IP=$\alpha\rho''$, S-wave impedance IS=$\beta\rho''$ and density ρ'' are given by:

$$\begin{aligned} K_{\mathrm{IP}} &= 2\left(1+\frac{4}{3}\frac{\mu}{\kappa}\right)K_\kappa = K_\alpha, \\ K_{\mathrm{IS}} &= 2\left(K_\mu - \frac{4}{3}\frac{\mu}{\kappa}K_\kappa\right) = K_\beta, \\ K_{\rho''} &= -K_\kappa - K_\mu + K_\rho = -K_\alpha - K_\beta + K_{\rho'}. \end{aligned} \tag{4.12}$$

Matrix multiplication of Newton equation system can be written with an integral formulation:

$$\nabla_{\mathbf{m}}\Phi(\mathbf{x}) = -\int_{\Omega(\mathbf{x}')} H(\mathbf{x},\mathbf{x}')\,\Delta m(\mathbf{x}')\,d\mathbf{x}', \tag{4.13}$$

where $H(\mathbf{x},\mathbf{x}')$ denotes one Hessian element described by positions $\mathbf{x}$ and $\mathbf{x}'$. In this chapter, to update the isotropic-elastic parameters simultaneously, a quasi-Newton l-BFGS optimization method is used. At each iteration, a line search approach is employed to obtain the step length for updating the model (Nocedal and Wright, 2006; Yuan et al., 2015).

4.3.2 Physical interpretation of multiparameter Hessian

In multiparameter FWI, the multiparameter Hessian has a block structure. For general anisotropic-elastic media, Hessian $\mathbf{H}$ can be expressed as:

$$\begin{aligned} \mathbf{H} = \int_{\Omega(\mathbf{x})}\int_{\Omega(\mathbf{x}')} [&\Delta m_\rho(\mathbf{x})\,H_{\rho\rho}(\mathbf{x},\mathbf{x}')\,\Delta m_\rho(\mathbf{x}') + \Delta m_\rho(\mathbf{x})\,H_{\rho\mathbf{c}}(\mathbf{x},\mathbf{x}')\,\mathbf{::}\Delta m_{\mathbf{c}}(\mathbf{x}') \\ &+ \Delta m_{\mathbf{c}}(\mathbf{x})\,\mathbf{::}H_{\mathbf{c}\rho}(\mathbf{x},\mathbf{x}')\,\Delta m_\rho(\mathbf{x}') + \Delta m_{\mathbf{c}}(\mathbf{x})\,\mathbf{::}H_{\mathbf{c}\rho}(\mathbf{x},\mathbf{x}')\,\mathbf{::}\Delta m_{\mathbf{c}}(\mathbf{x}')]\,d\mathbf{x}d\mathbf{x}', \end{aligned} \tag{4.14}$$

where $\Delta\mathbf{m}_{\mathbf{c}}$ indicates the perturbation of elastic constant tensor $\mathbf{c}$ and $\mathbf{::}$ means sequential contractions over the four nearest tensor indices (Luo, 2012). Because $\mathbf{H}$ is symmetric, then $\mathbf{H}_{\rho\mathbf{c}} = \mathbf{H}_{\mathbf{c}\rho}^{\dagger}$. Explicit expressions of diagonal blocks $\mathbf{H}_{\rho\rho}$ and $\mathbf{H}_{\mathbf{cc}}$ and off-diagonal block $\mathbf{H}_{\mathbf{c}\rho}$ are given by:

$$\begin{aligned} H_{\rho\rho}(\mathbf{x},\mathbf{x}') = &\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int\int \partial_{t'}^2 u_i(\mathbf{x},\mathbf{x}_s,t')\,G_{ni}(\mathbf{x}_g,\mathbf{x},T-t') \\ &\times G_{n'i'}(\mathbf{x}_g,\mathbf{x}',t-t')\,\partial_{t''}^2 u_{i'}(\mathbf{x}',\mathbf{x}_s,t'')\,dt'dt'', \end{aligned} \tag{4.15}$$

$$H_{\mathbf{cc}}(\mathbf{x},\mathbf{x}') = \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int\int \partial_k u_l(\mathbf{x},\mathbf{x}_s,t')\,\partial_j G_{ni}(\mathbf{x}_g,\mathbf{x},T-t') \times \partial_{j'} G_{n'i'}(\mathbf{x}_g,\mathbf{x}',t-t'')\,\partial_{k'} u_{l'}(\mathbf{x}',\mathbf{x}_s,t'')\,dt'dt'', \tag{4.16}$$

$$H_{\mathbf{c}\rho}(\mathbf{x},\mathbf{x}') = \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int\int \partial_k u_l(\mathbf{x},\mathbf{x}_s,t')\,\partial_j G_{ni}(\mathbf{x}_g,\mathbf{x},T-t') \times G_{n'i'}(\mathbf{x}_g,\mathbf{x}',t-t'')\,\partial_{t''}^2 u_{i'}(\mathbf{x}',\mathbf{x}_s,t'')\,dt'dt''. \tag{4.17}$$

For velocity-density parameterization in isotropic-elastic FWI, the Newton equation system for simultaneously updating P-wave velocity α, S-wave velocity β and density ρ' can be written as:

$$\begin{bmatrix} \mathbf{H}_{\alpha\alpha} & \mathbf{H}_{\alpha\beta} & \mathbf{H}_{\alpha\rho'} \\ \mathbf{H}_{\beta\alpha} & \mathbf{H}_{\beta\beta} & \mathbf{H}_{\beta\rho'} \\ \mathbf{H}_{\rho'\alpha} & \mathbf{H}_{\rho'\beta} & \mathbf{H}_{\rho'\rho'} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{m}_\alpha \\ \Delta\mathbf{m}_\beta \\ \Delta\mathbf{m}_{\rho'} \end{bmatrix} = - \begin{bmatrix} \nabla_\alpha \Phi \\ \nabla_\beta \Phi \\ \nabla_{\rho'} \Phi \end{bmatrix}, \tag{4.18}$$

where $\nabla_\alpha\Phi$, $\nabla_\beta\Phi$, and $\nabla_{\rho'}\Phi$ are gradient vectors of α, β and ρ' respectively. Multiparameter Hessian elements can be classified into 4 types: (A) diagonal elements of the diagonal blocks account for geometrical spreading (i.e., $H_{\alpha\alpha}(\mathbf{x},\mathbf{x})$); (B) off-diagonal elements of the diagonal blocks measure the spatial correlations of model parameters with the same physical signature (i.e., $H_{\alpha\alpha}(\mathbf{x},\mathbf{x}')$ with $\mathbf{x}\neq\mathbf{x}'$); (C) diagonals of off-diagonal blocks indicate the strength of interparameter coupling at the same location (i.e., $H_{\alpha\beta}(\mathbf{x},\mathbf{x})$); (D) off-diagonal elements of off-diagonal blocks describe both spatial and interparameter tradeoffs (i.e., $H_{\alpha\beta}(\mathbf{x},\mathbf{x}')$ with $\mathbf{x}\neq\mathbf{x}'$). One column of the multiparameter Hessian describes the blurring of an input delta function by the inverse operator, which is defined as multiparameter point spread function (MPSF) (Fichtner and Trampert, 2011c; Trampert et al., 2013; Fichtner and van Leeuwen, 2015; Tang and Lee, 2015; Zhu and Fomel, 2016). For example, the column $\mathbf{H}_\beta(\mathbf{x},\mathbf{x}_N)$ indicates the correlation model parameter β at position $\mathbf{x}_N$ with model parameters α, β and ρ' at all positions in the whole volume.

4.3.3 Parameter resolution analysis with scattering patterns

In recent years, researchers have made great efforts to understand the resolving abilities of different parameterizations in multiparameter FWI based on "scattering" (or "radiation") patterns (Tarantola, 1986; Gholami et al., 2013b; Operto et al., 2013; Alkhalifa and Plessix, 2014; Kamath and Tsvankin, 2014; Podgornova et al., 2015). Scattering pattern represents

the amplitude variations of Fréchet derivative wavefields with varying scattering or azimuthal angles. Parameter ambiguity appears when the scattering patterns duo to different physical parameters overlap at a certain range of scattering angle. Inversion sensitivity studies play a crucial role in characterizing interparameter tradeoffs, designing optimal parameterization and acquisition geometry for multiparameter FWI (Tarantola, 1986; Gholami et al., 2013b; Kamath and Tsvankin, 2014; Rusmanugroho et al., 2017).

4.3.3.1 Benefits and limitations of scattering patterns

Scattering patterns provide us with an efficient way to evaluate the coupling effects between different physical parameters. In Appendix B, the scattering coefficients of isotropic-elastic parameters within various parameterizations following the framework of scattering potentials developed by Stolt and Weglein (2012) are given. Figure 4.1 shows the scattering patterns due to the isotropic-elastic parameters, which provide important information for us to understand the parameter coupling effects with different wave modes. For example, subsurface density heterogeneities are essentially important for fluid reservoir characterization but are still poorly constrained in FWI, which may be caused by strong contaminations from velocity parameters (Kuo and Romanowicz, 2002; Xu and McMechan, 2014; Yuan et al., 2015) and the weak sensitivity of traveltime observations to density variations (Plonka et al., 2016). From the scattering patterns, it is observed that the scattered waves due to density perturbations propagate backward, which means that in transmission tomography, it will be very challenging to recover density properties. Examining P-P scattering patterns informs us that distinguishing P-wave velocity and density diffractors at near offset will be difficult.

However, scattering patterns may not be enough to evaluate the interparameter contaminations completely because of their oversimplification of the wave-medium interaction. First, the scattering patterns are derived based on the assumption that an incident plane-wave is scattered due to a local point heterogeneity or a horizontal reflector embedded in an isotropic-elastic and homogeneous background. This means that the complex physical background model and geometrical spreading effects are ignored. Second, scattering patterns only represent the amplitude variations of the Fréchet derivative wavefields. The influences of traveltime and finite-frequencies on interparameter tradeoffs should also be considered.

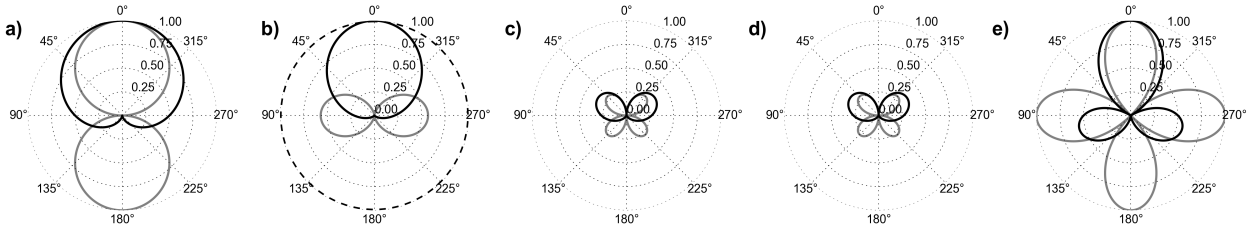

Figure 4.1: Figures (a-e) show the SH-SH, P-P, P-SV, SV-P and SV-SV scattering patterns due to the perturbations of isotropic-elastic parameters. Solid-grey and solid-black curves indicate the scattering patterns due to perturbations of S-wave velocity and density. In (a), the dash-black curve represents the P-P scattering pattern due the perturbation of P-wave velocity.

Furthermore, analysis based on scattering patterns also neglect the spatial correlations of different physical parameters (Alkhalifa and Plessix, 2014), which actually are very strong when using low frequencies for inversion or in transmission tomography. These defects of scattering patterns may result in misleading conclusions about the interparameter tradeoffs.

4.3.4 Quantifying interparameter tradeoffs via multiparameter Hessian probing

First this section shows that the unwanted interparameter tradeoff artifacts can be described by *interparameter contamination kernels* defined as products of multiparameter Hessian off-diagonal blocks with model perturbation vectors. Thus, the interparameter tradeoffs in isotropic-elastic FWI can be quantified by probing the multiparameter Hessian with various test vectors.

4.3.4.1 *Interparameter contamination kernels*

Interparameter contamination kernels may be first introduced starting from standard sensitivity kernels. According to equation (4.11), the sensitivity kernel $K_\alpha(\mathbf{x})$ is written explicitly as:

$$K_\alpha(\mathbf{x}) = -\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_0^T 2\rho'\alpha^2\partial_k u_k(\mathbf{x}_s,\mathbf{x},t)\int_0^{T-t}\partial_i G_{in}(\mathbf{x},\mathbf{x}_g,T-t-t')\,\tilde{f}_n(\mathbf{x},t')\,dt'dt, \tag{4.19}$$

where the adjoint source $\tilde{f}_n$ can be decomposed into three parts due to perturbations of $\Delta\mathbf{m}_\alpha$, $\Delta\mathbf{m}_\beta$ and $\Delta\mathbf{m}_{\rho'}$ respectively:

$$\tilde{f}_n\left(\mathbf{x},t';\Delta\mathbf{m}\right)=\tilde{f}_n\left(\mathbf{x},t';\Delta\mathbf{m}_\alpha\right)+\tilde{f}_n\left(\mathbf{x},t';\Delta\mathbf{m}_\beta\right)+\tilde{f}_n\left(\mathbf{x},t';\Delta\mathbf{m}_{\rho'}\right). \tag{4.20}$$

Ignoring multiple scattering components in the data residuals and following equations (4.4) and (4.10), the three adjoint sources in equation (4.20) can be expressed as:

$$\tilde{f}_{n'}\left(\mathbf{x},t';\Delta\mathbf{m}_\alpha\right)=\langle 2\rho'\alpha^2\partial_{i'}G_{i'n'}\left(\mathbf{x}'\right)\Delta m_\alpha\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\delta\left(\mathbf{x}-\mathbf{x}_g\right)\rangle, \tag{4.21}$$

$$\begin{aligned}\tilde{f}_{n'}\left(\mathbf{x},t';\Delta\mathbf{m}_\beta\right)=\langle 2\rho'\beta^2\left[\partial_{j'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_\beta\left(\mathbf{x}'\right)\left(\partial_{i'}u_{j'}\left(\mathbf{x}'\right)+\partial_{j'}u_{i'}\left(\mathbf{x}'\right)\right)\right.\\ \left.-2\partial_{i'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_\beta\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right]\delta\left(\mathbf{x}-\mathbf{x}_g\right)\rangle,\end{aligned} \tag{4.22}$$

$$\begin{aligned}\tilde{f}_{n'}\left(\mathbf{x},t';\Delta\mathbf{m}_{\rho'}\right)=\langle\rho'\left[\left(G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{t'}^2u_{i'}\left(\mathbf{x}'\right)+2\alpha^2\partial_{i'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right)\right.\\ +2\beta^2\left(\partial_{j'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\left(\partial_{i'}u_{j'}\left(\mathbf{x}'\right)+\partial_{j'}u_{i'}\left(\mathbf{x}'\right)\right)\right.\\ \left.\left.-2\partial_{i'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right)\right]\delta\left(\mathbf{x}-\mathbf{x}_g\right)\rangle,\end{aligned} \tag{4.23}$$

where the symbol $\langle\cdot\rangle$ indicates summation over sources, receivers, time and positions for sake of compactness. Inserting equations (4.21), (4.22) and (4.23) into equation (4.19) partitions the standard sensitivity kernel K_α into:

$$K_\alpha=K_{\alpha\leftrightarrow\alpha}+K_{\beta\to\alpha}+K_{\rho'\to\alpha}, \tag{4.24}$$

where the first term $K_{\alpha\leftrightarrow\alpha}$ represents the correct update kernel for α, and the second and third terms $K_{\beta\to\alpha}$ and $K_{\rho'\to\alpha}$ are defined as *interparameter contamination kernels* (Pan et al., 2017d,b,c), which represent the contaminations from β and ρ' to α:

$$K_{\alpha\leftrightarrow\alpha}\left(\mathbf{x}\right)=-\left\langle 2\rho'\alpha^2\partial_k u_k\left(\mathbf{x}\right)\partial_i G_{in}\left(\mathbf{x}\right)\left[2\rho'\alpha^2\partial_{i'}G_{i'n'}\left(\mathbf{x}'\right)\Delta m_\alpha\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right]\right\rangle, \tag{4.25}$$

$$\begin{aligned}K_{\beta\to\alpha}\left(\mathbf{x}\right)=-\left\langle 2\rho'\alpha^2\partial_k u_k\left(\mathbf{x}\right)\partial_i G_{in}\left(\mathbf{x}\right)2\rho'\beta^2\left[\partial_{j'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_\beta\left(\mathbf{x}'\right)\left(\partial_{i'}u_{j'}\left(\mathbf{x}'\right)\right.\right.\right.\\ \left.\left.\left.+\partial_{j'}u_{i'}\left(\mathbf{x}'\right)\right)-2\partial_{i'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_\beta\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right]\right\rangle,\end{aligned} \tag{4.26}$$

$$\begin{aligned}K_{\rho'\to\alpha}\left(\mathbf{x}\right)=-\left\langle 2\rho'\alpha^2\partial_k u_k\left(\mathbf{x}\right)\partial_i G_{in}\left(\mathbf{x}\right)\rho'\left[\left(G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{t'}^2u_{i'}\left(\mathbf{x}'\right)+2\alpha^2\partial_{i'}G_{n'i'}\left(\mathbf{x}'\right)\right.\right.\right.\\ \left.\times\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right)+2\beta^2\left(\partial_{j'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\left(\partial_{i'}u_{j'}\left(\mathbf{x}'\right)+\partial_{j'}u_{i'}\left(\mathbf{x}'\right)\right)\right.\\ \left.\left.\left.-2\partial_{i'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right)\right]\right\rangle.\end{aligned} \tag{4.27}$$

Interparameter contamination kernels can also be explained and obtained with Newton equation (4.18). The gradient vector $\nabla_\alpha \Phi(\mathbf{x})$ in equation (4.18) can be written as an integral formulation:

$$\begin{aligned}\nabla_\alpha \Phi(\mathbf{x}) = &-\int_{\Omega(\mathbf{x}')} H_{\alpha\alpha}(\mathbf{x},\mathbf{x}')\,\Delta m_\alpha(\mathbf{x}')\,d\mathbf{x}' \\ &-\int_{\Omega(\mathbf{x}')} H_{\alpha\beta}(\mathbf{x},\mathbf{x}')\,\Delta m_\beta(\mathbf{x}')\,d\mathbf{x}' \\ &-\int_{\Omega(\mathbf{x}')} H_{\alpha\rho'}(\mathbf{x},\mathbf{x}')\,\Delta m_{\rho'}(\mathbf{x}')\,d\mathbf{x}',\end{aligned} \tag{4.28}$$

where model perturbation vectors $\Delta\mathbf{m}_\beta$ and $\Delta\mathbf{m}_{\rho'}$ blurred by off-diagonal blocks $\mathbf{H}_{\alpha\beta}$ and $\mathbf{H}_{\alpha\rho'}$ in multiparameter Hessian are mapped into the update for parameter α. Equation (4.28) is equivalent to equation (4.24). Products of multiparameter Hessian block matrices with the model perturbation vectors are equivalent to the correct update kernel $K_{\alpha\leftrightarrow\alpha}$ and *interparameter contamination kernels* $K_{\beta\to\alpha}$ and $K_{\rho'\to\alpha}$ in equation (4.28). Similarly, gradient vectors $\nabla_\beta\Phi$ and $\nabla_{\rho'}\Phi$ can be written as:

$$\begin{aligned}\nabla_\beta \Phi(\mathbf{x}) &= a_\beta\left(K_{\alpha\to\beta}(\mathbf{x}) + K_{\beta\leftrightarrow\beta}(\mathbf{x}) + K_{\rho'\to\beta}(\mathbf{x})\right) \\ &= -\int_{\Omega(\mathbf{x}')} H_{\beta\alpha}(\mathbf{x},\mathbf{x}')\,\Delta m_\alpha(\mathbf{x}')\,d\mathbf{x}' \\ &\quad -\int_{\Omega(\mathbf{x}')} H_{\beta\beta}(\mathbf{x},\mathbf{x}')\,\Delta m_\beta(\mathbf{x}')\,d\mathbf{x}' \\ &\quad -\int_{\Omega(\mathbf{x}')} H_{\beta\rho'}(\mathbf{x},\mathbf{x}')\,\Delta m_{\rho'}(\mathbf{x}')\,d\mathbf{x}',\end{aligned} \tag{4.29}$$

$$\begin{aligned}\nabla_{\rho'} \Phi(\mathbf{x}) &= a_{\rho'}\left(K_{\alpha\to\rho'}(\mathbf{x}) + K_{\beta\to\rho'}(\mathbf{x}) + K_{\rho'\leftrightarrow\rho'}(\mathbf{x})\right) \\ &= -\int_{\Omega(\mathbf{x}')} H_{\rho'\alpha}(\mathbf{x},\mathbf{x}')\,\Delta m_\alpha(\mathbf{x}')\,d\mathbf{x}' \\ &\quad -\int_{\Omega(\mathbf{x}')} H_{\rho'\beta}(\mathbf{x},\mathbf{x}')\,\Delta m_\beta(\mathbf{x}')\,d\mathbf{x}' \\ &\quad -\int_{\Omega(\mathbf{x}')} H_{\rho'\rho'}(\mathbf{x},\mathbf{x}')\,\Delta m_{\rho'}(\mathbf{x}')\,d\mathbf{x}',\end{aligned} \tag{4.30}$$

where $K_{\beta\leftrightarrow\beta}$ and $K_{\rho'\leftrightarrow\rho'}$ are correct update kernels for β and ρ', $K_{\alpha\to\beta}$ and $K_{\rho'\to\beta}$ described by off-diagonal blocks $\mathbf{H}_{\beta\alpha}$ and $\mathbf{H}_{\beta\rho'}$ indicate contaminations from α and ρ' to β, $K_{\alpha\to\rho'}$ and $K_{\beta\to\rho'}$ described by off-diagonal blocks $\mathbf{H}_{\rho'\alpha}$ and $\mathbf{H}_{\rho'\beta}$ are contaminations from α and β to ρ'. Explicit expressions of these correct update kernels and *interparameter contamination kernels* are given in Appendix C.

According to equations (4.28), (4.29) and (4.30), gradient updates are linear combinations of correct model estimations and the contributions of the *interparameter contamination kernels*, which are determined by both of model perturbations and multiparameter Hessian off-diagonal blocks. In a generalized inversion framework, off-diagonal blocks of the multiparameter Hessian provide direct measurements of the parameter crosstalks, which are influenced by different wave modes, source-receiver illumination, parameterizations, etc. In large-scale inverse problems, it is always unaffordable to construct the whole Hessian matrix explicitly. One objective of this paper is to infer the characteristics of Hessian with matrix probing techniques by applying multiparameter Hessian to various types of vectors and quantify the interparameter tradeoffs in isotropic-elastic FWI. Products of multiparameter Hessian with an arbitrary vector can be calculated with adjoint-sate and finite-difference approaches, as explained in Appendix D.

4.3.4.2 Multiparameter point spread functions

The multiparameter Hessian is first applied to model perturbation vector $\Delta\mathbf{m}$:

$$\Delta\mathbf{m} = \left[\Delta\mathbf{m}_\alpha = 0 \ \Delta\mathbf{m}_\beta = A_\beta \delta\left(\mathbf{x}-\mathbf{z}\right) \ \Delta\mathbf{m}_{\rho'} = 0\right]^\dagger, \tag{4.31}$$

where perturbations of P-wave velocity α and density ρ' are zeros and perturbation of S-wave velocity β is point located at position $\mathbf{z}$ with a strength of A_β. According to equations (4.28), (4.29) and (4.30), the correct update kernel for S-wave velocity $K_{\beta\leftrightarrow\beta}$ is given by:

$$K_{\beta\leftrightarrow\beta}\left(\mathbf{x},\mathbf{z}\right) = -a_\beta^{-1} A_\beta \int_{\Omega(\mathbf{x}')} H_{\beta\beta}\left(\mathbf{x},\mathbf{x}'\right)\delta\left(\mathbf{x}'-\mathbf{z}\right)d\mathbf{x}'. \tag{4.32}$$

According to the sifting property of delta function:

$$K_{\beta\leftrightarrow\beta}\left(\mathbf{x},\mathbf{z}\right) = -a_\beta^{-1} A_\beta \mathbf{H}_{\beta\beta}\left(\mathbf{x},\mathbf{z}\right). \tag{4.33}$$

Similarly, *interparameter contamination kernels* $K_{\beta\to\alpha}$ and $K_{\beta\to\rho'}$ are given by:

$$K_{\beta\to\alpha}\left(\mathbf{x},\mathbf{z}\right) = -a_\beta^{-1} A_\beta \int_{\Omega(\mathbf{x}')} H_{\alpha\beta}\left(\mathbf{x},\mathbf{x}'\right)\delta\left(\mathbf{x}'-\mathbf{z}\right)d\mathbf{x}' = -a_\beta^{-1} A_\beta \mathbf{H}_{\alpha\beta}\left(\mathbf{x},\mathbf{z}\right) \tag{4.34}$$

$$K_{\beta\to\rho'}\left(\mathbf{x},\mathbf{z}\right) = -a_\beta^{-1} A_\beta \int_{\Omega(\mathbf{x}')} H_{\rho'\beta}\left(\mathbf{x},\mathbf{x}'\right)\delta\left(\mathbf{x}'-\mathbf{z}\right)d\mathbf{x}' = -a_\beta^{-1} A_\beta \mathbf{H}_{\alpha\beta}\left(\mathbf{x},\mathbf{z}\right), \tag{4.35}$$

where $K_{\beta\to\alpha}(\mathbf{x},\mathbf{z})$ and $K_{\beta\to\rho'}(\mathbf{x},\mathbf{z})$ are local contaminations from β to α and ρ'. Multiparameter Hessian-vector product preserves the column of multiparameter Hessian $\mathbf{H}_\beta(\mathbf{x},\mathbf{z}) = [\mathbf{H}_{\alpha\beta}(\mathbf{x},\mathbf{z}) \quad \mathbf{H}_{\beta\beta}(\mathbf{x},\mathbf{z}) \quad \mathbf{H}_{\rho'\beta}(\mathbf{x},\mathbf{z})]^\dagger$, which is referred to as a multiparameter point spread function (MPSF) following the common convention in exploration geophysics (Hu et al., 2001; Valenciano et al., 2006; Valenciano, 2008; Tang, 2009; Ren et al., 2011). Following equations (4.25), (4.26) and (4.27), the multiparameter point spread function $\mathbf{H}_\beta(\mathbf{x},\mathbf{z})$ can be expressed explicitly as:

$$\begin{aligned}\mathbf{H}_{\beta\beta}(\mathbf{x},\mathbf{z}) =&\langle -2\rho'\beta^2(\mathbf{x})\left[\partial_j G_{ni}(\mathbf{x})\left(\partial_i u_j(\mathbf{x}) + \partial_j u_i(\mathbf{x})\right)\right.\\ &\left.-2\partial_i G_{ni}(\mathbf{x})\,\partial_k u_k(\mathbf{x})\right]\mathfrak{J}_\beta(\mathbf{z})\rangle,\end{aligned} \tag{4.36}$$

$$\mathbf{H}_{\alpha\beta}(\mathbf{x},\mathbf{z}) = \langle -2\rho'\alpha^2(\mathbf{x})\,\partial_i G_{ni}(\mathbf{x})\,\partial_k u_k(\mathbf{x})\,\mathfrak{J}_\beta(\mathbf{z})\rangle, \tag{4.37}$$

$$\begin{aligned}\mathbf{H}_{\rho'\beta}(\mathbf{x},\mathbf{z}) = \langle -\rho'(\mathbf{x})\left[\left(G_{ni}(\mathbf{x})\,\partial_t^2 u_i(\mathbf{x}) + 2\alpha^2\partial_i G_{ni}(\mathbf{x})\,\partial_k u_k(\mathbf{x})\right)\right.\\ \left.+2\beta^2\left(\partial_j G_{ni}(\mathbf{x})\left(\partial_i u_j(\mathbf{x}) + \partial_j u_i(\mathbf{x})\right) - 2\partial_i G_{ni}(\mathbf{x})\,\partial_k u_k(\mathbf{x})\right)\right]\mathfrak{J}_\beta(\mathbf{z})\rangle,\end{aligned} \tag{4.38}$$

where $\mathfrak{J}_\beta(\mathbf{z})$ represents the product of Jacobian matrix due to parameter β with the point-localized model perturbation vector:

$$\begin{aligned}\mathfrak{J}_\beta(\mathbf{z}) = -\int_{\Omega(\mathbf{x}')} &2\rho'\beta^2(\mathbf{x}')\left[\partial_{j'} G_{n'i'}(\mathbf{x}')\left(\partial_{i'} u_{j'}(\mathbf{x}') + \partial_{j'} u_{i'}(\mathbf{x}')\right)\right.\\ &\left.-2\partial_{i'} G_{n'i'}(\mathbf{x}')\,\partial_{k'} u_{k'}(\mathbf{x}')\right] A_\beta\delta(\mathbf{x}'-\mathbf{z})\,d\mathbf{x}'.\end{aligned} \tag{4.39}$$

Applying multiparameter Hessian to spike model perturbation $\Delta\mathbf{m}_\alpha = A_\alpha\delta(\mathbf{x}-\mathbf{z})$ or $\Delta\mathbf{m}_{\rho'} = A_{\rho'}\delta(\mathbf{x}-\mathbf{z})$ allows us to calculate the MPSFs $\mathbf{H}_{\beta\alpha}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\rho'\alpha}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\alpha\rho'}(\mathbf{x},\mathbf{z})$, and $\mathbf{H}_{\beta\rho'}(\mathbf{x},\mathbf{z})$, which describe the local contaminations from α to β and ρ' and the contaminations from ρ' to α and β. With these MPSFs, the relative strengths, phase characteristics and spreading widths of the local interparameter contaminations are evaluated by taking finite-frequency effects and source-receiver illumination into consideration. Because it is used within the context of the Born approximation, the amplitude of the spike model perturbation vector should be chosen to be smaller than 10% of the background model.

4.3.4.3 Evaluating interparameter tradeoffs within the whole volume

Multiparameter point spread functions (MPSFs) are limited in their ability to characterize the parameter resolution because they are spatially local. To evaluate the coupling effects

of different physical parameters in the whole volume of interest, MPSFs would have to be computed for each type of model parameter at every spatial position, which gives rise to extensive computation requirements. An efficient stochastic probing approach is introduced to estimate the essential diagonals of subblock matrices in multiparameter Hessian. Diagonals of multiparameter Hessian off-diagonal blocks measure the coupling strengths of different physical parameters in the whole volume.

I first consider a function $v(\mathbf{x})$, which satisfies $\mathbf{v} \sim \mathcal{N}(\mathbb{E}[\mathbf{v}], \mathbf{\Sigma_{vv}})$ ($\mathcal{N}$ means Gaussian distribution). Expectation value $\mathbb{E}[\mathbf{v}]$ and variance-covariance matrix $\mathbf{\Sigma_{vv}}$ satisfy:

$$\mathbb{E}[v(\mathbf{x})] = 0, \tag{4.40}$$

$$\begin{aligned} \Sigma_{\mathbf{vv}}(v(\mathbf{x}), v(\mathbf{x}')) &= \mathbb{E}\left[(v(\mathbf{x}) - \mathbb{E}[v(\mathbf{x})])(v(\mathbf{x}') - \mathbb{E}[v(\mathbf{x}')])^{\dagger}\right] \\ &= \mathbb{E}[v(\mathbf{x})\, v(\mathbf{x}')] - \mathbb{E}[v(\mathbf{x})](\mathbb{E}[v(\mathbf{x}')])^{\dagger} \\ &= \delta(\mathbf{x} - \mathbf{x}'). \end{aligned} \tag{4.41}$$

Correlating this random function with its Hessian-vector product $\mathfrak{H} = \mathbf{Hv}$ gives:

$$\begin{aligned} v(\mathbf{x}) \odot \mathfrak{H}(\mathbf{x}) &= \int_{\Omega(\mathbf{x}')} v(\mathbf{x})\, H(\mathbf{x}, \mathbf{x}')\, v(\mathbf{x}')\, d\mathbf{x}' \\ &= v(\mathbf{x})\, H(\mathbf{x}, \mathbf{x})\, v(\mathbf{x}) + \int_{\Omega(\mathbf{x}')} v(\mathbf{x})\, H_{\mathbf{x}\neq\mathbf{x}'}(\mathbf{x}, \mathbf{x}')\, v(\mathbf{x}')\, d\mathbf{x}', \end{aligned} \tag{4.42}$$

where $\odot$ indicates element-wise multiplication, $H(\mathbf{x}, \mathbf{x})$ and $H_{\mathbf{x}\neq\mathbf{x}'}(\mathbf{x}, \mathbf{x}')$ represents Hessian diagonals and off-diagonals. Applying expectation operator $\mathbb{E}$ on both sides of equation (4.42) gives (Sacchi et al., 2007; Trampert et al., 2013):

$$\begin{aligned} \mathbb{E}[v(\mathbf{x}) \odot \mathfrak{H}(\mathbf{x})] &= \int_{\Omega(\mathbf{x}')} H(\mathbf{x}, \mathbf{x}')\, \mathbb{E}[v(\mathbf{x})\, v(\mathbf{x}')]\, d\mathbf{x}' \\ &= \int_{\Omega(\mathbf{x}')} H(\mathbf{x}, \mathbf{x}') \left(\Sigma_{\mathbf{vv}}(v(\mathbf{x}), v(\mathbf{x}')) + \mathbb{E}[v(\mathbf{x})](\mathbb{E}[v(\mathbf{x}')])^{\dagger}\right) d\mathbf{x}' \\ &= \int_{\Omega(\mathbf{x}')} H(\mathbf{x}, \mathbf{x}')\, \delta(\mathbf{x} - \mathbf{x}')\, d\mathbf{x}' \\ &= H(\mathbf{x}, \mathbf{x}), \end{aligned} \tag{4.43}$$

where it can be seen that taking the expectation operation, the second term in equation (4.42), which represents off-diagonal elements, vanishes (Hutchinson, 1990; Trampert et al.,

2013). The theoretical expectation operation can be approximated by averaging the cross-correlation results $\mathbf{v} \odot \mathfrak{H}$ with a finite number of independent zero-mean random vectors:

$$\mathbf{H}^{\text{diag}} \approx \sum_{nr=1}^{NR} \mathbf{v}_{nr} \odot \mathbf{H}\mathbf{v}_{nr} \oslash \sum_{nr}^{NR} \mathbf{v}_{nr} \odot \mathbf{v}_{nr} \tag{4.44}$$

where $\oslash$ represents element-wise division, nr is the index of random vector, NR indicates the maximum number of random vectors and $\mathbf{v}_{nr} \odot \mathbf{v}_{nr}$ is normalization term (MacCarthy et al., 2011). In a multiparameter inverse problem, the random vector $\mathbf{v}$ can be partitioned into N_p subvectors and multiparameter Hessian is divided into $N_p \times N_p$ subblock matrices, as illustrated in equation (4.18). Applying multiparameter Hessian to the random vector gives N_p sub-Hessian-vector products. Diagonals of the Hessian subblock matrices can be estimated by:

$$\mathbf{H}_{pq}^{\text{diag}} = \mathbb{E}\left[\mathbf{v}_p \odot \mathfrak{H}_p\right] = \mathbb{E}\left[\mathbf{v}_p \odot \sum_{q=1}^{N_p} \mathbf{H}_{pq}\mathbf{v}_q\right], \tag{4.45}$$

where p and q are indexes for subvectors representing different physical parameters, and $\mathfrak{H}_p$ is the sub-Hessian-vector product. Considering that zero-mean random vectors $\mathbf{v}_p$ and $\mathbf{v}_q$ for two different physical parameters are independent, equation (4.45) becomes:

$$\mathbf{H}_{pq}^{\text{diag}} = \sum_{q=1}^{N_p} \mathbf{H}_{pq}\mathbb{E}\left[\mathbf{v}_p\mathbf{v}_q\right] = \mathbf{H}_{pq}\mathbb{E}\left[\mathbf{v}_p\mathbf{v}_p\right]. \tag{4.46}$$

Proof of equation (4.46) is given in Appendix E. With a series of random vectors, diagonals of the Hessian subblock matrices can be obtained approximately by:

$$\mathbf{H}_{pq}^{\text{diag}} \approx \sum_{nr=1}^{NR} \mathbf{v}_{p,nr} \odot \mathbf{H}_{pq}\mathbf{v}_{p,nr} \oslash \sum_{nr=1}^{NR} \mathbf{v}_{p,nr} \odot \mathbf{v}_{p,nr}. \tag{4.47}$$

In isotropic-elastic FWI, the random vector is given by $\mathbf{v} = [\mathbf{v}_\alpha\ \mathbf{v}_\beta\ \mathbf{v}_{\rho'}]^\dagger$, where $\mathbf{v}_\alpha$, $\mathbf{v}_\beta$, and $\mathbf{v}_{\rho'}$ are independent zero-mean subvectors. Applying multiparameter Hessian to this random vector gives three sub-Hessian-vector products $\mathfrak{H}_\alpha$, $\mathfrak{H}_\beta$ and $\mathfrak{H}_{\rho'}$:

$$\mathfrak{H}_\alpha = \mathbf{H}_{\alpha\alpha}\mathbf{v}_\alpha + \mathbf{H}_{\alpha\beta}\mathbf{v}_\beta + \mathbf{H}_{\alpha\rho'}\mathbf{v}_{\rho'}, \tag{4.48}$$

$$\mathfrak{H}_\beta = \mathbf{H}_{\beta\alpha}\mathbf{v}_\alpha + \mathbf{H}_{\beta\beta}\mathbf{v}_\beta + \mathbf{H}_{\beta\rho'}\mathbf{v}_{\rho'}, \tag{4.49}$$

$$\mathfrak{H}_{\rho'} = \mathbf{H}_{\rho'\alpha}\mathbf{v}_\alpha + \mathbf{H}_{\rho'\beta}\mathbf{v}_\beta + \mathbf{H}_{\rho'\rho'}\mathbf{v}_{\rho'}. \tag{4.50}$$

With a series of independent zero-mean random vectors, diagonals of the Hessian subblocks $\mathbf{H}_{\alpha\alpha}$ and $\mathbf{H}_{\beta\alpha}$ can be estimated approximately by:

$$\begin{aligned}\mathbf{H}_{\alpha\alpha}^{\text{diag}} &= \mathbb{E}\left[\mathbf{v}_\alpha \odot \mathfrak{H}_\alpha\right] \\ &= \mathbb{E}\left[\mathbf{v}_\alpha \odot \mathbf{H}_{\alpha\alpha}\mathbf{v}_\alpha\right] + \mathbb{E}\left[\mathbf{v}_\alpha \odot \mathbf{H}_{\alpha\beta}\mathbf{v}_\beta\right] + \mathbb{E}\left[\mathbf{v}_\alpha \odot \mathbf{H}_{\alpha\rho'}\mathbf{v}_{\rho'}\right] \\ &\approx \sum_{nr=1}^{NR} \mathbf{v}_{\alpha,nr} \odot \mathbf{H}_{\alpha\alpha}\mathbf{v}_{\alpha,nr} \oslash \sum_{nr=1}^{NR} \mathbf{v}_{\alpha,nr} \odot \mathbf{v}_{\alpha,nr},\end{aligned} \tag{4.51}$$

$$\begin{aligned}\mathbf{H}_{\beta\alpha}^{\text{diag}} &= \mathbb{E}\left[\mathbf{v}_\alpha \odot \mathfrak{H}_\beta\right] \\ &= \mathbb{E}\left[\mathbf{v}_\alpha \odot \mathbf{H}_{\beta\alpha}\mathbf{v}_\alpha\right] + \mathbb{E}\left[\mathbf{v}_\alpha \odot \mathbf{H}_{\beta\beta}\mathbf{v}_\beta\right] + \mathbb{E}\left[\mathbf{v}_\alpha \odot \mathbf{H}_{\beta\rho'}\mathbf{v}_{\rho'}\right] \\ &\approx \sum_{nr=1}^{NR} \mathbf{v}_{\alpha,nr} \odot \mathbf{H}_{\beta\alpha}\mathbf{v}_{\alpha,nr} \oslash \sum_{nr=1}^{NR} \mathbf{v}_{\alpha,nr} \odot \mathbf{v}_{\alpha,nr}.\end{aligned} \tag{4.52}$$

A similar approach can be used to estimate the diagonals of $\mathbf{H}_{\beta\beta}$, $\mathbf{H}_{\rho'\rho'}$, $\mathbf{H}_{\alpha\rho'}$ and $\mathbf{H}_{\beta\rho'}$. The choice of maximum random vectors NR depends on the desired accuracy of the estimated diagonals, which can be evaluated by statistically examining repeated estimates with independent random vectors (MacCarthy et al., 2011). Generally, more random probes give better estimations. If the sublocks of multiparameter Hessian are diagonally dominant, much less random probes are needed (Trampert et al., 2013). Sacchi et al. (2007) estimated the diagonal Hessian preconditioner with 5 random realizations using a phase shift approach. In this chapter, we show that diagonals of multiparameter Hessian can be estimated stochastically with 1 or 2 random Hessian-vector applications using spectral-element method. Stochastic estimations of Hessian diagonals can also be used as effective preconditioners in the inversion process.

4.3.5 Reducing interparameter tradeoffs with approximate contamination kernels

Non-uniqueness due to interparameter tradeoffs will increase nonlinearity and uncertainties within multiparameter inverse problems significantly. Different strategies including Newton-based optimization methods (Métivier et al., 2015; Liu et al., 2015), subspace optimization methods (Kennett et al., 1988; Baumstein, 2014; Bernauer et al., 2014), and wave mode decomposition strategies (Wang and Cheng, 2017), have been proposed to reduce the influences of interparameter tradeoffs in multiparameter FWI. However, most of these strategies have

some limitations, as discussed in introduction section. In the numerical modelling section, with the proposed probing strategies, I find that S-wave velocity dominates the inversion process and produces relatively strong contaminations into density and P-wave velocity updates but suffers very weak contaminations from other parameters. Based on these observations, a novel inversion strategy has been developed to reduce the contaminations from S-wave velocity to other parameters especially density by approximating the contamination kernels.

From equation (4.30), it can be seen that the standard sensitivity kernel $K_{\rho'}$ is just linear summation of the correct update kernel $K_{\rho'\leftrightarrow\rho'}$ with two contamination kernels $K_{\alpha\rightarrow\rho'}$ and $K_{\beta\rightarrow\rho'}$. The interparameter mappings from α and β to ρ' can be removed completely by simply summing the Hessian-vector products $\mathfrak{H}_{\rho'\alpha}=\mathbf{H}_{\rho'\alpha}\Delta\mathbf{m}_{\alpha}$ and $\mathfrak{H}_{\rho'\beta}=\mathbf{H}_{\rho'\beta}\Delta\mathbf{m}_{\beta}$ with standard sensitivity kernel $K_{\rho'}$. However, true model perturbation vectors $\Delta\mathbf{m}_{\alpha}$ and $\Delta\mathbf{m}_{\beta}$ are unknown variables. Because S-wave velocity suffers little contaminations from other parameters, the model parameters can be updated simultaneously for a finite number of k' iterations and then the inverted P-wave velocity and density models are dropped. The estimated S-wave velocity $\mathbf{m}_{\beta}^{k'}$ is kept. The inversion is then started from initial models by simultaneously updating three model parameters. At the $\tilde{k}$th iteration, the approximate contamination kernels $\tilde{K}_{\beta\rightarrow\alpha}^{\tilde{k}}$ and $\tilde{K}_{\beta\rightarrow\rho'}^{\tilde{k}}$ are constructed:

$$\tilde{K}_{\beta\rightarrow\alpha}^{\tilde{k}}\left(\mathbf{x}\right)=-\int_{\Omega(\mathbf{x}')}H_{\alpha\beta}^{\tilde{k}}\left(\mathbf{x},\mathbf{x}'\right)\Delta\tilde{m}_{\beta}^{\tilde{k}}\left(\mathbf{x}'\right)d\mathbf{x}', \tag{4.53}$$

$$\tilde{K}_{\beta\rightarrow\rho'}^{\tilde{k}}\left(\mathbf{x}\right)=-\int_{\Omega(\mathbf{x}')}H_{\rho'\beta}^{\tilde{k}}\left(\mathbf{x},\mathbf{x}'\right)\Delta\tilde{m}_{\beta}^{\tilde{k}}\left(\mathbf{x}'\right)d\mathbf{x}', \tag{4.54}$$

where $\Delta\tilde{\mathbf{m}}_{\beta}^{\tilde{k}}=\mathbf{m}_{\beta}^{k'}-\mathbf{m}_{\beta}^{\tilde{k}}$ is the approximate model perturbation vector. Subtracting the approximate contamination kernels from the standard sensitivity kernels $K_{\alpha}^{\tilde{k}}$ and $K_{\rho'}^{\tilde{k}}$ will remove the contaminations partially and give the new update kernels for α, β and ρ':

$$\tilde{K}_{\alpha}^{\tilde{k}}\left(\mathbf{x}\right)=K_{\alpha}^{\tilde{k}}\left(\mathbf{x}\right)-\tilde{K}_{\beta\rightarrow\alpha}^{\tilde{k}}\left(\mathbf{x}\right),\tilde{K}_{\beta}^{\tilde{k}}\left(\mathbf{x}\right)=K_{\beta}^{\tilde{k}}\left(\mathbf{x}\right),\tilde{K}_{\rho'}^{\tilde{k}}\left(\mathbf{x}\right)=K_{\rho'}^{\tilde{k}}\left(\mathbf{x}\right)-\tilde{K}_{\beta\rightarrow\rho'}^{\tilde{k}}\left(\mathbf{x}\right), \tag{4.55}$$

in which the S-wave velocity kernel $\tilde{K}_{\beta}^{\tilde{k}}$ is kept unchanged. A better approximation of the model perturbation vector $\Delta\tilde{\mathbf{m}}_{\beta}$ removes the contaminations more completely but at the cost of more computation requirements. Table 4.1 illustrates the basic work-flow for this inversion strategy. In traditional simultaneous inversion strategy, the computational cost of

Notations: $\tilde{k}_{max}$ is the maximum iteration; ϕ is the normalized misfit; ϕ_{min} is the minimum normalized misfit; $\mathbf{N}_f$ is the frequency band.

Input: $\leftarrow \mathbf{m}^0 = [\mathbf{m}^0_\alpha\ \mathbf{m}^0_\beta\ \mathbf{m}^0_{\rho'}]^\dagger$, ϕ_{min}, $\tilde{k}_{max}$, $\tilde{\tilde{k}}_{max}$, k', $\mathbf{N}_f$, $\mathbf{d}_{\text{obs}}$

Output: $\rightarrow \mathbf{m}^{\text{est}}$, ϕ

Initialization: $\tilde{k} = 0$

For $\tilde{k} < \tilde{k}_{max}$ or $\phi_{\tilde{k}} > \phi_{min}$

1. $\leftarrow \mathbf{m}^{\tilde{k}}$, k', $\mathbf{N}_f$, $\mathbf{d}_{\text{obs}}$ \\ Iteratively estimate β by k' iterations $\rightarrow \mathbf{m}^{k'}_\beta$;
2. **For** $\tilde{\tilde{k}} < \tilde{\tilde{k}}_{max}$

2.1. $\leftarrow \mathbf{m}^{\tilde{k}}$, $\mathbf{N}_f$, $\mathbf{d}_{\text{obs}}$ \\ Calculate sensitivity kernels $\rightarrow K^{\tilde{\tilde{k}}}_\alpha$, $K^{\tilde{\tilde{k}}}_\beta$ and $K^{\tilde{\tilde{k}}}_{\rho'}$;

2.2. $\leftarrow \Delta\tilde{\mathbf{m}}^{\tilde{\tilde{k}}}_\beta = \mathbf{m}^{k'}_\beta - \mathbf{m}^{\tilde{\tilde{k}}}_\beta$ \\ Calculate approximate contamination kernels:
$\rightarrow \tilde{K}^{\tilde{\tilde{k}}}_{\beta\rightarrow\alpha} = -\mathbf{H}^{\tilde{\tilde{k}}}_{\alpha\beta}\Delta\tilde{\mathbf{m}}^{\tilde{\tilde{k}}}_\beta$, $\tilde{K}^{\tilde{\tilde{k}}}_{\beta\rightarrow\rho'} = -\mathbf{H}^{\tilde{\tilde{k}}}_{\rho'\beta}\Delta\tilde{\mathbf{m}}^{\tilde{\tilde{k}}}_\beta$

2.3. $\leftarrow \tilde{K}^{\tilde{\tilde{k}}}_{\beta\rightarrow\alpha}$ and $\tilde{K}^{\tilde{\tilde{k}}}_{\beta\rightarrow\rho'}$ \\ Calculate new update kernels:
$\rightarrow \tilde{K}^{\tilde{\tilde{k}}}_\alpha = K^{\tilde{\tilde{k}}}_\alpha - \tilde{K}^{\tilde{\tilde{k}}}_{\beta\rightarrow\alpha}$, $\tilde{K}^{\tilde{\tilde{k}}}_\beta = K^{\tilde{\tilde{k}}}_\beta$, $\tilde{K}^{\tilde{\tilde{k}}}_{\rho'} = K^{\tilde{\tilde{k}}}_{\rho'} - \tilde{K}^{\tilde{\tilde{k}}}_{\beta\rightarrow\rho'}$

2.4. Apply stochastic estimations of diagonal Hessian preconditioners:

2.5. Get step length $\mu_{\tilde{\tilde{k}}}$ with a line search method;

2.6. Update the model vector: $\mathbf{m}^{\tilde{\tilde{k}}+1} = \mathbf{m}^{\tilde{\tilde{k}}} + \mu_{\tilde{\tilde{k}}}\Delta\mathbf{m}^{\tilde{\tilde{k}}}$;

2.7. Calculate misfit $\phi_{\tilde{\tilde{k}}}$ and $\tilde{\tilde{k}} = \tilde{\tilde{k}} + 1$;

End

3. Update parameters: $\tilde{k} = \tilde{k} + \tilde{\tilde{k}}_{max}$, $\phi_{\tilde{k}} = \phi_{\tilde{\tilde{k}}}$, $\mathbf{m}^{\text{est}} = \mathbf{m}^{\tilde{k}} = \mathbf{m}^{\tilde{\tilde{k}}_{max}}$;

End

Table 4.1: Work-flow of the new inversion strategy for isotropic-elastic FWI with approximate contamination kernels.

$\tilde{k}_{max}$ iterations is equivalent to number of $2 \times N_s \times \tilde{k}_{max}$ forward and adjoint simulations. This new inversion strategy will be more expensive for obtaining $\mathbf{m}^{k'}_\beta$ and constructing approximate contamination kernels. For $\tilde{k}_{max}$ iterations, the number of forward and adjoint simulations is equivalent to $\left(2 \times N_s \times k' \times N_{k'} + 8 \times N_s \times \tilde{k}_{max}\right)$, where $N_{k'}$ is number of loops for obtaining $\mathbf{m}^{k'}_\beta$.

4.3.6 Resolution analysis

Resolution analysis is long lasting issue for geophysical inverse problems and have been studied by many researchers (Backus and Gilbert, 1968; Spakman, 1991; Rawlinson et al., 2014;

Rawlinson and Spakman, 2016). Quantifying resolution and uncertainties of the inverted models due to interparameter tradeoffs is a key aspect of for multiparameter FWI. Assuming that an optimal model $\mathbf{m}$ has been obtained with least-squares optimization framework, applying model perturbation $\Delta\mathbf{m}$ gives perturbed model $\mathbf{m}' = \mathbf{m} + \Delta\mathbf{m}$, which is close to model $\mathbf{m}$. The reconstructed model $\tilde{\mathbf{m}}$ can be obtained by $\tilde{\mathbf{m}} = \mathbf{m} + \Delta\mathbf{m} + \Delta\tilde{\mathbf{m}}$, where $\Delta\tilde{\mathbf{m}}$ represents the estimated model perturbation vector:

$$\Delta\tilde{\mathbf{m}} = -\mathbf{H}^{-g}\nabla_{\mathbf{m}}\Phi = \mathbf{H}^{-g}\mathbf{H}\Delta\mathbf{m} = \mathbf{R}\Delta\mathbf{m}, \tag{4.56}$$

where $\mathbf{H}^{-g}$ is the generalized inverse of $\mathbf{H}$ and $\mathbf{R} = \mathbf{H}^{-g}\mathbf{H}$ is the resolution matrix, which describes how the estimated model perturbation $\Delta\tilde{\mathbf{m}}$ relates to the true model perturbation $\Delta\mathbf{m}$. Ideally $\mathbf{R}$ should be an identity matrix $\mathbf{I}$ meaning that the model is perfectly recovered (Backus and Gilbert, 1968). However, if the resolution matrix deviates significantly from the identity matrix, the inverted model suffers from tradeoffs (Luo, 2012). A column of $\mathbf{R}$ measures the local resolution and uncertainties of the inverted model (Oldenborger and Routh, 2009; Fichtner and Trampert, 2011c). However, explicitly constructing and inverting $\mathbf{H}$ are computationally unaffordable for large-scale inverse problems. In recent years, researchers evaluated the local resolution of the inverted model with point spread functions by approximating $\mathbf{H}^{-g}$ with an identity matrix $\mathbf{I}$ (Fichtner and Trampert, 2011c; Rickers et al., 2013; Zhu et al., 2015; Bozdağ et al., 2016). Thus, the column of multiparameter Hessian (i.e., $\mathbf{H}(\mathbf{x}, \mathbf{z})$ equation (4.33)) only represents an conservative estimation of the column in resolution matrix (i.e., $\mathbf{R}(\mathbf{x}, \mathbf{z})$) (Fichtner and van Leeuwen, 2015). In this section, the similarities and differences between $\mathbf{H}(\mathbf{x}, \mathbf{z})$ and $\mathbf{R}(\mathbf{x}, \mathbf{z})$ in resolution analysis are investigated and the potential benefits by applying approximate inverse Hessian operators to PSFs are explored.

The symmetric and positive semi-definite Hessian matrix $\mathbf{H}$ can be decomposed as:

$$\mathbf{H} = \Xi\Pi\Xi^{-1}, \tag{4.57}$$

where $\Xi = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, ..., \mathbf{a}_M]$ is an orthogonal matrix consisting of M column eigenvectors $\mathbf{a}_{\dot{k}}, \dot{k} = 1, 2, ..., M$ of $\mathbf{H}$ and Π is a diagonal matrix with corresponding eigenvalues $\lambda_{\dot{k}}$. The generalized inverse of $\mathbf{H}$ is given by:

$$\mathbf{H}^{-g} = \left[\Xi\left(\Pi + \lambda_0\mathbf{I}\right)\Xi^{-1}\right]^{-1} = \Xi\left(\Pi + \lambda_0\mathbf{I}\right)^{-1}\Xi^{-1}, \tag{4.58}$$

where $\lambda_0\mathbf{I}$ is the damping term added to the eigenvalues. The resolution matrix $\mathbf{R}$ can be obtained by:

$$\mathbf{R} = \Xi\tilde{\Pi}^{-1}\Pi\Xi^{-1} = \Xi\left(\Pi + \lambda_0\mathbf{I}\right)^{-1}\Pi\Xi^{-1}, \tag{4.59}$$

where $\tilde{\Pi} = \left(\Pi + \lambda_0\mathbf{I}\right)^{-1}\Pi$ is diagonal matrix with eigenvalues of $\tilde{\lambda} = \left(1 + \lambda_0/\lambda_{\dot{k}}\right)^{-1}$. The Hessian matrix and resolution matrix have the same eigenvectors but different eigenvalues. Because the orthogonal eigenvectors of $\mathbf{H}$ span the model space, the model perturbation vector $\Delta\mathbf{m}$ can also be written as a sum of M eigenvectors $\mathbf{a}_{\dot{k}}$:

$$\Delta\mathbf{m} = \sum_{\dot{k}=1}^{M} h_{\dot{k}}\mathbf{a}_{\dot{k}} = h_1\mathbf{a}_1 + h_2\mathbf{a}_2 + \mathbf{...} + h_M\mathbf{a}_M, \tag{4.60}$$

where $h_{\dot{k}}$ are the model expansion coefficients. Combining equation (4.60) and equation (4.57), Hessian-vector product $\mathbf{H}\Delta\mathbf{m}$ can be expressed in terms of eigenvalues and eigenvectors of $\mathbf{H}$:

$$\mathbf{H}\Delta\mathbf{m} = \sum_{\dot{k}=1}^{M} \lambda_{\dot{k}}h_{\dot{k}}\mathbf{a}_{\dot{k}} = \lambda_1h_1\mathbf{a}_1 + \lambda_2h_2\mathbf{a}_2 + \mathbf{...} + \lambda_Mh_M\mathbf{a}_M. \tag{4.61}$$

Substituting equations (4.60) and (4.59) into equation (4.56) gives:

$$\begin{aligned}\tilde{\mathbf{m}} &= \sum_{\dot{k}=1}^{M}\left(1 + \lambda_0/\lambda_{\dot{k}}\right)^{-1}h_{\dot{k}}\mathbf{a}_{\dot{k}}, \\ &= \left(1 + \lambda_0/\lambda_1\right)^{-1}h_1\mathbf{a}_1 + \left(1 + \lambda_0/\lambda_2\right)^{-1}h_2\mathbf{a}_2 + \mathbf{...} + \left(1 + \lambda_0/\lambda_M\right)^{-1}h_M\mathbf{a}_M.\end{aligned} \tag{4.62}$$

Assuming that the eigenvalues of $\mathbf{H}$ are constant $\lambda_{\dot{k}} \approx \lambda$, equations (4.61) and (4.62) become:

$$\begin{aligned}\mathbf{H}\Delta\mathbf{m} &\approx \sum_{\dot{k}=1}^{M}\lambda h_{\dot{k}}\mathbf{a}_{\dot{k}} = \lambda\Delta\mathbf{m}, \\ \tilde{\mathbf{m}} &\approx \sum_{\dot{k}=1}^{M}\left(1 + \lambda_0/\lambda\right)^{-1}h_{\dot{k}}\mathbf{a}_{\dot{k}} = \left(1 + \lambda_0/\lambda\right)^{-1}\Delta\mathbf{m}.\end{aligned} \tag{4.63}$$

Magnitudes of the PSFs directly measure the magnitudes of eigenvalues. Larger eigenvalues mean well constrained eigenvectors. Smaller eigenvalues mean poorly constrained eigenvectors. PSFs determine resolution of the inverted models with eigenvalues and mimic the shape of the true model perturbation but have distinct magnitudes. Because $\lambda_{\dot{k}} \gg \lambda_0$, then $\left(1 + \lambda_0/\lambda_{\dot{k}}\right)^{-1} \approx 1$, the closer of eigenvalues of the resolution matrix approach 1, the better of the resolution is. If the Hessian is diagonally dominant, eigenvalues of the resolution matrix

within the whole volume can be approximated by diagonals of the resolution matrix (Luo, 2012; Zhu et al., 2015), referred to as approximate eigenvalue volume:

$$\mathbf{Eig} = \left[\mathbf{H}^{\mathrm{diag}} + \tilde{\epsilon} \times \max\left(\mathbf{H}^{\mathrm{diag}}\right)\right]^{-1} \mathbf{H}^{\mathrm{diag}}, \tag{4.64}$$

where $\max\left(\mathbf{H}^{\mathrm{diag}}\right)$ represents the maximum value of diagonal Hessian $\mathbf{H}^{\mathrm{diag}}$ and $\tilde{\epsilon}$ is a small constant value.

Here, this chapter proposes to quantify the local spatial and interparameter tradeoffs of the inverted models with extended multiparameter point spread functions (EMPSFs) by applying approximate inverse Hessian to MPSFs with conjugate-gradient algorithm preconditioned by stochastic estimations of diagonal Hessian. Considering a point-localized model perturbation vector $\Delta\mathbf{m} = [\Delta\mathbf{m}_\alpha = 0\ \Delta\mathbf{m}_\beta = A_\beta\delta\left(\mathbf{x}-\mathbf{z}\right)\ \Delta\mathbf{m}_{\rho'} = 0]^\dagger$, equation (4.56) can be written as:

$$\Delta\tilde{m}_\beta\left(\mathbf{z}\right) = \int_{\Omega(\mathbf{x})}\int_{\Omega(\mathbf{x}')} A_\beta\tilde{R}_\beta\left(\mathbf{x},\mathbf{x}'\right)\delta\left(\mathbf{x}'-\mathbf{z}\right)d\mathbf{x}'d\mathbf{x} = \int_{\Omega(\mathbf{x})} A_\beta\tilde{R}_\beta\left(\mathbf{x},\mathbf{z}\right)d\mathbf{x}, \tag{4.65}$$

where $\tilde{\mathbf{R}}_\beta\left(\mathbf{x},\mathbf{z}\right) = \mathbb{H}^{-1}\left(\mathbf{H}_\beta\left(\mathbf{x},\mathbf{z}\right)\right)$ indicates the extended MPSF (EMPSF) and $\mathbb{H}^{-1}$ represents the approximate inverse Hessian by preconditioned conjugate-gradient algorithm. Applying the inverse Hessian approximately will re-scale the magnitudes and de-blur the MPSFs. Furthermore, approximate inverse multiparameter Hessian will also suppress the interparameter contaminations to a certain extent (Innanen, 2014b; Métivier et al., 2015; Pan et al., 2016; Wang and Cheng, 2017). Thus, the EMPSFs will provide more accurate measurements of the local spatial and interparameter tradeoffs. To evaluate the interparameter tradeoffs of the inverted models obtained by new inversion strategy, the EMPSFs with a variant of preconditioned conjugate-gradient approach are constructed following the work-flow illustrated in Table 4.1.

4.4 Numerical examples

In the numerical modelling section, the proposed strategies are applied to quantify and reduce the interparameter tradeoffs in isotropic-elastic FWI. Spectral-element methods are employed for forward and adjoint simulations with the open-source software package *SPECFEM2D*

(Komatitsch and Tromp, 2005). Influences of surface waves are currently not considered in the numerical examples presented in this chapter.

4.4.1 Spike probing test with MPSFs

The relative strengths and characteristics of the interparameter contaminations are first investigated with multiparameter point spread functions (MPSFs) in isotropic-elastic FWI using x-z component data. Inversion experiments with Gaussian-anomaly examples are given to verify the predictions and examine the effectiveness of this new inversion strategy in reducing the interparameter contaminations.

Figure 4.2 shows the 2D isotropic-elastic model with one spike model perturbation embedded in a homogeneous background. P-wave velocity, S-wave velocity and density of the background model are 2.0 km/s, 1.4 km/s and 1.2 g/cm^3. A P-SV mode source with Ricker wavelet (dominant frequency f_{dom}=8Hz) is used for modeling. A total of 60 sources and 200 receivers are arranged along all boundaries of the model with a regular source spacing of 62.5 m and a regular receiver spacing of 20 m. I first apply a positive spike model perturbation of P-wave velocity at position $\mathbf{z}$ (the model center): $\Delta\mathbf{m}_\alpha(\mathbf{z}) = 0.1$ km/s. Multiparameter point spread functions (MPSFs) $\mathbf{H}_{\alpha\alpha}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\beta\alpha}(\mathbf{x},\mathbf{z})$, and $\mathbf{H}_{\rho'\alpha}(\mathbf{x},\mathbf{z})$ are calculated with x-z component data, where $\mathbf{H}_{\beta\alpha}(\mathbf{x},\mathbf{z})$ and $\mathbf{H}_{\rho'\alpha}(\mathbf{x},\mathbf{z})$ describe the mappings from α to β and ρ'. Then, spike model perturbations $\Delta\mathbf{m}_\beta(\mathbf{z}) = 0.1$ km/s and $\Delta\mathbf{m}_{\rho'}(\mathbf{z}) = 0.1$ g/cm^3 are applied respectively. MPSFs $\mathbf{H}_{\alpha\beta}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\beta\beta}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\rho'\beta}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\alpha\rho'}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\beta\rho'}(\mathbf{x},\mathbf{z})$, and $\mathbf{H}_{\rho'\rho'}(\mathbf{x},\mathbf{z})$ are obtained. $\mathbf{H}_{\alpha\beta}(\mathbf{x},\mathbf{z})$ and $\mathbf{H}_{\rho'\beta}(\mathbf{x},\mathbf{z})$ describe the mappings from β to α and ρ'. $\mathbf{H}_{\alpha\rho'}(\mathbf{x},\mathbf{z})$ and $\mathbf{H}_{\beta\rho'}(\mathbf{x},\mathbf{z})$ describe the mappings from ρ' to α and β.

These MPSFs are plotted in model space and arranged in a block structure in consistent with their positions in multiparameter Hessian, as shown in Figure 4.3a. Figure 4.3a is also equivalent to a sparse representation of multiparameter Hessian with 3 columns, which measure finite-frequency features of the interparameter tradeoffs. A positive α perturbation produces a negative contamination in β described by $\mathbf{H}_{\beta\alpha}(\mathbf{x},\mathbf{z})$ and vice versa. However, both positive α and β perturbations result in positive contaminations in density ρ' described by $\mathbf{H}_{\rho'\alpha}(\mathbf{x},\mathbf{z})$ and $\mathbf{H}_{\rho'\beta}(\mathbf{x},\mathbf{z})$ and vice versa. Furthermore, regards to spatial spreading, the MPSFs representing contaminations to density (i.e., $\mathbf{H}_{\rho'\alpha}(\mathbf{x},\mathbf{z})$ and $\mathbf{H}_{\rho'\beta}(\mathbf{x},\mathbf{z})$) experience

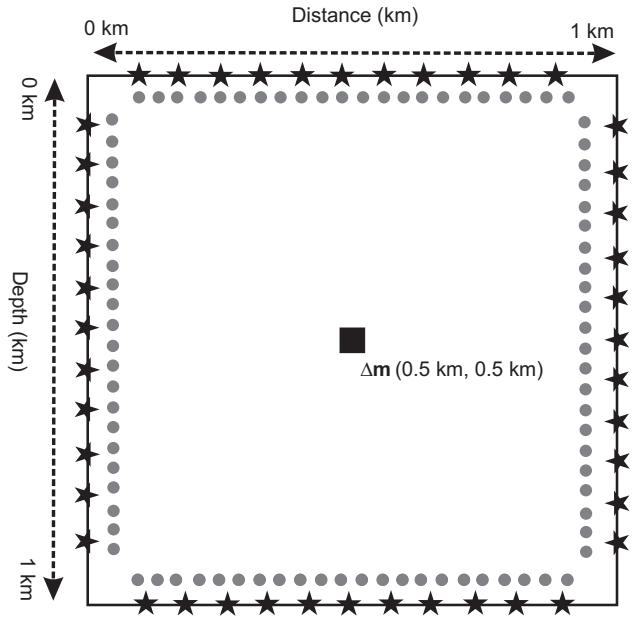

Figure 4.2: Acquisition geometry for spike probing test. Black stars and gray circles represent sources and receivers positions. The 2D model is discretized into 50 and 50 uniform mesh nodes in horizontal and vertical directions with 1 km in width and 1 km in depth. The black square located at the center of the model indicates the spike model perturbation at position $\mathbf{z} = (0.5 \text{ km}, 0.5 \text{ km})$.

oscillatory side-lobes, which may distort the correct density updates.

As indicated by A_{PSF} in Figure 4.3a, magnitudes of the MPSFs, which describe relative strengths of the eigenvalues, also differ significantly. Magnitude of $\mathbf{H}_{\beta\beta}(\mathbf{x}, \mathbf{z})$ is larger than the magnitudes of $\mathbf{H}_{\alpha\alpha}(\mathbf{x}, \mathbf{z})$ and $\mathbf{H}_{\rho'\rho'}(\mathbf{x}, \mathbf{z})$ meaning that the eigenvectors associated with S-wave velocity will be better recovered than those associated with P-wave velocity and density. To evaluate relative strengths of the interparameter contaminations, the contaminations are normalized with the MPSFs representing correct model updates. For example, the MPSFs $\mathbf{H}_{\alpha\beta}(\mathbf{x}, \mathbf{z})$ and $\mathbf{H}_{\alpha\rho'}(\mathbf{x}, \mathbf{z})$ are normalized by the maximum absolute value of $\mathbf{H}_{\alpha\alpha}(\mathbf{x}, \mathbf{z})$. Normalized MPSFs are shown in Figure 4.3b. Contaminations from α to β and ρ' appear to be relatively weak. Density ρ' perturbations also produce moderate unwanted artifacts in α and β. S-wave velocity β suffers from the least amount of contaminations but produces strong mappings to α and ρ', which may make density under- or overestimated and cancel the updates for P-wave velocity. Geological features in the inverted P-wave velocity and density models may be contaminations from the S-wave velocity, which increases the uncertainties of the inverse problems significantly. These information helps us understand

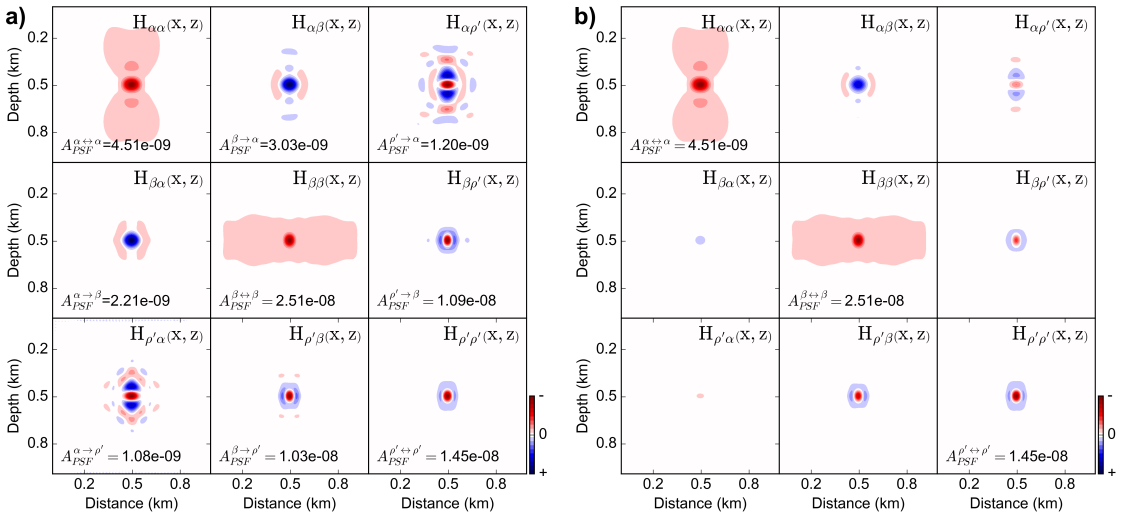

Figure 4.3: Multiparameter point spread functions (MPSFs) of isotropic-elastic parameters in velocity-density parameterization. (a) shows the MPSFs $\mathbf{H}_{\alpha\alpha}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\beta\alpha}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\rho'\alpha}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\alpha\beta}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\beta\beta}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\rho'\beta}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\alpha\rho'}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\beta\rho'}(\mathbf{x},\mathbf{z})$, and $\mathbf{H}_{\rho'\rho'}(\mathbf{x},\mathbf{z})$ with x-z component data; (b) shows the corresponding normalized MPSFs. A_{PSF} indicate the maximum magnitudes of the MPSFs.

how the interparameter tradeoffs affect the inversion process.

To verify our analysis and predictions with MPSFs, inversion experiments with a Gaussian-anomaly model is carried out. Figures 4.4a, 4.4b, and 4.4c show the true P-wave velocity, S-wave velocity and density models with 3 isolated Gaussian anomalies. The initial models are homogeneous with $\alpha = 2.0$ km/s, $\beta = 1.2$ km/s and $\rho' = 1.2$ g/cm^3. The acquisition arrangement is the same with previous example. A l-BFGS optimization method is employed for updating α, β and ρ' simultaneously. This inversion experiment can be considered as an extended version of spike probing test with 3 Gaussian model perturbation vectors. Relative

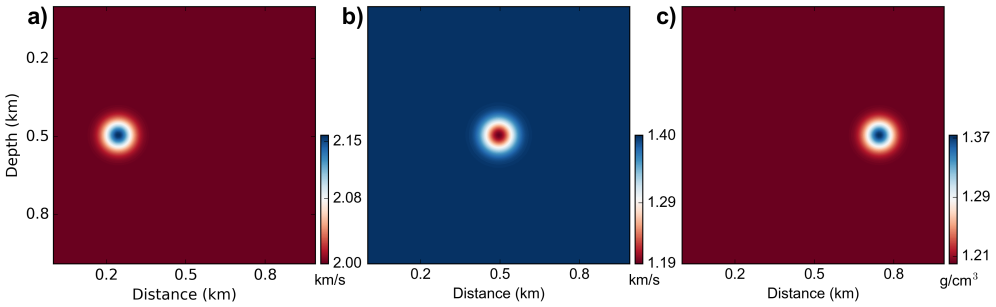

Figure 4.4: Figures (a-c) show the true P-wave velocity, S-wave velocity and density of the Gaussian-anomaly model: $\mathbf{m}_{\alpha}^{\text{true}}$, $\mathbf{m}_{\beta}^{\text{true}}$ and $\mathbf{m}_{\rho'}^{\text{true}}$.

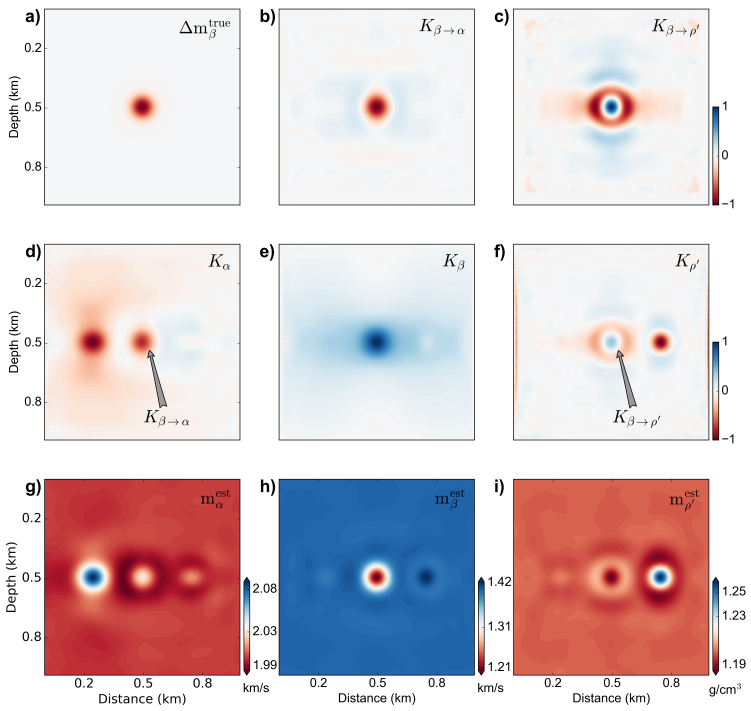

Figure 4.5: Figures (a-c) show the true S-wave velocity model perturbation vector $\Delta\mathbf{m}_\beta^{\text{true}}$, true contamination kernels $K_{\beta\to\alpha}$ and $K_{\beta\to\rho'}$ respectively; Figures (d-f) illustrate the standard sensitivity kernels K_α, K_β, and $K_{\rho'}$; Figures (g-i) are the inverted models $\mathbf{m}_\alpha^{\text{est}}$ ($\tilde{\epsilon}_\alpha$=0.47), $\mathbf{m}_\beta^{\text{est}}$ ($\tilde{\epsilon}_\beta$=0.15) and $\mathbf{m}_{\rho'}^{\text{est}}$ ($\tilde{\epsilon}_{\rho'}$=0.77) after 10 iterations with traditional simultaneous inversion strategy.

least-squares error (RLSE) (equation (2.42)) is used to evaluate the quality of the inverted model. Figure 4.5a shows the true S-wave velocity model perturbation $\Delta\mathbf{m}_\beta^{\text{true}}$. Figures 4.5b and 4.5c show the true contamination kernels $K_{\beta\to\alpha}$ and $K_{\beta\to\rho'}$ calculated by multiplying multiparameter Hessian off-diagonal blocks $\mathbf{H}_{\alpha\beta}$ and $\mathbf{H}_{\rho'\beta}$ with true model perturbation vector $\Delta\mathbf{m}_\beta^{\text{true}}$. Figures 4.5d, 4.5e and 4.5f show the standard sensitivity kernels K_α, K_β and $K_{\rho'}$ in the first iteration. Strengths and characteristics of the interparameter contaminations generally match our predictions using MPSFs shown in Figure 4.3. A negative S-wave velocity perturbation produces strong positive and negative contaminations into the updates for α and ρ', as indicated by the *interparameter contamination kernels* $K_{\beta\to\alpha}$ and $K_{\beta\to\rho'}$. Figures 4.5g, 4.5h and 4.5i show the inverted P-wave velocity, S-wave velocity and density models after 10 iterations using traditional simultaneous inversion strategy. S-wave velocity suffers from limited contaminations and is best inverted. P-wave velocity and density suffer strong contaminations from S-wave velocity. As iteration proceeds, S-wave velocity is estimated

fastest. The interparameter contaminations due to S-wave velocity perturbations are also reduced iteratively and if a sufficient number of iterations are performed, the contaminations are expected to be removed almost completely.

Figure 4.6a shows the estimated S-wave velocity model perturbation vector $\Delta\mathbf{m}_{\beta}^{\text{est}}$ after $k' = 3$ iterations. Figures 4.6b and 4.6c show the approximate contamination kernels $\tilde{K}_{\beta\to\alpha}$ and $\tilde{K}_{\beta\to\rho'}$ calculated by multiplying multiparameter Hessian off-diagonal blocks $\mathbf{H}_{\alpha\beta}$ and $\mathbf{H}_{\rho'\beta}$ with estimated model perturbation vector $\Delta\mathbf{m}_{\beta}^{\text{est}}$. The features of the approximate contamination kernels match those of true contamination kernels (Figures 4.5b and 4.5c) very well. Figures 4.6d, 4.6e and 4.6f are the new update kernels following equation (4.55). Figures 4.6g, 4.6h and 4.6i show the inverted P-wave velocity, S-wave velocity and density models with the new inversion strategy. As indicated by the arrows, contaminations from S-wave velocity to P-wave velocity and density have been suppressed. Figure 4.7 shows the convergence histories of traditional simultaneous inversion strategy and new inversion strategy for the Gauss-anomaly example. The new inversion strategy provides faster convergence compared to traditional simultaneous inversion strategy but it is 2.5 times more expensive.

4.4.2 Marmousi model example

The proposed stochastic probing strategy is first applied to evaluate the strengths of the interparameter tradeoffs within the whole volume. The new inversion strategy with approximate contamination kernels is employed to invert the isotropic-elastic parameters in comparison with traditional simultaneous inversion strategy. Approximate eigenvalue volumes and extended multiparameter point spread functions (EMPSFs) are used to evaluate resolution of the inverted models.

Figures 4.8a, 4.8b and 4.8c show the true P-wave velocity, S-wave velocity and density models. Figures 4.8d, 4.8e and 4.8f show the initial P-wave velocity, S-wave velocity and density models. Figure 4.8g, 4.8h and 4.8i show the corresponding true model perturbations. The model is 3.4 km wide and 1.2 km deep. Number of 33 sources and 330 receivers are deployed regularly with a source spacing of 100 m and a receiver spacing of 10 m along top surface of the model. A Ricker wavelet with dominant frequency of 6 Hz is used for forward modelling.

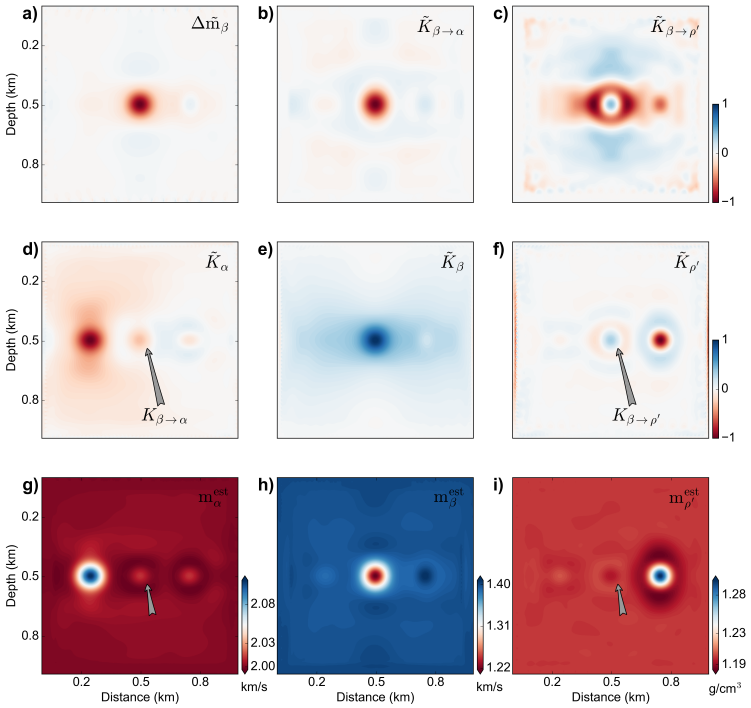

Figure 4.6: Figures (a-c) show the estimated S-wave velocity model perturbation vector $\Delta\tilde{\mathbf{m}}_\beta$ ($k' = 3$), approximate contamination kernels $\tilde{K}_{\beta\to\alpha}$ and $\tilde{K}_{\beta\to\rho'}$ respectively; Figures (d-f) illustrate the new update kernels $\tilde{K}_\alpha$, $\tilde{K}_\beta$, and $\tilde{K}_{\rho'}$; Figures (g-i) are the inverted models $\mathbf{m}_\alpha^{\text{est}}$ ($\tilde{\epsilon}_\alpha$=0.32), $\mathbf{m}_\beta^{\text{est}}$ ($\tilde{\epsilon}_\beta$=0.14) and $\mathbf{m}_{\rho'}^{\text{est}}$ ($\tilde{\epsilon}_{\rho'}$=0.61) after 10 iterations with new inversion strategy.

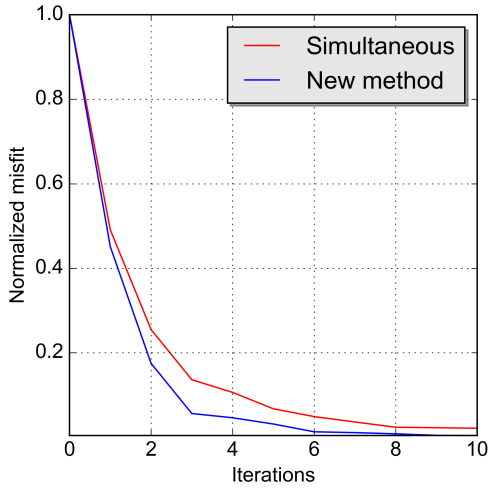

Figure 4.7: Convergence histories comparison of traditional simultaneous inversion strategy (red curve) and new inversion strategy (blue curve) for the Gaussian-anomaly inversion example.

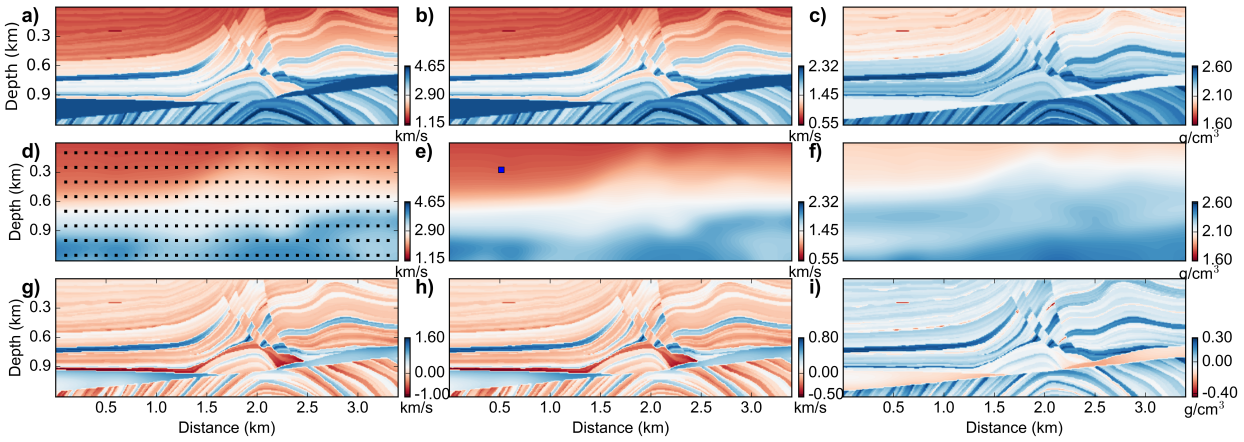

Figure 4.8: (a-c) show true P-wave velocity, S-wave velocity and density models; (d-f) show initial P-wave velocity, S-wave velocity and density models; Figures (g-i) show true P-wave velocity, S-wave velocity and density model perturbations. The regularly distributed black squares in (d) represent the vector $\mathbf{v}'$ for interparameter tradeoffs analysis within the whole model. The blue square in (e) indicates the location $\mathbf{z}_1 = (0.515 \text{ km}, 0.275 \text{ km})$ for quantifying local spatial and interparameter tradeoffs of the inverted models.

4.4.2.1 Evaluating the strengths of interparameter tradeoffs within the whole volume

Diagonals of subblock matrices in multiparameter Hessian are first estimated with the stochastic probing approach following equation (4.47). Figure 4.9a shows the Hessian diagonals $\mathbf{H}^{\text{diag, aj}}_{\rho'\rho'}$ calculated with adjoint-state method (Shin et al., 2001a). The computation cost is equivalent to 363 forward simulations. Figures 4.9b and 4.9c show the stochastic estimations of Hessian diagonals with 1 and 2 random vector applications respectively. Computation costs are equivalent to 66 and 198 forward simulations. Energy distributions in the stochastic estimations generally match those calculated with the adjoint-state method, which verifies the effectiveness of stochastic probing approach. Figures 4.10a show the stochastic estimations of the Hessian diagonals $\mathbf{H}^{\text{diag}}_{\alpha\alpha}$, $\mathbf{H}^{\text{diag}}_{\beta\beta}$ and $\mathbf{H}^{\text{diag}}_{\rho'\rho'}$ after normalization. Energy distributions in these Hessian diagonals for different parameters differ significantly. Stronger elements of the Hessian diagonals mean that the model parameters are well constrained. However, energies of $\mathbf{H}^{\text{diag}}_{\beta\beta}$ are constrained in the shallow parts of the model. Maximum magnitudes of $\mathbf{H}^{\text{diag}}_{\beta\beta}$ are approximately 11.0 times and 6.3 times stronger than those of $\mathbf{H}^{\text{diag}}_{\alpha\alpha}$ and $\mathbf{H}^{\text{diag}}_{\rho'\rho'}$, which means that S-wave velocity will be better recovered than P-wave velocity and density. The Hessian diagonals are also used as preconditioners in the inversion process.

Figures 4.10b show the stochastic estimations of the Hessian diagonals $\mathbf{H}^{\text{diag}}_{\alpha\beta}$, $\mathbf{H}^{\text{diag}}_{\alpha\rho'}$ and

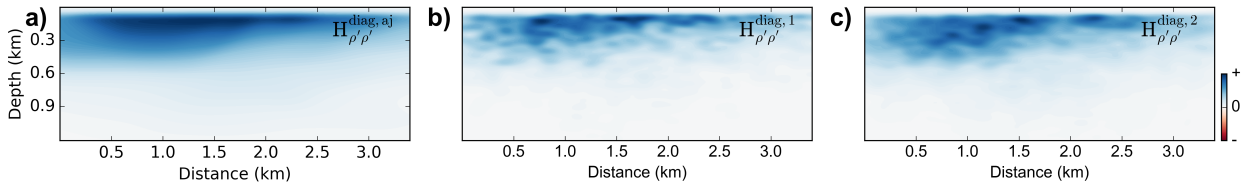

Figure 4.9: (a) shows the Hessian diagonals $\mathbf{H}^{\text{diag, aj}}_{\rho'\rho'}$ calculated with adjoint-state method; (b-c) show the stochastic estimation of Hessian diagonals $\mathbf{H}^{\text{diag, 1}}_{\rho'\rho'}$ and $\mathbf{H}^{\text{diag, 2}}_{\rho'\rho'}$ with 1 and 2 random vector applications respectively.

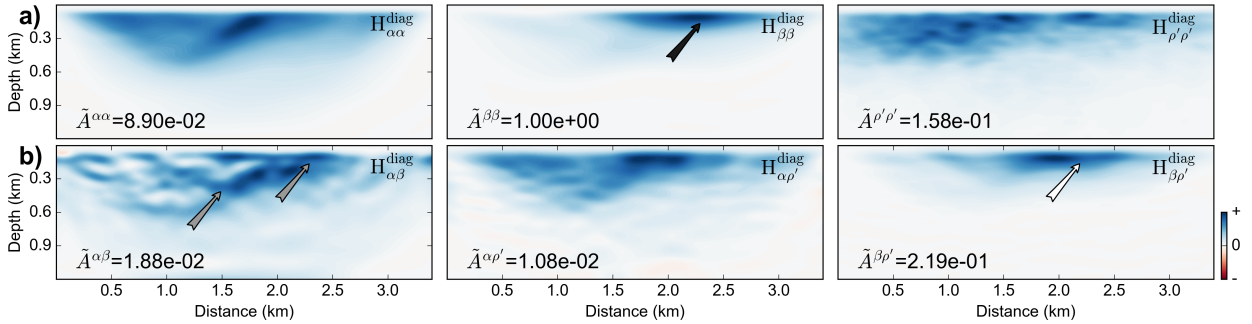

Figure 4.10: (a) shows the the stochastic estimations of Hessian diagonals $\mathbf{H}^{\text{diag}}_{\alpha\alpha}$, $\mathbf{H}^{\text{diag}}_{\beta\beta}$ and $\mathbf{H}^{\text{diag}}_{\rho'\rho'}$ with 2 random vector applications; (b) shows the stochastic estimations of Hessian diagonals $\mathbf{H}^{\text{diag}}_{\alpha\beta}$, $\mathbf{H}^{\text{diag}}_{\alpha\rho'}$ and $\mathbf{H}^{\text{diag}}_{\beta\rho'}$ with 2 random vector applications. $\tilde{A}$ mean the maximum magnitude of the Hessian diagonals after normalization.

$\mathbf{H}^{\text{diag}}_{\beta\rho'}$ which measure the coupling strengths of the isotropic-elastic parameters in the whole volume. The coupling strengths change within the whole volume significantly, that is to say, they are influenced by inhomogeneity of the model and source-receive illumination. In earlier iterations, strong interparameter tradeoffs appear at the shallow parts of the model, as indicated by the grey and white arrows. Magnitudes of the diagonals of off-diagonal blocks associated with different physical parameters are quite different. $\mathbf{H}^{\text{diag}}_{\beta\rho'}$ is much stronger than $\mathbf{H}^{\text{diag}}_{\alpha\beta}$ and $\mathbf{H}^{\text{diag}}_{\alpha\rho'}$ meaning that the interparameter tradeoffs among the isotropic-elastic parameters mainly come from the coupling effects between S-wave velocity and density. $\mathbf{H}^{\text{diag}}_{\beta\rho'}$ is very similar to $\mathbf{H}^{\text{diag}}_{\beta\beta}$ meaning that the coupling effects between S-wave velocity and density are dominated by S-wave velocity.

In this research, a vector $\mathbf{v}'$, which consists of regularly distributed spikes with a constant magnitude of 0.2, is designed, as indicated by the black squares in Figure 4.8d. Products of the multiparameter Hessian subblocks approximate row summations of the multiparameter

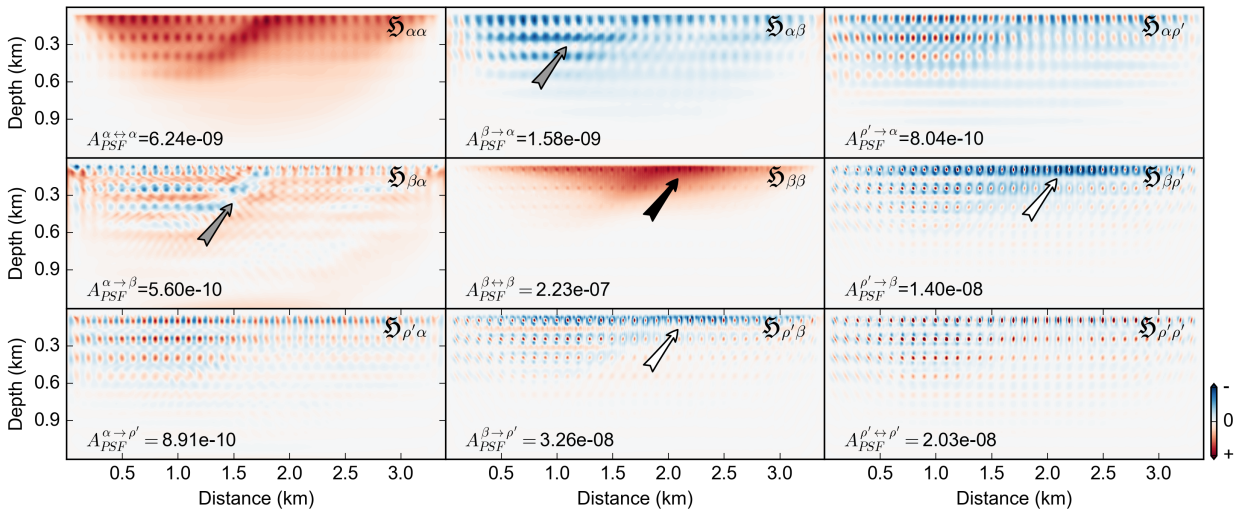

Figure 4.11: Products of multiparmaeter Hessian with vector $\mathbf{v}'$. The first column show the multiparameter Hessian-vector products $\mathfrak{H}_{\alpha\alpha} = \mathbf{H}_{\alpha\alpha}\mathbf{v}'$, $\mathfrak{H}_{\beta\alpha} = \mathbf{H}_{\beta\alpha}\mathbf{v}'$, and $\mathfrak{H}_{\beta\alpha} = \mathbf{H}_{\beta\alpha}\mathbf{v}'$. The second column show the multiparameter Hessian-vector products $\mathfrak{H}_{\alpha\beta} = \mathbf{H}_{\alpha\beta}\mathbf{v}'$, $\mathfrak{H}_{\beta\beta} = \mathbf{H}_{\beta\beta}\mathbf{v}'$, and $\mathfrak{H}_{\rho'\beta} = \mathbf{H}_{\rho'\beta}\mathbf{v}'$. The third column show the multiparameter Hessian-vector products $\mathfrak{H}_{\alpha\rho'} = \mathbf{H}_{\alpha\rho'}\mathbf{v}'$, $\mathfrak{H}_{\beta\rho'} = \mathbf{H}_{\beta\rho'}\mathbf{v}'$, and $\mathfrak{H}_{\rho'\rho'} = \mathbf{H}_{\rho'\rho'}\mathbf{v}'$.

Hessian, as illustrated in Figure 4.11. Strengths of interparameter contaminations generally match our predictions with multiparameter Hessian diagonals. Areas with strong interparameter tradeoffs are indicated by the grey and white arrows in Figure 4.11. Furthermore, comparing strengths of the off-diagonal Hessian-vector products (i.e., $\mathfrak{H}_{\beta\alpha}$) with those of diagonal Hessian-vector products (i.e., $\mathfrak{H}_{\beta\beta}$), it is concluded that the contaminations from α to β and ρ' are relatively weak and can be ignored. Contaminations from ρ' to α and β are also not very strong. However, the contaminations from β to α may degrade the update for α. Contaminations from β to ρ' may boost the density update by 1.8 times.

To verify these predictions and conclusions, the true *interparameter contamination kernels* are calculated by applying multiparameter Hessian to the true model perturbation vectors $\Delta\mathbf{m}_\alpha$, $\Delta\mathbf{m}_\beta$ and $\Delta\mathbf{m}_{\rho'}$ as shown in Figures 4.8g, 4.8h and 4.8i. The first row in Figure 4.12 show the standard sensitivity kernels K_α, K_β and $K_{\rho'}$, which are contaminated by mappings from other parameters. The second row in Figure 4.12 show the correct update kernel $K_{\alpha\leftrightarrow\alpha}$ and contamination kernels $K_{\beta\to\alpha}$ and $K_{\rho'\to\alpha}$. In the third row of Figure 4.12, the contamination kernel $K_{\alpha\to\beta}$, correct update kernel $K_{\beta\leftrightarrow\beta}$, and $K_{\rho'\to\beta}$ are illustrated from left to right. In the forth row of Figure 4.12, contamination kernels $K_{\alpha\to\rho'}$ and $K_{\beta\to\rho'}$ and

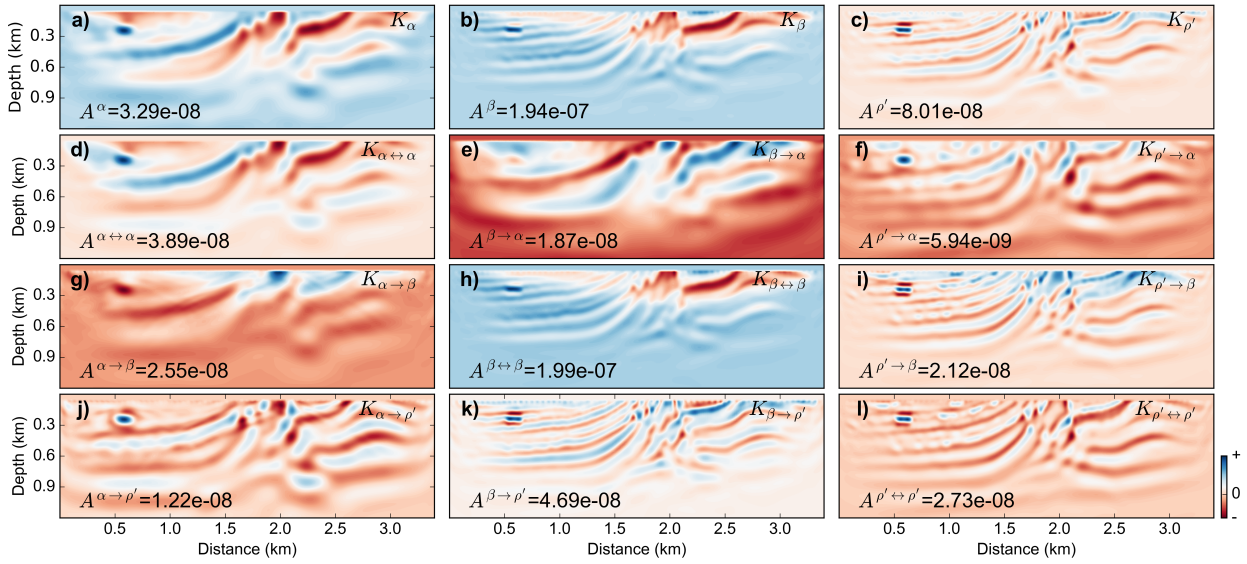

Figure 4.12: (a-c) illustrate the standard sensitivity kernels K_α, K_β, and $K_{\rho'}$; (d-f) show the correct update kernel $K_{\alpha\leftrightarrow\alpha}$ and contamination kernels $K_{\beta\to\alpha}$ and $K_{\rho'\to\alpha}$; (g-i) show the contamination kernel $K_{\alpha\to\beta}$, correct update kernels $K_{\beta\leftrightarrow\beta}$ and contamination kernel $K_{\rho'\to\beta}$; (j-l) show contamination kernels $K_{\beta\to\rho'}$ and $K_{\alpha\to\rho'}$ and correct update kernel $K_{\rho'\leftrightarrow\rho'}$. A represent maximum magnitudes of the kernels.

correct update kernel $K_{\rho'\to\rho'}$ are given.

Since magnitudes of the true model perturbation vectors change within the whole volume and the strengths of P-wave velocity perturbation are approximately 2 times and 4 times larger than those of S-wave velocity and density perturbations, the contamination kernels are not entirely consistent with the predictions by Hessian diagonals and Hessian-vector products shown in Figures 4.10 and 4.11 exactly. I interpret this is an indication of the complexity of the resolution problem in general. However, areas with strong elements in $K_{\beta\leftrightarrow\beta}$ generally match those of Hessian diagonals $\mathbf{H}^{\text{diag}}_{\beta\beta}$ (Figure 4.10) and Hessian-vector product $\boldsymbol{\mathfrak{H}}_{\beta\beta}$ (Figure 4.11), as indicated by the black arrows. Examining the contamination kernels $K_{\alpha\to\beta}$ and $K_{\beta\to\alpha}$ tells us that Hessian diagonals ($\mathbf{H}^{\text{diag}}_{\alpha\beta}$ in Figure 4.10) and Hessian-vector products ($\boldsymbol{\mathfrak{H}}_{\alpha\beta}$ and $\boldsymbol{\mathfrak{H}}_{\beta\alpha}$ in Figure 4.10) predict energy distributions of the interparameter tradeoffs, as indicated by the grey arrows. White arrows in $K_{\rho'\to\beta}$ and $K_{\beta\to\rho'}$ also indicate the areas with strong interparameter tradeoffs between S-wave velocity and density.

Comparing magnitudes of the correct updates and *interparameter contamination kernels*, it can be observed that $K_{\beta\to\beta}$ is very close to K_β meaning that the S-wave velocity suffers

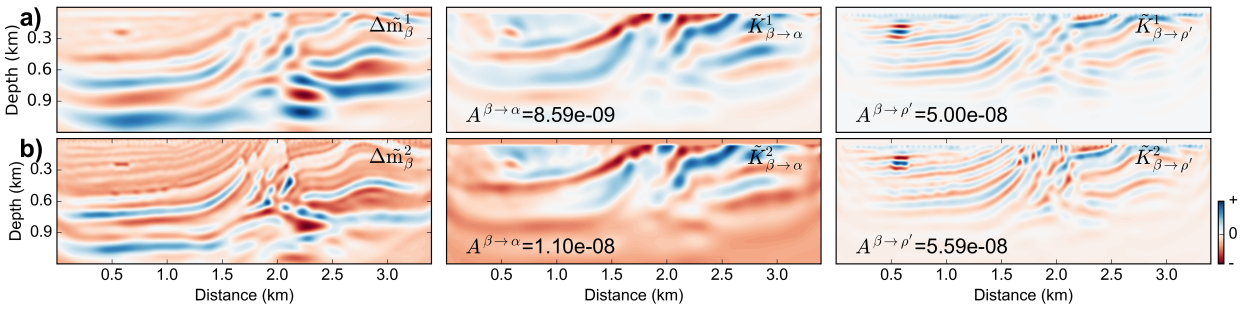

Figure 4.13: (a) shows the estimated model perturbation vector $\Delta\tilde{\mathbf{m}}^1_\beta$, approximate contamination kernels $\tilde{K}^1_{\beta\to\alpha}$ and $\tilde{K}^1_{\beta\to\rho'}$; (b) shows the estimated model perturbation vector $\Delta\tilde{\mathbf{m}}^2_\beta$, approximate contamination kernels $\tilde{K}^2_{\beta\to\alpha}$ and $\tilde{K}^2_{\beta\to\rho'}$.

limited contamination from α and ρ'. Furthermore, the correct update kernel $K_{\alpha\leftrightarrow\alpha}$ will be degraded by the contamination kernel $K_{\beta\to\alpha}$. Contamination kernel $K_{\beta\to\rho'}$ is approximately 1.7 times stronger than the correct update kernel $K_{\rho'\leftrightarrow\rho'}$, which will make density highly under- or overestimated. The contaminations from β to ρ' will dominate the estimated density structures. Note: during the inversion process, the contaminations can be reduced partially and the energy distributions of the interparameter contaminations may also change.

4.4.2.2 Mitigating the interparameter tradeoffs

To mitigate the contamination of S-wave velocity into other parameters, a novel inversion strategy is proposed with approximate contamination kernels. I first carry out inversion experiments by $k' = 8$ and 15 iterations, which provide estimated model perturbation vectors $\Delta\tilde{\mathbf{m}}^1_\beta$ and $\Delta\tilde{\mathbf{m}}^2_\beta$, as shown in Figures 4.13a and 4.13b. The estimated P-wave and density perturbations are dropped. Contamination kernels $\tilde{K}^1_{\beta\to\alpha}$, $\tilde{K}^1_{\beta\to\rho'}$, $\tilde{K}^2_{\beta\to\alpha}$, and $\tilde{K}^2_{\beta\to\rho'}$ are constructed by applying multiparameter Hessian off-diagonal blocks $\mathbf{H}_{\alpha\beta}$ and $\mathbf{H}_{\rho'\beta}$ to the estimated model vectors, as shown in Figures 4.13a and 4.13b. Magnitudes and characteristics of the approximate contamination kernels match the true contamination kernels $\tilde{K}_{\beta\to\alpha}$ and $\tilde{K}_{\beta\to\rho'}$ shown in Figure 4.12 very well. Because $\Delta\tilde{\mathbf{m}}^2_\beta$ is more resolved and better recovered than $\Delta\tilde{\mathbf{m}}^1_\beta$, $\tilde{K}^2_{\beta\to\alpha}$ and $\tilde{K}^2_{\beta\to\rho'}$ represent better approximations than $\tilde{K}^1_{\beta\to\alpha}$ and $\tilde{K}^1_{\beta\to\rho'}$.

The new updates kernels $\tilde{K}^1_\alpha$, $\tilde{K}^1_\beta$, $\tilde{K}^1_{\rho'}$, $\tilde{K}^2_\alpha$, $\tilde{K}^2_\beta$ and $\tilde{K}^2_{\rho'}$ are calculated by subtracting the approximate contamination kernels from the standard sensitivity kernels following equation (4.55), as shown in Figure 4.14. Magnitudes of the new updates kernels $\tilde{K}^1_{\rho'}$ and $\tilde{K}^2_{\rho'}$ have been

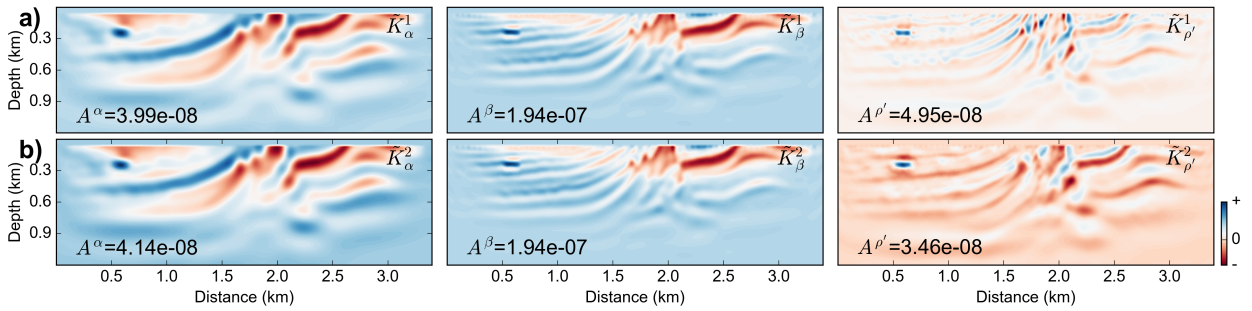

Figure 4.14: (a) show the new update kernels $\tilde{K}^1_\alpha$, $\tilde{K}^1_\beta$ and $\tilde{K}^1_{\rho'}$; (b) show the new update kernels $\tilde{K}^2_\alpha$, $\tilde{K}^2_\beta$ and $\tilde{K}^2_{\rho'}$.

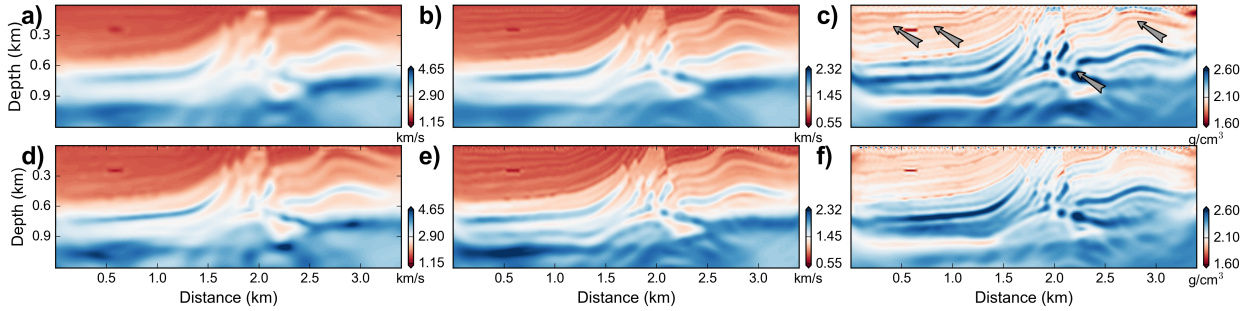

Figure 4.15: (a-c) show inverted P-wave velocity ($\tilde{\epsilon}_\alpha$=0.83), S-wave velocity ($\tilde{\epsilon}_\beta$=0.72) and density ($\tilde{\epsilon}_{\rho'}$=1.04) models with traditional simultaneous inversion strategy; (d-f) show the inverted P-wave velocity ($\tilde{\epsilon}_\alpha$=0.76), S-wave velocity ($\tilde{\epsilon}_\beta$=0.67) and density ($\tilde{\epsilon}_{\rho'}$=0.83) models using new inversion method with approximate contamination kernels.

reduced by approximately 38.2% and 56.8%. In particular, it is observed that the features of new update kernel $\tilde{K}^2_{\rho'}$ for density are very close to the characteristics of true update kernel $K_{\rho'\leftrightarrow\rho'}$ shown in Figure 4.12, which means that the contaminations from S-wave velocity to density have been suppressed.

The P-wave velocity α, S-wave velocity β and density ρ' are inverted with traditional simultaneous inversion strategy. Multiscale approach is adopted for reducing the nonlinearity by expanding the frequency band from [3 Hz, 5Hz] to [3 Hz, 8Hz]. With each frequency band, 20 iterations are performed. Figures 4.15a, 4.15b and 4.15c show the inverted α, β and ρ' models after 40 iterations with traditional simultaneous inversion strategy. Following the work-flow illustrated in Table 4.1, I first simultaneously update α, β and ρ' models by $k' = 15$ iterations and the inverted models are dropped but only keep the estimated model perturbation $\Delta\tilde{\mathbf{m}}_\beta$. Then, model parameters α, β and ρ' are inverted again from initial

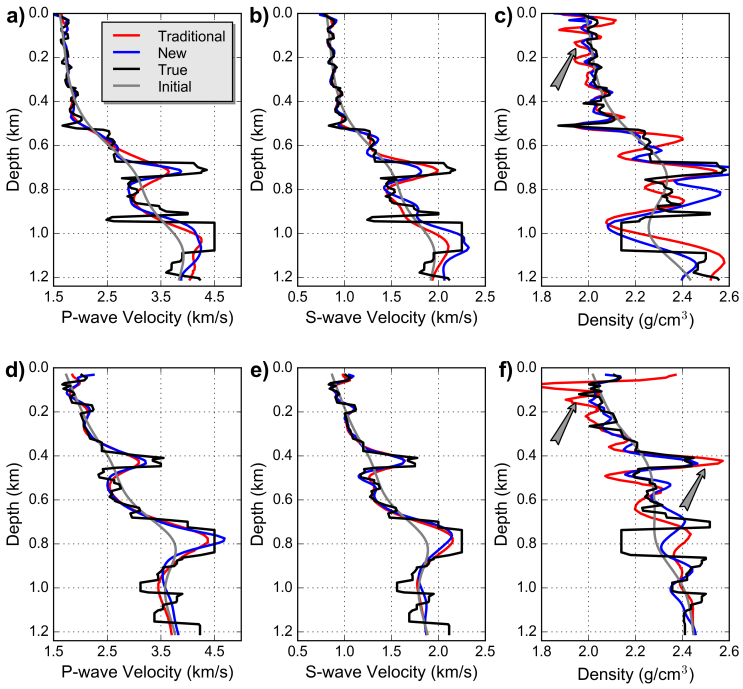

Figure 4.16: (a-c) show the well log data of P-wave velocity, S-wave velocity, and density models at 0.5 km; (d-f) show the well log data at 3.0 km. The red and grey curves indicate the true and initial models. The blue and green lines indicate the inverted models by traditional simultaneous inversion strategy and new inversion strategy.

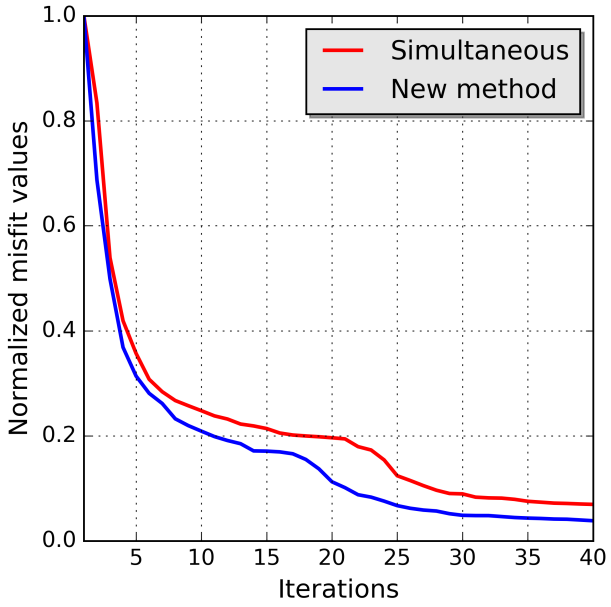

Figure 4.17: Convergence history comparison of traditional simultaneous inversion strategy (red curve) and new inversion strategy (blue curve) for Marmousi model example.

models by 10 iterations. In this inversion loop, at each iteration approximate contamination kernels are constructed and the models are updated with new kernels as indicated in equation (4.55). This process is then repeated every 10 iterations. Figures 4.15d, 4.15e and 4.15f show the inverted P-wave velocity, S-wave velocity and density models with the new inversion strategy after 40 iterations. The computation cost is 2.5 times more expensive than traditional simultaneous inversion strategy.

S-wave velocity is best inverted and more resolved than P-wave velocity. P-wave velocity is poorly recovered but limited interparameter contamination artifacts can be observed. With traditional simultaneous inversion strategy, the S-wave velocity structures are mapped into the estimated density model as indicated by the arrows in Figures 4.15c, 4.16c and 4.16f. Positive S-wave velocity perturbations make density overestimated and negative S-wave velocity perturbations make density underestimated. With the new inversion strategy, the imprints in the inverted density model have been suppressed effectively, as indicated by arrows in Figures 4.15f, 4.16c and 4.16f. Furthermore, the inverted P-wave velocity (Figure 4.15d) and S-wave velocity (Figure 4.15e) are also enhanced. The new inversion approach is also able to provide faster convergence, as shown in Figure 4.17.

4.4.2.3 Resolution analysis

Approximate eigenvalue volume (equation (4.64)) and extended multiparameter point spread functions (EMPSFs) are used to quantify resolution of the inverted models with different inversion strategies. Figures 4.18a, 4.18b and 4.18c show the approximate eigenvalue volumes obtained with 2 random Hessian-vector applications. The approximate eigenvalue volumes of S-wave velocity ($\mathbf{Eig}_{\beta\beta}$) are closer to 1 than those of P-wave velocity and density, which means that S-wave velocity is better recovered than P-wave velocity and density. Magnitudes of the approximate eigenvalue volumes decrease with increasing depths meaning that shallow parts of models are better recovered than deep parts.

Spike model perturbations $\Delta\mathbf{m}_\alpha = 0.2$ km/s, $\Delta\mathbf{m}_\beta = 0.2$ km/s and $\Delta\mathbf{m}_{\rho'} = 0.2$ g/cm^3 are applied at local position $\mathbf{z}_1 = (0.515 \text{ km}, 0.275 \text{ km})$ (as shown in Figure 4.8e) respectively, which are used to measure the local spatial and interparameter tradeoffs of the inverted models. The traditional MPSFs are plotted in Figure 4.19a. The MPSFs representing interparameter contaminations are normalized with those representing correct model estimations. Strong contaminations from S-wave velocity to P-wave velocity and density are observed as indicated by the grey arrows. The spike model perturbations are then reconstructed with 10 conjugate-gradient iterations, which gives the EMPSFs, as presented in Figures 4.19b. Contaminations from S-wave velocity to P-wave velocity and density are reduced by approximately 23.8% and 47.1%, as indicated by the grey arrows. Furthermore, compared to MPSFs, EMPSFs are more de-blurred. This is strongly suggestive that the local spatial and interparameter tradeoffs provided by traditional MPSFs may not be accurate and the inverted models by traditional simultaneous inversion strategy suffer strong interparameter tradeoffs. In Figures 4.19c, the EMPSFs obtained with conjugate-gradient method following the work-flow with approximate contamination kernels are given. The contaminations from S-wave velocity to P-wave velocity and density are reduced effectively, as indicated by grey arrows. This means that the inverted models (Figures 4.15a, 4.15b and 4.15c) by new inversion strategy suffer little interparameter contaminations.

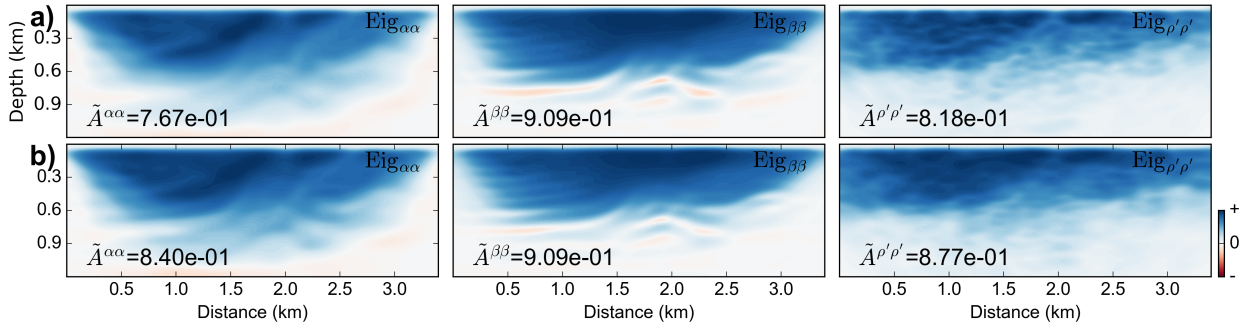

Figure 4.18: Approximate eigenvalue volumes of the inverted models. Figures (a-c) show approximate eigenvalue volumes $\mathbf{Eig}_{\alpha\alpha}$, $\mathbf{Eig}_{\beta\beta}$, and $\mathbf{Eig}_{\rho'\rho'}$ for inverted P-wave velocity, S-wave velocity and density by traditional simultaneous inversion strategy; Figures (d-f) show the corresponding approximate eigenvalue volumes by the new inversion strategy.

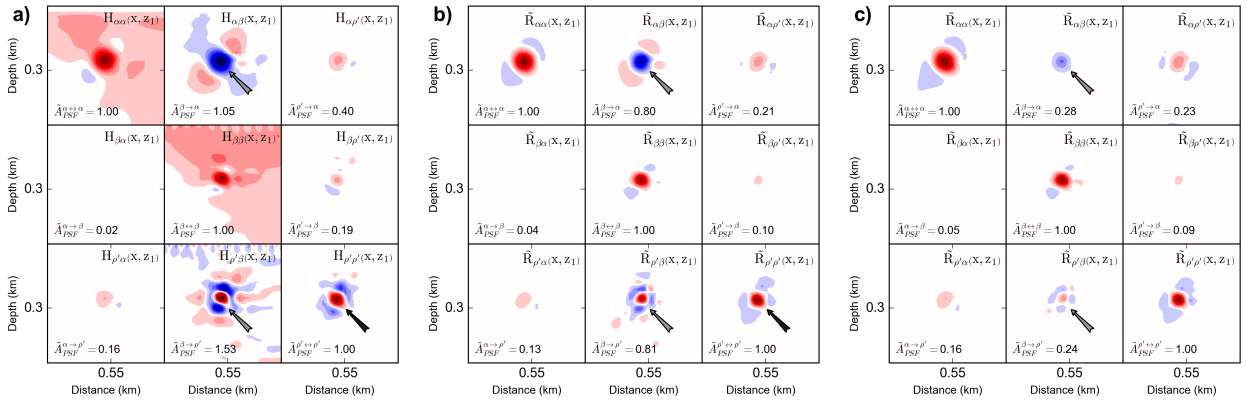

Figure 4.19: (a) show the traditional MPSFs after normalization at position $\mathbf{z}_1$; (b) show the normalized EMPSFs with 10 conjugate-gradient iterations; (c) show the normalized EMPSFs constructed with new inversion work-flow (Table 4.1).

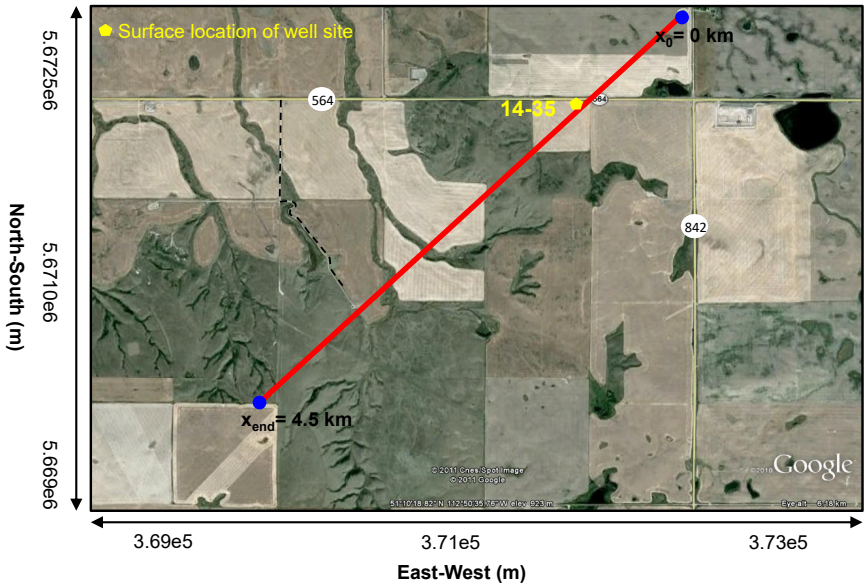

Figure 4.20: The location of seismic line and well (14-35) in Hussar experiment (Margrave et al., 2012). Note: I have reset the coordinate of the seismic line for FWI. I assume that initial location of the seismic line starts at x_0=0 km and ends at x_{end}=4.5 km, as indicated by the blue circles.

4.4.2.4 Hussar dataset application

At the end, the proposed strategies are applied to invert isotropic-elastic parameters with Hussar practical seismic dataset and quantify resolution of the inverted models. In September 2011, CREWES (Consortium for Research in Elastic Wave Exploration Seismology) initiated a seismic experiment in Hussar area, which is about 100 km east of Calgary, Alberta, Canada. The objective of this experiment was to maximize the low frequency content of the seismic data (Margrave et al., 2012), and to acquire a land dataset maximally suitable for full-waveform inversion methods. The 2D seismic survey line is 4.5 km in length. Figure 4.20 show the locations of the seismic line and well log 14-35. The seismic experiments were carried out with dynamite and vibroseis sources and different receiver types. In this research, I use the multicomponent data recorded by 10 Hz 3C (three-component) geophones with dynamite sources for inversion. A total of 269 sources (2 kg charge at 15 m in depth) are arranged regularly with a spacing of 20 m. A total of 448 geophones are distributed with a spacing of 10 m.

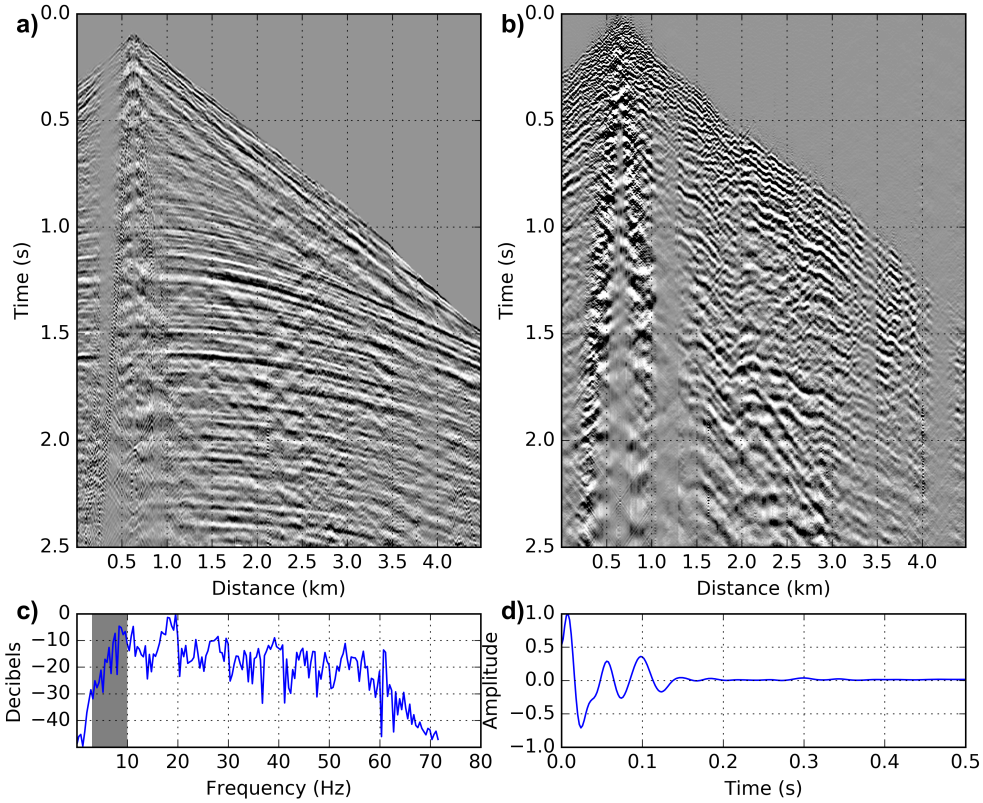

Figure 4.21: (a) and (b) show the preprocessed vertical (z) and radial (x) component shot gathers at position of 0.6 km in horizontal distance; (c) shows the amplitude spectrum of the data. The shaded area means frequency band of [3Hz, 10Hz]. (d) shows the estimated minimum phase wavelet with dominant frequency of 25 Hz.

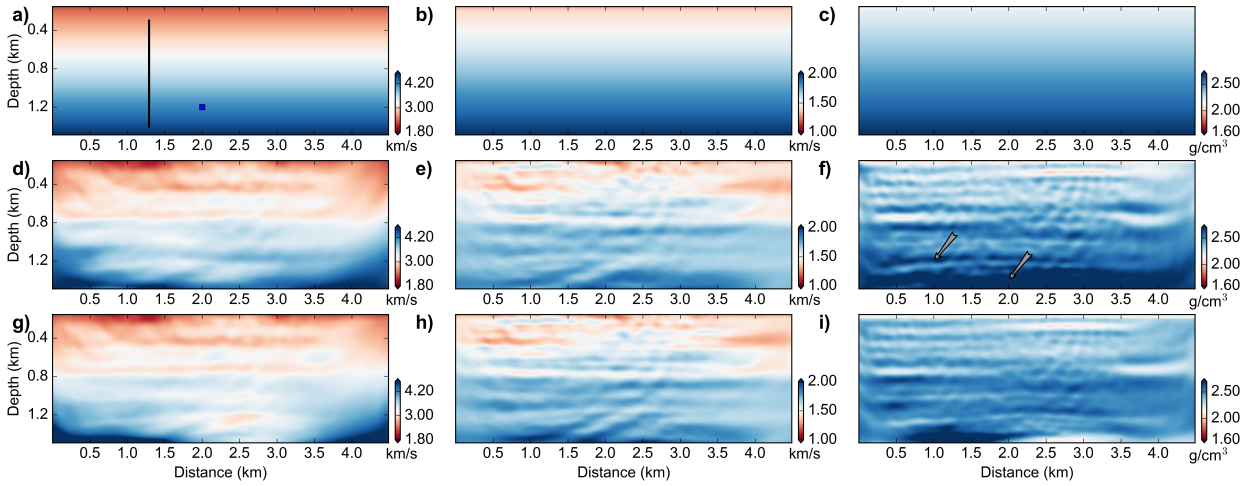

Figure 4.22: (a-c) show the initial P-wave velocity, S-wave velocity and density models; (d-f) show the inverted P-wave velocity, S-wave velocity and density models using traditional simultaneous inversion strategy; (g-i) show the inverted P-wave velocity, S-wave velocity and density models using new inversion method. The black line in (a) indicates the position of well log 14-35. The blue square in (a) indicates the location $\mathbf{z}_1$=(2.0 km, 1.2 km) for local spatial and interparameter tradeoffs analysis.

The raw seismic shot gathers are preprocessed with a series of steps. Automatic gain control (AGC) is first applied for amplitude recovery. Surface waves and monochromatic noise are suppressed with F-K filtering. Elevation statics and residual statics are applied to compensate the topographic variations and near-surface lithological variations. The seismic data is finally band-pass filtered within the frequency band of [3Hz, 60Hz]. Figures 4.21a and 4.21b show the preprocessed vertical (z) and radial (x) component data. Figure 4.21c shows the amplitude spectrum of the data. Frequency band of [3Hz, 10Hz] is used for inversion, as indicated by the shaded area. A minimum phase wavelet with dominant frequency of 25 Hz is estimated from seismic data and used for forward modelling, as illustrated in Figure 4.21d.

Figures 4.22a, 4.22b and 4.22c show the linear initial P-wave velocity, S-wave velocity and density models. The well log (14-35) is located at about 1.29 km in horizontal distance, as indicated by the black line in Figure 4.22a. I first simultaneously update P-wave velocity, S-wave velocity and density by expanding the frequency band from [3Hz, 5Hz] to [3Hz, 8Hz] and then [3Hz, 10Hz] with 10 iterations for each frequency band. The inverted P-wave velocity, S-wave velocity and density models using this traditional inversion strategy are illustrated

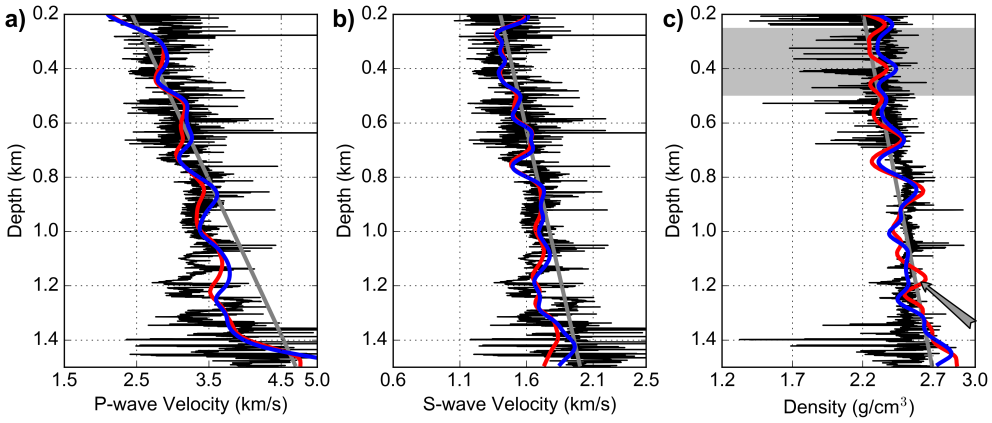

Figure 4.23: (a-c) show the well log data comparison of P-wave velocity, S-wave velocity and density respectively. Black and gray curves are true well log data and initial models. Red and blue curves are the inverted models by traditional simultaneous inversion method and new inversion method respectively.

in Figures 4.22d, 4.22e and 4.22f. I then carry out inversion experiments using new inversion strategy with approximate contamination kernels following the work-flow shown in Table 4.1. At each frequency band, I need to simultaneously update the model parameters by k'=15 iterations for estimating the approximate S-wave velocity perturbation vector $\Delta\tilde{\mathbf{m}}_\beta$. Figures 4.22g, 4.22h and 4.22i show the inverted P-wave velocity, S-wave velocity and density models using the new inversion method. Some artifacts appear in the inverted models. The geological layers are resolved and most of them are flat, which are consistent with the previous studies of impedance inversion (Lloyd, 2013; Cui, 2015; Esmaeili, 2016). Figures 4.23a, 4.23b and 4.23c show the well log data comparison of P-wave velocity, S-wave velocity and density models respectively. The inverted P-wave velocity and S-wave velocity models generally match the well log data. However, it appears that the shallow parts of inverted density model (Figure 4.22f) by traditional simultaneous inversion strategy are underestimated, as indicated by the shaded area in Figure 4.23c. Furthermore, artifacts appear in the deep parts of the inverted density model as indicated by the arrows in Figures 4.22f and 4.23c. In the inverted density model (4.22i) by new inversion method, the shallow parts are better recovered and the artifacts in deep parts are suppressed. The new inversion method also provides faster convergence as shown in Figure 4.24.

For resolution analysis, in Figure 4.25a, approximate eigenvalue volumes $\mathbf{Eig}_{\alpha\alpha}$, $\mathbf{Eig}_{\beta\beta}$

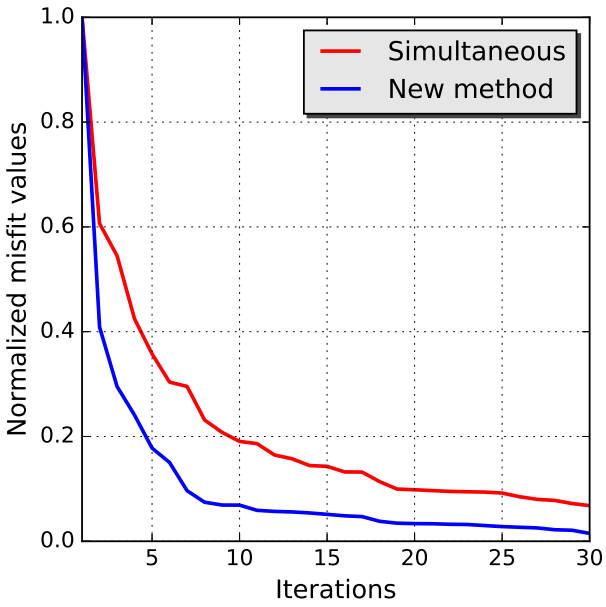

Figure 4.24: Convergence history comparison of traditional simultaneous inversion method (red) and new inversion method (blue) for Hussar seismic dataset.

and $\mathbf{Eig}_{\rho'\rho'}$ of the inverted P-wave velocity, S-wave velocity and density models by traditional simultaneous inversion method are plotted. In Figure 4.25b, the approximate eigenvalue volumes of the inverted models generated using the new inversion method are given. Both of these methods are able to recover amplitudes of the model parameters very well. The spatial and interparameter tradeoffs of the inverted models at location of $\mathbf{z}_1$=(2.0 km, 1.2 km) are evaluated, as indicated by blue square in Figure 4.22a. Figure 4.26a show the normalized EMPSFs with 10 conjugate-gradient iterations. Figure 4.26b show the normalized EMPSFs with 10 conjugate-gradient iterations following the work-flow of new inversion method. In Figure 4.26a, it is observed that the EMPSFs $\mathbf{R}_{\alpha\beta}(\mathbf{x},\mathbf{z})$ and $\mathbf{R}_{\rho'\beta}(\mathbf{x},\mathbf{z})$ representing the contaminations from S-wave velocity to P-wave velocity and density are still strong, which means the inverted P-wave velocity and density models (Figures 4.22d and 4.22f) by traditional simultaneous inversion method still suffer interparameter contaminations. In Figure 4.26b, the EMPSFs $\mathbf{R}_{\alpha\beta}(\mathbf{x},\mathbf{z})$ and $\mathbf{R}_{\rho'\beta}(\mathbf{x},\mathbf{z})$ are very weak, which means that with the new inversion method, the interparameter contaminations from S-wave velocity to P-wave velocity and density have been reduced. These numerical experiments show that the new inversion method is able to provide high resolution P-wave velocity and S-wave velocity

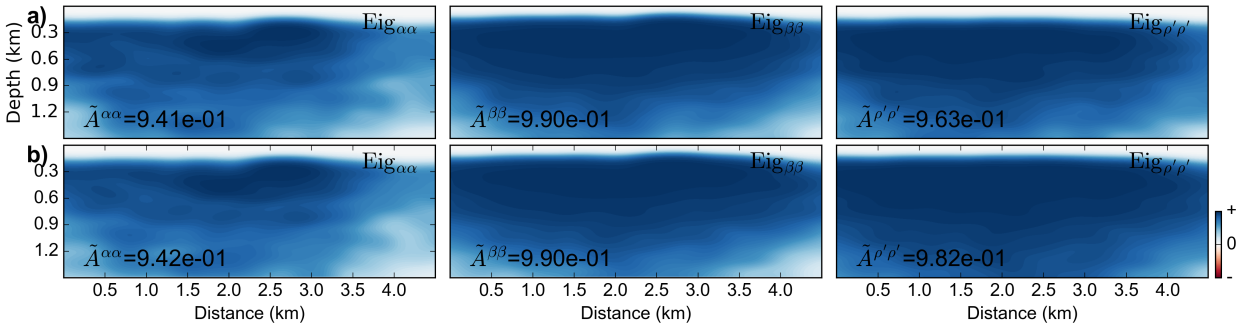

Figure 4.25: (a) illustrate the approximate eigenvalue volumes $\mathbf{Eig}_{\alpha\alpha}$, $\mathbf{Eig}_{\beta\beta}$ and $\mathbf{Eig}_{\rho'\rho'}$ of the inverted models by traditional simultaneous inversion method; (b) illustrate the approximate eigenvalue volumes $\mathbf{Eig}_{\alpha\alpha}$, $\mathbf{Eig}_{\beta\beta}$ and $\mathbf{Eig}_{\rho'\rho'}$ of the inverted models generated using the new inversion method.

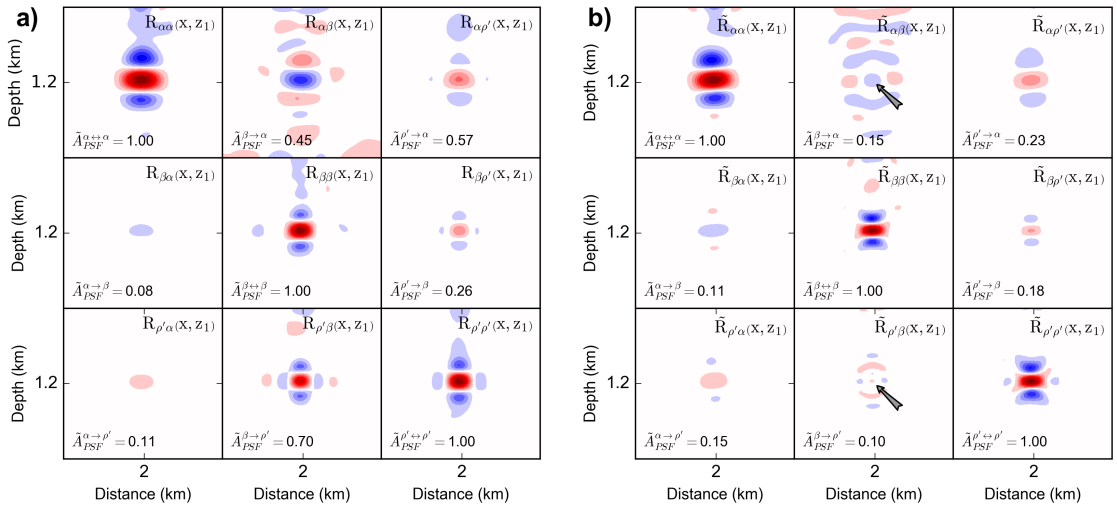

Figure 4.26: (a) show the collection of EMPSFs at local position $\mathbf{z}_1 = (2.0\text{ km}, 1.2\text{ km})$ with conjugate-gradient algorithm; (b) show the collection of EMPSFs at $\mathbf{z}_1$ with conjugate-gradient algorithm following the work-flow of new inversion method (Table 4.1).

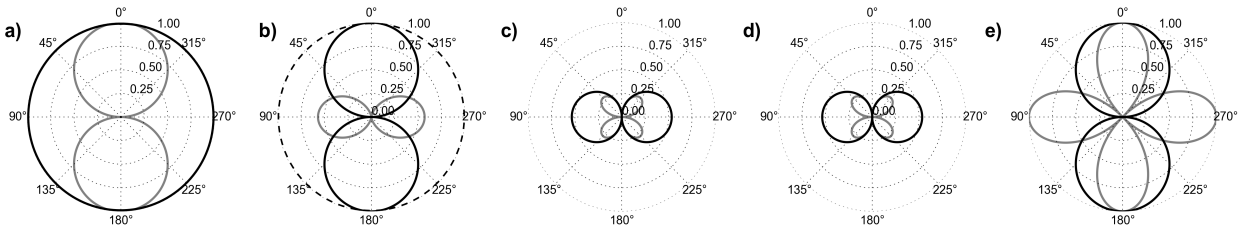

Figure 4.27: (a-e) show the SH-SH, P-P, P-SV, SV-P and SV-SV scattering patterns due to the perturbations of isotropic-elastic parameters. Solid-grey and solid-black curves indicate the scattering patterns due to perturbations of shear modulus μ and density ρ. In (a), the dash-black curve represents the P-P scattering pattern due the perturbation of bulk modulus κ.

models. Furthermore, the interparameter contaminations from S-wave velocity to P-wave velocity and density can be suppressed, which will provide more convincing isotropic-elastic parameters for reservoir characterization.

4.4.3 The influence of different parameterizations in isotropic-elastic FWI

In this section, the inverted isotropic-elastic models with another two types of parameterizations: modulus-density parameterization (bulk modulus κ, shear modulus μ and density ρ) and impedance-density parameterization (P-wave impedance IP=$\alpha\rho''$, S-wave impedance IS=$\beta\rho''$ and density ρ'') are given. Figures 4.27 and 4.28 show the scattering patterns of isotropic-elastic parameters within modulus-density and impedance-density parameterizations. It can be seen that for modulus-density parameterization, the scattering patterns of different parameterizations overlap significantly, which means that the coupling effects between different parameters are strong. The coupling effects will decrease the convergence rate for inverting the model parameters. For impedance-density parameterization, the scattered wavefields due to density perturbations are mostly forward propagated, which means that when using reflection survey, it will be very difficult to recover density structures with impedance-density parameterization.

Inversion experiments with velocity-density, modulus-density and impedance-density parameterizations are then carried out for comparison. The true models and initial models are shown in Figure 4.8. Figure 4.29 show the standard sensitivity kernels and true *interparameter contamination kernels* among bulk modulus κ, shear modulus μ and density ρ within

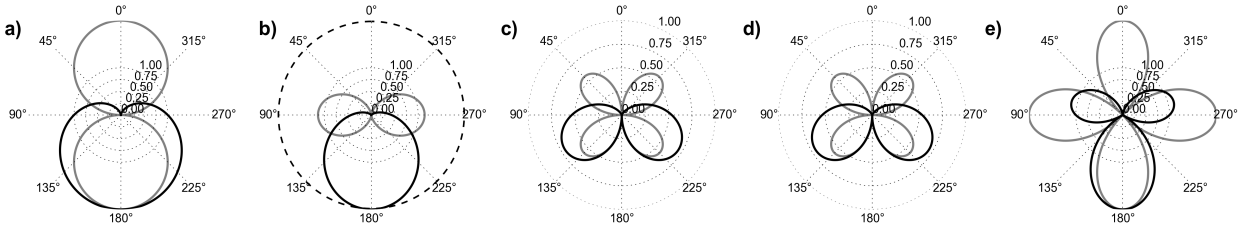

Figure 4.28: (a-e) show the SH-SH, P-P, P-SV, SV-P and SV-SV scattering patterns due to the perturbations of isotropic-elastic parameters. Solid-grey and solid-black curves indicate the scattering patterns due to perturbations of S-wave impedance IS and density ρ''. In (a), the dash-black curve represents the P-P scattering pattern due the perturbation of P-wave impedance IP.

modulus-density parameterization. It can be seen that the standard sensitivity kernel K_ρ is quite different from the correct update kernel $K_{\rho\leftrightarrow\rho}$. However, the contamination kernel $K_{\mu\rightarrow\rho}$ is very similar to the standard sensitivity kernel K_ρ. This is strongly suggestive that the update for density ρ is dominated by the contaminations from shear modulus μ to density ρ. Figure 4.30 shows the standard sensitivity kernels and true *interparameter contamination kernels* among P-wave impedance IP, S-wave impedance IS and density ρ'' within impedance-density parameterization. Similarly, the standard update kernel for density ρ'' is dominated by the contamination from S-wave impedance to density ρ''.

The inversion experiments are carried out using three frequency bands of [3Hz, 5Hz], [3Hz, 8Hz] and [3Hz, 10Hz] with 60 iterations at each frequency band. The same true models and initial models are used for inversion with the 3 different parameterizations. Figures 4.31, 4.32 and 4.33 show the corresponding true models, initial models and inverted models after 180 iterations. Note: here I use non-linear conjugate-gradient method for inversion. In Figure 4.34, the convergence histories at frequency band of [3Hz, 5Hz] are plotted for comparison. In Figure 4.31, P-wave velocity and S-wave velocity are well reconstructed. The density structures are also reconstructed even though there are still some interparameter contaminations. In Figure 4.32, the shear modulus μ is recovered best. The inverted density model is distorted, which may be caused by the contaminations from shear modulus. Magnitudes of the recovered bulk modulus model are very weak. In Figure 4.33, the reconstructed density structures are distorted significantly, which may be caused by the strong interparameter contaminations from S-wave impedance. This observation also verifies our analysis with scat-

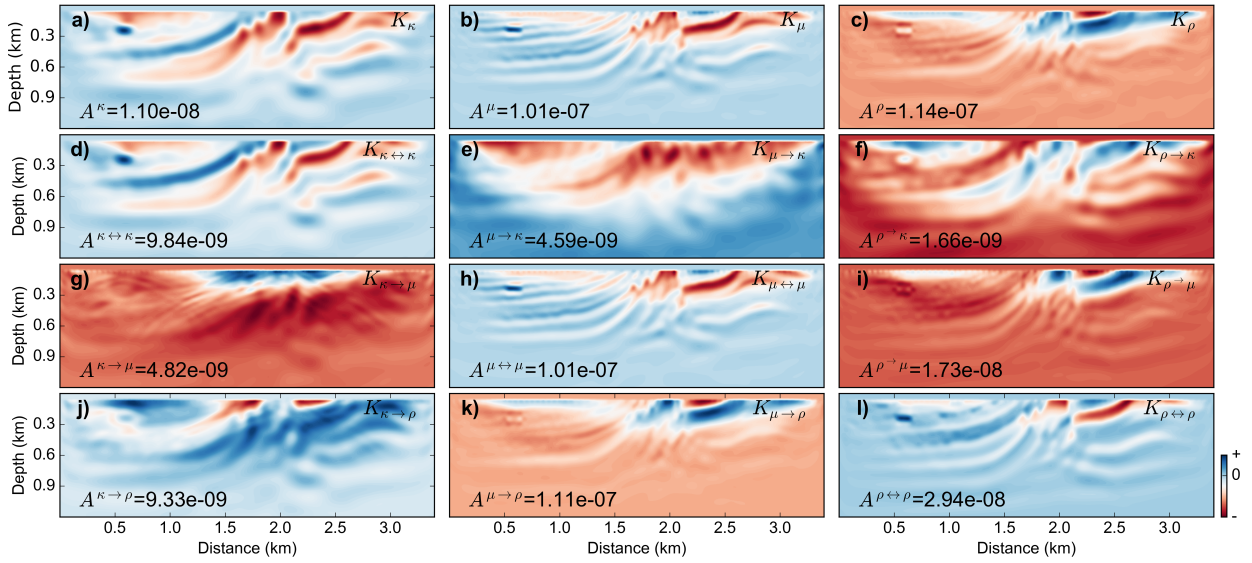

Figure 4.29: (a-c) illustrate the standard sensitivity kernels K_κ, K_μ, and K_ρ; (d-f) show the correct update kernel $K_{\kappa\leftrightarrow\kappa}$ and contamination kernels $K_{\mu\rightarrow\kappa}$ and $K_{\rho\rightarrow\kappa}$; (g-i) show the contamination kernel $K_{\kappa\rightarrow\mu}$, correct update kernels $K_{\mu\leftrightarrow\mu}$ and contamination kernel $K_{\rho\rightarrow\mu}$; (j-l) show contamination kernels $K_{\kappa\rightarrow\rho}$ and $K_{\mu\rightarrow\rho}$ and correct update kernel $K_{\rho\leftrightarrow\rho}$. A represent maximum magnitudes of the kernels.

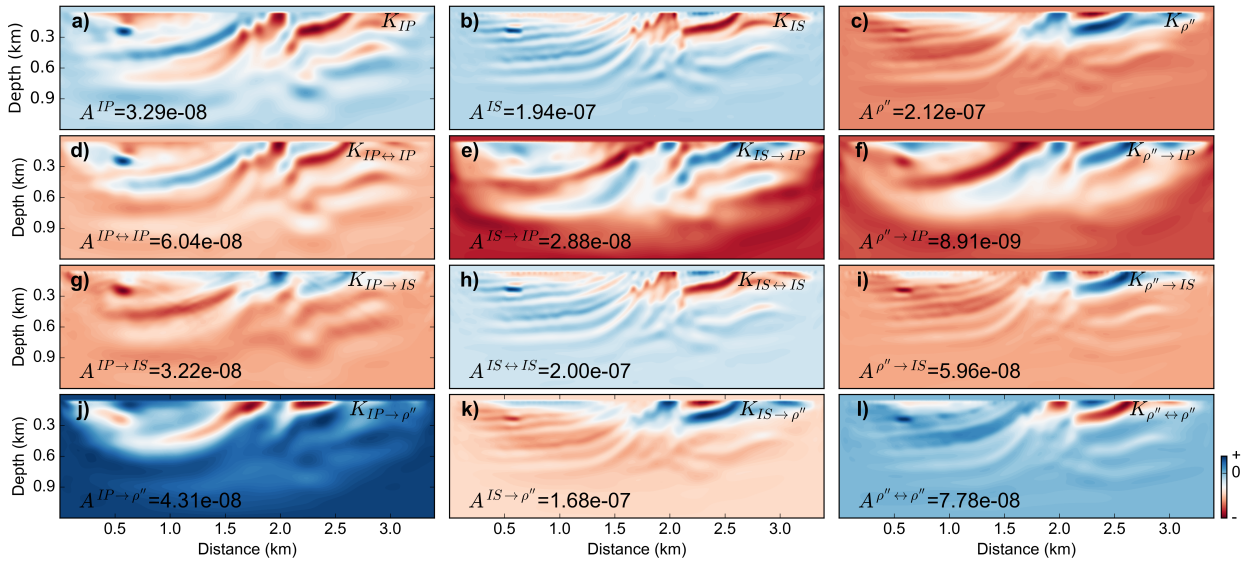

Figure 4.30: (a-c) illustrate the standard sensitivity kernels K_{IP}, K_{IS}, and $K_{\rho''}$; (d-f) show the correct update kernel $K_{\mathrm{IP}\leftrightarrow\mathrm{IP}}$ and contamination kernels $K_{\mathrm{IS}\rightarrow\mathrm{IP}}$ and $K_{\rho''\rightarrow\mathrm{IP}}$; (g-i) show the contamination kernel $K_{\mathrm{IP}\rightarrow\mathrm{IS}}$, correct update kernels $K_{\mathrm{IS}\leftrightarrow\mathrm{IS}}$ and contamination kernel $K_{\rho''\rightarrow\mathrm{IS}}$; (j-l) show contamination kernels $K_{\mathrm{IP}\rightarrow\rho''}$ and $K_{\mathrm{IS}\rightarrow\rho''}$ and correct update kernel $K_{\rho''\leftrightarrow\rho''}$. A represent maximum magnitudes of the kernels.

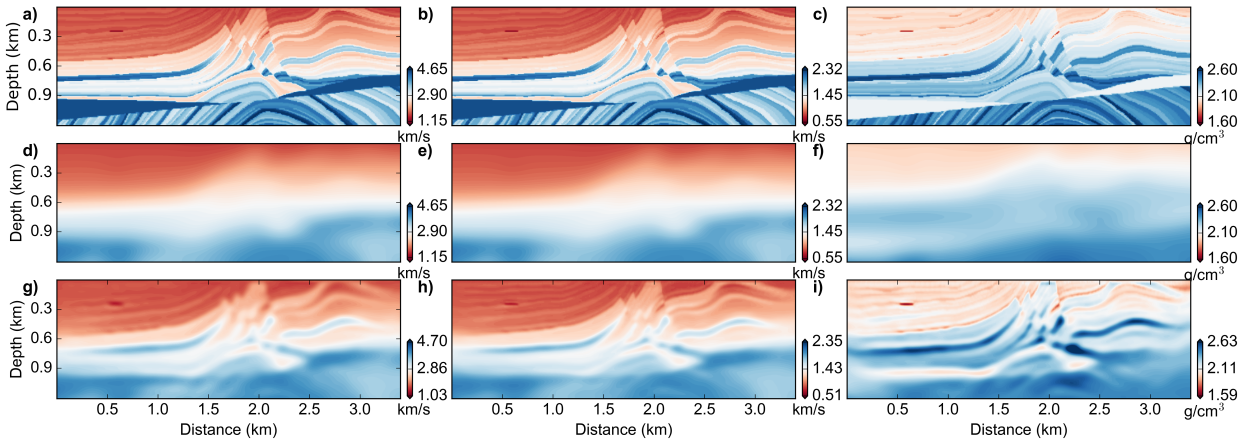

Figure 4.31: (a-c) show the true P-wave velocity, S-wave velocity and density models; (d-f) show the corresponding initial P-wave velocity, S-wave velocity and density models; (g-i) show the corresponding inverted P-wave velocity, S-wave velocity and density models.

tering patterns in Figure 4.28. Because the scattered wavefields due to density perturbations in impedance-density parameterization mostly forward scattered, recorded data on surface is mainly caused by P-wave impedance and S-wave impedance perturbations, which makes it more difficult to recover density structures. In Figure 4.34, velocity-density parameterization provides the fastest convergence rate. Hence, velocity-density parameterization is still the best choice to recover isotropic-elastic parameters among these three parameterizations.

4.5 Discussion

Interparameter tradeoffs are strongly influenced by source-receive illumination (or acquisition geometry). In this chapter, the interparameter tradeoffs with perfect acquisition geometry are studied using a simple Gaussian-anomaly model and reflection acquisition geometry using Marmousi model. In transmission tomography (i.e., cross-well survey or earthquake seismology), the strengths and characteristics of the interparameter contaminations may be different from the conclusions and results presented in this chapter. Hence, for inverse problems with different models and acquisition geometries, the interparameter tradeoffs should be reevaluated following the strategies presented in this chapter.

Various misfit functions (i.e., envelope, instantaneous phase and traveltime misfit functions) based on different measurements have been studied for full-waveform inversion (Bozdag

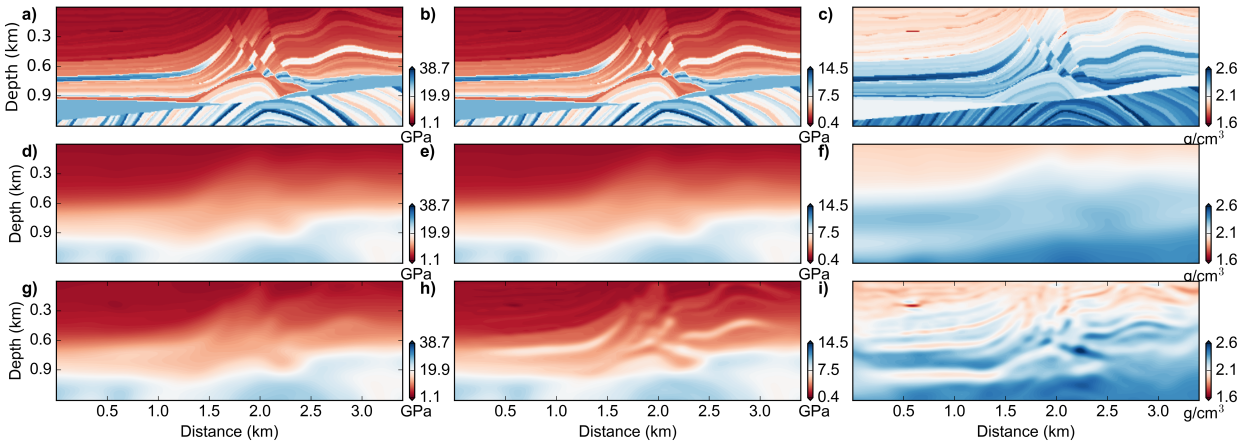

Figure 4.32: (a-c) show the true bulk modulus κ, shear modulus μ and density ρ models; (d-f) show the corresponding initial bulk modulus κ, shear modulus μ and density ρ models; (g-i) show the corresponding inverted bulk modulus κ, shear modulus μ and density ρ models.

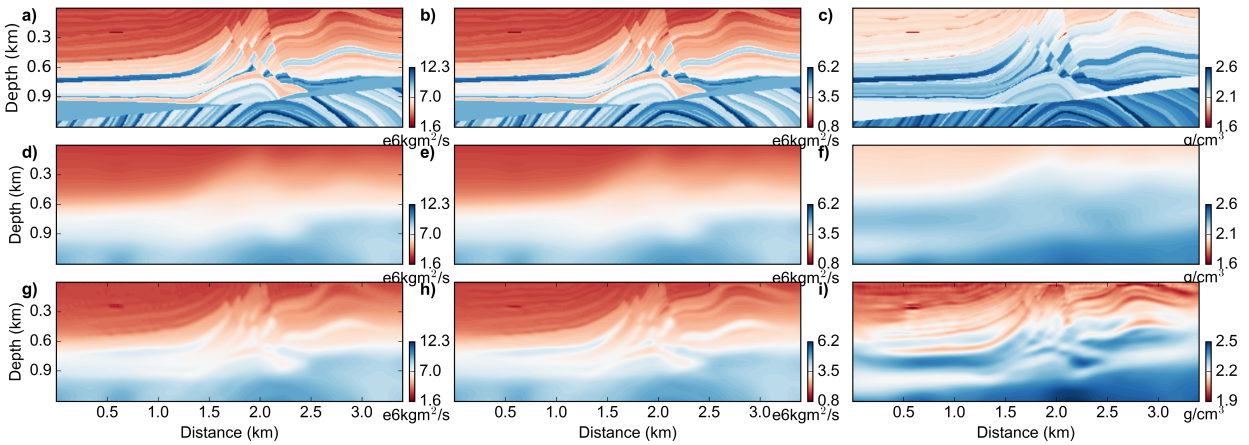

Figure 4.33: (a-c) show the true P-wave impedance IP, S-wave impedance IS and density ρ'' models; (d-f) show the corresponding initial P-wave impedance IP, S-wave impedance IS and density ρ'' models; (g-i) show the corresponding inverted P-wave impedance IP, S-wave impedance IS and density ρ'' models.

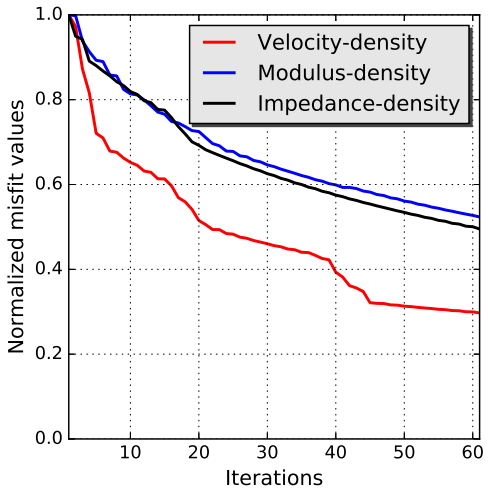

Figure 4.34: Convergence rates comparison for various parameterizations in isotropic-elastic FWI ([3Hz, 5Hz] frequency band). The red, blue and black curves indicate velocity-density, modulus-density and impedance-density parameterization respectively.

et al., 2011). Different physical parameters are sensitive to different measurements (i.e., amplitude and traveltime). The interparameter tradeoffs with the common waveform difference based misfit function are studied. It is also necessary to assess the interparameter tradeoffs in isotropic-elastic FWI for different misfit functions.

4.6 Conclusions

Origins of interparameter tradeoffs in isotropic-elastic FWI have been revealed with *interparameter contamination kernels*. Strengths and characteristics of the interparameter contaminations in isotropic-elastic FWI are quantified locally or within the whole volume by applying multiparameter Hessian to various types of probes. Two approaches (adjoint-state and finite-difference) are examined to construct the multiparameter Hessian matrix-vector products. This chapter reveals that S-wave velocity perturbations produce relatively strong contaminations into density updates and phase-revered contaminations into P-wave velocity updates. These contaminations make density structures highly under- and overestimated.

A novel inversion strategy has been recommended to reduce the contaminations from S-wave velocity to other parameters based on approximate contamination kernels. Numerical examples are given to illustrate that this new inversion strategy is able to provide more

convincing and reliable density estimations in isotropic-elastic FWI. Approximate eigenvalue volume is employed to evaluate resolution of inverted models within the whole volume. Both of local spatial and interparameter tradeoffs of the inverted models are evaluated with extended multiparameter point spread functions (EMPSFs), which provide more accurate measurements of the local resolution compared to traditional MPSFs. The proposed strategies are finally applied on Hussar practical seismic dataset. According to the inverted models with different parameterizations, it is concluded that the velocity-density parameterization is still a better choice than modulus-density and impedance-density parameterizations.

Appendix B

Scattering patterns of isotropic-elastic parameters with various parameterizations

Scattering coefficients are derived based on the assumption that an incident plane-wave is scattered due to a local heterogeneity embedded in an isotropic-elastic background. Scattering potential is defined as the difference between the wave operators in the perturbed and unperturbed medium and always reside inside Born approximation integral (Matson, 1997; Stolt and Weglein, 2012; Innanen, 2012):

$$d_n\left(\mathbf{x}_r,\mathbf{x}_s,t\right)\approx\int_0^t\int_{\Omega(\mathbf{x})}G_{nj}\left(\mathbf{x}_r,\mathbf{x},t\right)\mathcal{V}_{jk}\left(\mathbf{x}\right)f_i\left(\mathbf{x}_s,t'\right)G_{ki}\left(\mathbf{x},\mathbf{x}_s,t-t'\right)d\mathbf{x}dt', \tag{B.1}$$

where f_i indicates the source function at the ith direction and $\mathcal{V}_{jk}\left(\mathbf{x}\right)$ is the scattering potential: $\mathcal{V}_{jk}\left(\mathbf{x}\right)=\sum_p \acute{\mathrm{A}}_j^p a_p\left(\mathbf{x}\right)\grave{\mathrm{A}}_k^p$, where $a_p\left(\mathbf{x}\right)$ is the relative perturbation of model parameter p at $\mathbf{x}$, $\grave{\mathrm{A}}_k^p$ and $\acute{\mathrm{A}}_j^p$ are weighting coefficients associated with time and spatial derivatives for incident and scattered waves respectively. With integration by parts, derivatives in $\grave{\mathrm{A}}_k^p$ and $\acute{\mathrm{A}}_j^p$ can be applied on the Green's functions:

$$d_n\left(\mathbf{x}_r,\mathbf{x}_s,t\right)\approx-\sum_p\int_0^t\int_{\Omega(\mathbf{x})}\left(\acute{\mathrm{A}}_j^pG_{nj}\left(\mathbf{x}_r,\mathbf{x},t\right)\right)a_p\left(\mathbf{x}\right)f_i\left(\mathbf{x}_s,t'\right)\left(\grave{\mathrm{A}}_k^pG_{ki}\left(\mathbf{x},\mathbf{x}_s,t-t'\right)\right)d\mathbf{x}dt'. \tag{B.2}$$

Equation of motion in isotropic-elastic medium is given by: $\boldsymbol{\mathcal{L}}\left(\mathbf{x},t\right)\mathbf{u}\left(\mathbf{x},\mathbf{x}_s,t\right)=0$, where $\mathbf{u}=[u_x,u_y,u_z]^\dagger$ is the displacement vector and wave operator $\boldsymbol{\mathcal{L}}$ is a 3×3 matrix containing spatial and time derivatives:

$$\mathcal{L}_{ij}=\partial_i\left(\kappa+\frac{4}{3}\mu\right)\partial_j-\delta_{ij}\partial_t^2\rho+\delta_{ij}\sum_q\partial_q\mu\partial_q-2\partial_i\mu\partial_j+\partial_j\mu\partial_i, \tag{B.3}$$

where q also takes on the value of x, y and z. The isotropic-elastic scattering potential operator $\boldsymbol{\mathcal{V}}$ can be expressed in terms of wavenumbers in displacement space (Moradi and Innanen, 2015):

$$\begin{aligned}\mathcal{V}_{ij}&=\mathcal{L}_{ij}-\tilde{\mathcal{L}}_{ij}\\&=-\tilde{\kappa}\acute{\mathrm{k}}_i a_\kappa\grave{\mathrm{k}}_j-\frac{4}{3}\tilde{\mu}\acute{\mathrm{k}}_i a_\mu\grave{\mathrm{k}}_j+\tilde{\rho}\delta_{ij}\omega^2a_\rho+\tilde{\mu}\delta_{ij}\sum_q\acute{\mathrm{k}}_q a_\mu\grave{\mathrm{k}}_q+2\tilde{\mu}\acute{\mathrm{k}}_i a_\mu\grave{\mathrm{k}}_j+\tilde{\mu}\acute{\mathrm{k}}_j a_\mu\grave{\mathrm{k}}_i.\end{aligned} \tag{B.4}$$

where $\mathcal{L}_{ij}$ and $\tilde{\mathcal{L}}_{ij}$ indicate the wave operators in the perturbed and reference medium. $\grave{\mathrm{k}}$ and $\acute{\mathrm{k}}$ are wavenumbers describing incident and scattered waves respectively.

The displacement domain scattering potential operator $\boldsymbol{\mathcal{V}}$ (equation (B.4)) can be diagonalized into P-wave and S-wave components by pre-multiplying the partial derivative matrix $\boldsymbol{\Lambda}$ and post-multiplying its inverse $\boldsymbol{\Lambda}^{-1} = \nabla^{-2}\boldsymbol{\Lambda}^{\dagger}$: $\boldsymbol{\mathcal{V}}_{\mathrm{E}} = \boldsymbol{\Lambda}\boldsymbol{\mathcal{V}}\boldsymbol{\Lambda}^{-1}$, where $\boldsymbol{\Lambda}$ is a 4×3 matrix containing divergence and curl operations:

$$\boldsymbol{\Lambda} = \mathrm{i}\left[\begin{array}{cc} \acute{\mathbf{k}}_{\alpha}\cdot & \acute{\mathbf{k}}_{\beta}\times \end{array}\right]^{\dagger}, \tag{B.5}$$

where $\acute{\mathbf{k}}_{\alpha}$ and $\acute{\mathbf{k}}_{\beta}$ represent wavenumber vectors of scattered P-wave and S-wave. Inverse of the partial derivative matrix can be written as:

$$\boldsymbol{\Lambda}^{-1} = \nabla^{-2}\boldsymbol{\Lambda}^{\dagger} = -\mathrm{i}\left[\begin{array}{cc} \frac{\tilde{\alpha}^2}{\omega^2}\grave{\mathbf{k}}_{\alpha}^{\dagger}\cdot & \frac{\tilde{\beta}^2}{\omega^2}\grave{\mathbf{k}}_{\beta}^{\dagger}\times \end{array}\right], \tag{B.6}$$

where $\grave{\mathbf{k}}_{\alpha}$ and $\grave{\mathbf{k}}_{\beta}$ represent wavenumber vectors of incident P-wave and S-wave. The P-S decomposed elastic scattering potentials $\boldsymbol{\mathcal{V}}_{\mathrm{E}}$ can be subdivided into SV and SH components by pre-multiplying $\grave{\mathbf{E}}$ and post-multiplying the transpose of $\acute{\mathbf{E}}$ (Stolt and Weglein, 2012): $\tilde{\boldsymbol{\mathcal{V}}} = \acute{\mathbf{E}}\boldsymbol{\mathcal{V}}_{\boldsymbol{E}}\grave{\mathbf{E}}^{\dagger} = \acute{\mathbf{E}}\boldsymbol{\Lambda}\boldsymbol{\mathcal{V}}\boldsymbol{\Lambda}^{-1}\grave{\mathbf{E}}^{\dagger}$, where $\grave{\mathbf{E}}$ and $\acute{\mathbf{E}}$ are rotation matrices for incident and scattered waves respectively:

$$\grave{\mathbf{E}} = \begin{bmatrix} 1 & \mathbf{0} \\ 0 & \grave{\mathbf{p}}_{\mathrm{SV}} \\ 0 & -\grave{\mathbf{p}}_{\mathrm{SH}} \end{bmatrix}, \acute{\mathbf{E}} = \begin{bmatrix} 1 & \mathbf{0} \\ 0 & \acute{\mathbf{p}}_{\mathrm{SV}} \\ 0 & -\acute{\mathbf{p}}_{\mathrm{SH}} \end{bmatrix}, \tag{B.7}$$

where $\mathbf{0}$ is a zero vector, $\grave{\mathbf{p}}_{\mathrm{SV}}$ and $\acute{\mathbf{p}}_{\mathrm{SV}}$ are unit vectors describing particle motions of incident SV-wave and scattered SV-wave, $\grave{\mathbf{p}}_{\mathrm{SH}}$ and $\acute{\mathbf{p}}_{\mathrm{SH}}$ are unit vectors describing particle motions of incident SH-wave and scattered SH-wave respectively. The right term $\boldsymbol{\Lambda}^{-1}\grave{\mathbf{E}}^{\dagger}$ and left term $\acute{\mathbf{E}}\boldsymbol{\Lambda}$ of the displacement space scattering potential matrix $\boldsymbol{\mathcal{V}}$ are expressed as:

$$\boldsymbol{\Lambda}^{-1}\grave{\mathbf{E}}^{\dagger} = -\mathrm{i}\begin{bmatrix} \frac{\omega}{\tilde{\alpha}}\grave{\mathbf{p}}_{\mathrm{P}} & \frac{\omega}{\tilde{\beta}}\grave{\mathbf{p}}_{\mathrm{SH}} & \frac{\omega}{\tilde{\beta}}\grave{\mathbf{p}}_{\mathrm{SV}} \end{bmatrix}, \tag{B.8}$$

$$\acute{\mathbf{E}}\boldsymbol{\Lambda} = -\mathrm{i}\begin{bmatrix} \frac{\omega}{\tilde{\alpha}}\acute{\mathbf{p}}_{\mathrm{P}}^{\dagger} & \frac{\omega}{\tilde{\beta}}\acute{\mathbf{p}}_{\mathrm{SH}}^{\dagger} & \frac{\omega}{\tilde{\beta}}\acute{\mathbf{p}}_{\mathrm{SV}}^{\dagger} \end{bmatrix}^{\dagger}, \tag{B.9}$$

where $\grave{\mathbf{p}}_{\mathrm{P}}$ and $\acute{\mathbf{p}}_{\mathrm{P}}$ are unit vectors indicating particle motions of incident P-wave and scattered

P-wave. The transformed scattering potentials are finally obtained as:

$$\tilde{\boldsymbol{\mathcal{V}}} = \begin{bmatrix} \tilde{\mathcal{V}}_{\mathrm{PP}} & \tilde{\mathcal{V}}_{\mathrm{PSH}} & \tilde{\mathcal{V}}_{\mathrm{PSV}} \\ \tilde{\mathcal{V}}_{\mathrm{SHP}} & \tilde{\mathcal{V}}_{\mathrm{SHSH}} & \tilde{\mathcal{V}}_{\mathrm{SHSV}} \\ \tilde{\mathcal{V}}_{\mathrm{SVP}} & \tilde{\mathcal{V}}_{\mathrm{SVSH}} & \tilde{\mathcal{V}}_{\mathrm{SVSV}} \end{bmatrix} = \begin{bmatrix} \tilde{\mathcal{V}}_{\mathrm{PP}} & 0 & \tilde{\mathcal{V}}_{\mathrm{PSV}} \\ 0 & \tilde{\mathcal{V}}_{\mathrm{SHSH}} & 0 \\ \tilde{\mathcal{V}}_{\mathrm{SVP}} & 0 & \tilde{\mathcal{V}}_{\mathrm{SVSV}} \end{bmatrix}, \tag{B.10}$$

where diagonal elements of $\tilde{\boldsymbol{\mathcal{V}}}$ describe scattering that preserves wave type and off-diagonal elements describe scattering that convert wave type. Furthermore, in isotropic-elastic media, the terms associated with P-SH, SV-SH, SH-P and SH-SV scatterings are zeros. The elements of scattering potential matrix $\tilde{\boldsymbol{\mathcal{V}}}$ are obtained as:

$$\begin{aligned} \tilde{\mathcal{V}}_{\mathrm{SHSH}} &= \tilde{\rho}\omega^2 \left(a_\kappa \mathbb{R}^{\kappa}_{\mathrm{SHSH}} + a_\mu \mathbb{R}^{\mu}_{\mathrm{SHSH}} + a_\rho \mathbb{R}^{\rho}_{\mathrm{SHSH}}\right) \\ &= \tilde{\rho}\omega^2 \left(0 + a_\mu \cos\sigma + a_\rho\right), \end{aligned} \tag{B.11}$$

$$\begin{aligned} \tilde{\mathcal{V}}_{\mathrm{PP}} &= \tilde{\rho}\omega^2 \left(a_\kappa \mathbb{R}^{\kappa}_{\mathrm{PP}} + a_\mu \mathbb{R}^{\mu}_{\mathrm{PP}} + a_\rho \mathbb{R}^{\rho}_{\mathrm{PP}}\right) \\ &= \tilde{\rho}\omega^2 \left(-a_\kappa + 2a_\mu \frac{\tilde{\mu}}{\tilde{\kappa}} \sin^2\sigma - a_\rho \cos\sigma\right), \end{aligned} \tag{B.12}$$

$$\begin{aligned} \tilde{\mathcal{V}}_{\mathrm{PSV}} &= \tilde{\rho}\omega^2 \left(a_\kappa \mathbb{R}^{\kappa}_{\mathrm{PSV}} + a_\mu \mathbb{R}^{\mu}_{\mathrm{PSV}} + a_\rho \mathbb{R}^{\rho}_{\mathrm{PSV}}\right) \\ &= \tilde{\rho}\omega^2 \left(0 + 2a_\mu \frac{\tilde{\mu}}{\tilde{\kappa}} |\sin\sigma| \cos\sigma + a_\rho \sqrt{\frac{\tilde{\mu}}{\tilde{\kappa}}} |\sin\sigma|\right), \end{aligned} \tag{B.13}$$

$$\begin{aligned} \tilde{\mathcal{V}}_{\mathrm{SVSV}} &= \tilde{\rho}\omega^2 \left(a_\kappa \mathbb{R}^{\kappa}_{\mathrm{SVSV}} + a_\mu \mathbb{R}^{\mu}_{\mathrm{SVSV}} + a_\rho \mathbb{R}^{\rho}_{\mathrm{SVSV}}\right) \\ &= \tilde{\rho}\omega^2 \left(0 - a_\mu \cos 2\sigma - a_\rho \cos\sigma\right), \end{aligned} \tag{B.14}$$

$$\begin{aligned} \tilde{\mathcal{V}}_{\mathrm{SVP}} &= \tilde{\rho}\omega^2 \left(a_\kappa \mathbb{R}^{\kappa}_{\mathrm{SVP}} + a_\mu \mathbb{R}^{\mu}_{\mathrm{SVP}} + a_\rho \mathbb{R}^{\rho}_{\mathrm{SVP}}\right) \\ &= \tilde{\rho}\omega^2 \left(0 - 2a_\mu \frac{\tilde{\mu}}{\tilde{\kappa}} |\sin\sigma| \cos\sigma - a_\rho \sqrt{\frac{\tilde{\mu}}{\tilde{\kappa}}} |\sin\sigma|\right), \end{aligned} \tag{B.15}$$

where σ indicates the opening angle between incident wave and scattering wave and $\mathbb{R}$ represents the frequency-independent scattering coefficient depending on opening angle σ. Scattering coefficients of velocity parameters α, β and ρ' can be obtained by replacing the relation perturbations a_κ, a_μ and a_ρ with: $a_\kappa = a_\alpha + a_{\rho'}$, $a_\mu = a_\beta + a_{\rho'}$ and $a_\rho = a_{\rho'}$. The corresponding frequency-independent scattering coefficients for velocity-density and impedance-density parameterizations are given in Table B.1.

	Velocity-density			Impedance-density		
	α	β	ρ'	IP	IS	ρ''
$\mathbb{R}_{\text{SHSH}}$	0	$\cos\sigma$	$1+\cos\sigma$	0	$\cos\sigma$	$1-\cos\sigma$
$\mathbb{R}_{\text{PP}}$	-1	$2\gamma^2\sin^2\sigma$	$2\gamma^2\sin^2\sigma-1-\cos\sigma$	-1	$2\sin\sigma\tilde{\gamma}^2$	$1-\cos\sigma-2\sin\sigma^2\tilde{\gamma}^2$
$\mathbb{R}_{\text{PSV}}$	0	$2\gamma^2\lvert\sin\sigma\rvert\cos\sigma$	$\lvert\sin\sigma\rvert(\gamma+2\gamma^2\cos\sigma)$	0	$2\lvert\sin\sigma\rvert\cos\sigma\tilde{\gamma}^2$	$\lvert\sin\sigma\rvert(\tilde{\gamma}-2\cos\sigma\tilde{\gamma}^2)$
$\mathbb{R}_{\text{SVSV}}$	0	$-\cos 2\sigma$	$-(\cos\sigma+\cos 2\sigma)$	0	$-\cos 2\sigma$	$\cos 2\sigma-\cos\sigma$
$\mathbb{R}_{\text{SVP}}$	0	$-\gamma^2\lvert\sin\sigma\rvert\cos\sigma$	$-\lvert\sin\sigma\rvert(\gamma+2\gamma^2\cos\sigma)$	0	$-2\lvert\sin\sigma\rvert\cos\sigma\tilde{\gamma}^2$	$\lvert\sin\sigma\rvert(2\cos\sigma\tilde{\gamma}^2-\tilde{\gamma})$

Table B.1: Scattering coefficients of isotropic-elastic parameters within velocity-density and impedance-density parameterizations. $\gamma=\tilde{\beta}/\tilde{\alpha}$ and $\tilde{\gamma}$ =IS/IP.

Appendix C

Explicit expressions of interparameter contamination kernels in isotropic-elastic FWI

In this appendix, explicit expressions of correct update kernels and *interparameter contamination kernels* in equations (4.29) and (4.30) are given. $K_{\beta\leftrightarrow\beta}$ and $K_{\rho'\leftrightarrow\rho'}$ are correct update kernels for β and ρ':

$$\begin{aligned}
K_{\beta\leftrightarrow\beta} &= -a_{\beta}^{-1}\int_{\Omega(\mathbf{x}')} H_{\beta\beta}\left(\mathbf{x},\mathbf{x}'\right)\Delta m_{\beta}\left(\mathbf{x}'\right)d\mathbf{x}' \\
&= -\langle 2\rho'\beta^{2}\left[\partial_{j}G_{ni}\left(\mathbf{x}\right)\left(\partial_{i}u_{j}\left(\mathbf{x}\right)+\partial_{j}u_{i}\left(\mathbf{x}\right)\right)-2\partial_{i}G_{ni}\left(\mathbf{x}\right)\partial_{k}u_{k}\left(\mathbf{x}\right)\right] \\
&\times 2\rho'\beta^{2}\left[\partial_{j'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\beta}\left(\mathbf{x}'\right)\left(\partial_{i'}u_{j'}\left(\mathbf{x}'\right)+\partial_{j'}u_{i'}\left(\mathbf{x}'\right)\right)\right. \\
&\left.-2\partial_{i'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\beta}\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right]\rangle,
\end{aligned} \tag{C.1}$$

$$\begin{aligned}
K_{\rho'\leftrightarrow\rho'}\left(\mathbf{x}\right) &= -a_{\rho'}^{-1}\int_{\Omega(\mathbf{x}')} H_{\rho'\rho'}\left(\mathbf{x},\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)d\mathbf{x}' \\
&= -\langle\rho'\left[\left(G_{ni}\left(\mathbf{x}\right)\partial_{t}^{2}u_{i}\left(\mathbf{x}\right)+2\alpha^{2}\partial_{i}G_{ni}\left(\mathbf{x}\right)\partial_{k}u_{k}\left(\mathbf{x}\right)\right)\right. \\
&\left.+2\beta^{2}\left(\partial_{j}G_{ni}\left(\mathbf{x}\right)\left(\partial_{i}u_{j}\left(\mathbf{x}\right)+\partial_{j}u_{i}\left(\mathbf{x}\right)\right)-2\partial_{i}G_{ni}\left(\mathbf{x}\right)\partial_{k}u_{k}\left(\mathbf{x}\right)\right)\right] \\
&\times\rho'\left[\left(G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{t'}^{2}u_{i'}\left(\mathbf{x}'\right)+2\alpha^{2}\partial_{i'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right)\right. \\
&+2\beta^{2}\left(\partial_{j'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\left(\partial_{i'}u_{j'}\left(\mathbf{x}'\right)+\partial_{j'}u_{i'}\left(\mathbf{x}'\right)\right)\right. \\
&\left.\left.-2\partial_{i'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right)\right]\rangle.
\end{aligned} \tag{C.2}$$

Interparameter contamination kernels $K_{\alpha\rightarrow\beta}$ and $K_{\rho'\rightarrow\beta}$ represent the mappings from α and ρ' to β respectively:

$$\begin{aligned}
K_{\alpha\rightarrow\beta} &= -a_{\beta}^{-1}\int_{\Omega(\mathbf{x}')} H_{\beta\alpha}\left(\mathbf{x},\mathbf{x}'\right)\Delta m_{\alpha}\left(\mathbf{x}'\right)d\mathbf{x}' \\
&= -\langle 2\rho'\beta^{2}\left[\partial_{j}G_{ni}\left(\mathbf{x}\right)\left(\partial_{i}u_{j}\left(\mathbf{x}\right)+\partial_{j}u_{i}\left(\mathbf{x}\right)\right)-2\partial_{i}G_{ni}\left(\mathbf{x}\right)\partial_{k}u_{k}\left(\mathbf{x}\right)\right] \\
&\times\left[2\rho'\alpha^{2}\partial_{i'}G_{i'n'}\left(\mathbf{x}'\right)\Delta m_{\alpha}\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right]\rangle,
\end{aligned} \tag{C.3}$$

$$
\begin{aligned}
K_{\rho'\to\beta} &= -a_{\beta}^{-1}\int_{\Omega(\mathbf{x}')} H_{\beta\rho'}\left(\mathbf{x},\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)d\mathbf{x}' \\
&= -\langle 2\rho'\beta^2\left[\partial_j G_{ni}\left(\mathbf{x}\right)\left(\partial_i u_j\left(\mathbf{x}\right)+\partial_j u_i\left(\mathbf{x}\right)\right)-2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\right] \\
&\times\rho'\left[\left(G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{t'}^2 u_{i'}\left(\mathbf{x}'\right)+2\alpha^2\partial_{i'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right)\right. \\
&+2\beta^2\left(\partial_{j'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\left(\partial_{i'}u_{j'}\left(\mathbf{x}'\right)+\partial_{j'}u_{i'}\left(\mathbf{x}'\right)\right)\right. \\
&\left.\left.-2\partial_{i'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right)\right]\rangle.
\end{aligned}
\tag{C.4}
$$

Interparameter contamination kernels $K_{\alpha\to\rho'}$ and $K_{\beta\to\rho'}$ represent the mappings from α and β to ρ' respectively:

$$
\begin{aligned}
K_{\alpha\to\rho'} &= -a_{\rho'}^{-1}\int_{\Omega(\mathbf{x}')} H_{\rho'\alpha}\left(\mathbf{x},\mathbf{x}'\right)\Delta m_{\alpha}\left(\mathbf{x}'\right)d\mathbf{x}' \\
&= -\langle\rho'\left[\left(G_{ni}\left(\mathbf{x}\right)\partial_t^2 u_i\left(\mathbf{x}\right)+2\alpha^2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\right)\right. \\
&\left.+2\beta^2\left(\partial_j G_{ni}\left(\mathbf{x}\right)\left(\partial_i u_j\left(\mathbf{x}\right)+\partial_j u_i\left(\mathbf{x}\right)\right)-2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\right)\right] \\
&\times\left[2\rho'\alpha^2\partial_{i'}G_{i'n'}\left(\mathbf{x}'\right)\Delta m_{\alpha}\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right]\rangle,
\end{aligned}
\tag{C.5}
$$

$$
\begin{aligned}
K_{\beta\to\rho'} &= -a_{\rho'}^{-1}\int_{\Omega(\mathbf{x}')} H_{\rho'\beta}\left(\mathbf{x},\mathbf{x}'\right)\Delta m_{\beta}\left(\mathbf{x}'\right)d\mathbf{x}' \\
&= -\langle\rho'\left[\left(G_{ni}\left(\mathbf{x}\right)\partial_t^2 u_i\left(\mathbf{x}\right)+2\alpha^2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\right)\right. \\
&\left.+2\beta^2\left(\partial_j G_{ni}\left(\mathbf{x}\right)\left(\partial_i u_j\left(\mathbf{x}\right)+\partial_j u_i\left(\mathbf{x}\right)\right)-2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\right)\right] \\
&\times\rho'\beta^2\left[\partial_{j'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\beta}\left(\mathbf{x}'\right)\left(\partial_{i'}u_{j'}\left(\mathbf{x}'\right)+\partial_{j'}u_{i'}\left(\mathbf{x}'\right)\right)\right. \\
&\left.-2\partial_{i'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\beta}\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right]\rangle.
\end{aligned}
\tag{C.6}
$$

Appendix D

Multiparameter Hessian-vector product calculation in time domain

Constructing Hessian-vector products is an essential step for implementing Hessian-free optimization methods and quantifying uncertainties of the inverse problems (Santosa and Symes, 1988; Fichtner and Trampert, 2011a; Métivier et al., 2012; Pan et al., 2017a). One popular approach for Hessian-vector product construction in time domain is known as adjoint-state method. To understand the mechanism of the adjoint-state method for Hessian-vector calculation, I first consider minimizing the misfit function (Métivier et al., 2013):

$$\Psi\left(\mathbf{m}\right)=\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_0^T u_n^*\left(\mathbf{x}_g,\mathbf{x}_s,t'\right)\boldsymbol{\nu}dt', \tag{D.1}$$

where $\boldsymbol{\nu}$ is an arbitrary function and the gradient is given by $\nabla_{\mathbf{m}}\Psi=\nabla_{\mathbf{m}}\mathbf{u}^*\boldsymbol{\nu}$. Minimizing this misfit function subject to that wavefield $\mathbf{u}$ satisfies the wave equation gives the augmented Lagrangian functional (Liu et al., 2006; Métivier et al., 2013):

$$\begin{aligned}\boldsymbol{\chi}\left(\mathbf{m},\mathbf{u},\boldsymbol{\lambda}\right)=&\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_0^T\left[u_n^*\left(\mathbf{x}_g,\mathbf{x}_s,t'\right)\boldsymbol{\nu}\right.\\&\left.-\lambda_i\left(\rho\partial_{t'}^2u_i\left(\mathbf{x}_g,\mathbf{x}_s,t'\right)-\partial_j\left(c_{ijkl}\partial_l u_k\left(\mathbf{x}_g,\mathbf{x}_s,t'\right)\right)-f_i\right)\right]dt',\end{aligned} \tag{D.2}$$

where f_i is the source term and λ_i is the Lagrangian multiplier. Variation of functional due to the perturbations of model parameter $\Delta\mathbf{m}$ and wavefield $\Delta\mathbf{u}$ is given by (Liu et al., 2006):

$$\begin{aligned}\Delta\boldsymbol{\chi}\left(\mathbf{m},\mathbf{u},\boldsymbol{\lambda}\right)=&\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_0^T-\left[\Delta\rho\lambda_i\partial_{t'}^2u_i\left(\mathbf{x}_g,\mathbf{x}_s,t'\right)+\Delta c_{ijkl}\partial_j\lambda_i\left(\partial_l u_k\left(\mathbf{x}_g,\mathbf{x}_s,t'\right)\right)\right]\\&+\left[\boldsymbol{\nu}-\left(\rho\partial_{t'}^2\lambda_i-\partial_j\left(c_{ijkl}\partial_l\lambda_k\right)\right)\right]\Delta u_n dt'.\end{aligned} \tag{D.3}$$

Equation (D.3) is stationary with respect to wavefield perturbation $\Delta\mathbf{u}$ when its coefficient is zero, which gives the adjoint-state equation: $\rho\partial_{t'}^2\lambda_i-\partial_j\left(c_{ijkl}\partial_l\lambda_k\right)=\boldsymbol{\nu}$, where $\boldsymbol{\nu}$ serves as the adjoint source. Thus, gradients of the misfit function with respect to density ρ and elastic constants $\mathbf{c}$ become:

$$\nabla_\rho\boldsymbol{\chi}=\nabla_\rho\mathbf{u}^*\boldsymbol{\nu}=-\langle\tilde{u}_i\partial_{t'}^2u_i\rangle,\nabla_{\mathbf{c}}\boldsymbol{\chi}=\nabla_{\mathbf{c}}\mathbf{u}^*\boldsymbol{\nu}=-\langle\partial_j\tilde{u}_i\partial_l u_k\rangle, \tag{D.4}$$

where $\tilde{u}_i(\mathbf{x}_g, \mathbf{x}, T-t') = \lambda_i(\mathbf{x}_g, \mathbf{x}, T-t') = G_{ni}(\mathbf{x}_g, \mathbf{x}, T-t')\boldsymbol{\nu}$ is the adjoint wavefield. Product of Jacobian matrix with an arbitrary vector $\mathbf{v}$ is given by:

$$\mathfrak{J} = \nabla_{\mathbf{m}}\mathbf{u}\mathbf{v} = -\langle G_{n'i'}(\mathbf{x}')\, v(\mathbf{x}')\, \partial_{t''}^2 u_{i'}(\mathbf{x}') + \partial_{j'} G_{n'i'}(\mathbf{x}')\, v(\mathbf{x}')\, \partial_{l'} u_{k'}(\mathbf{x}')\rangle, \tag{D.5}$$

Replacing $\boldsymbol{\nu}$ in equation (D.4) with Jacobian-vector product $\mathfrak{J}$ gives the multiparameter Gauss-Newton Hessian-vector product:

$$\begin{aligned}\mathfrak{H} &= \nabla_{\mathbf{m}}\mathbf{u}^*\nabla_{\mathbf{m}}\mathbf{u}\mathbf{v} \\ &= \langle \big(G_{ni}(\mathbf{x})\, \partial_{t'}^2 u_i(\mathbf{x}) + \partial_j G_{ni}(\mathbf{x})\, \partial_l u_k(\mathbf{x})\big) \\ &\times \big(G_{n'i'}(\mathbf{x}')\, v(\mathbf{x}')\, \partial_{t''}^2 u_{i'}(\mathbf{x}') + \partial_{j'} G_{n'i'}(\mathbf{x}')\, v(\mathbf{x}')\, \partial_{l'} u_{k'}(\mathbf{x}')\big)\rangle.\end{aligned} \tag{D.6}$$

For example, product of off-diagonal block $\mathbf{H}_{\mathbf{c}\rho}$ in multiparameter Gauss-Newton Hessian with an arbitrary perturbation vector $\mathbf{v}_\rho$ due to density can be written explicitly as:

$$\begin{aligned}\mathfrak{H}_{\mathbf{c}\rho}(\mathbf{x}) &= \int_{\Omega(\mathbf{x}')} H_{\mathbf{c}\rho}(\mathbf{x}, \mathbf{x}')\, v_\rho(\mathbf{x}')\, d\mathbf{x}' \\ &= \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g} \int_{\Omega(\mathbf{x}')} \int\int \partial_k u_l(\mathbf{x}, \mathbf{x}_s, t')\, \partial_j G_{ni}(\mathbf{x}_g, \mathbf{x}, T-t') \\ &\times G_{n'i'}(\mathbf{x}_g, \mathbf{x}', t-t'')\, v_\rho(\mathbf{x}')\, \partial_{t''}^2 u_{i'}(\mathbf{x}', \mathbf{x}_s, t'')\, dt' dt'' d\mathbf{x}',\end{aligned} \tag{D.7}$$

where Jacobian-vector product $\mathfrak{J}_\rho$ is:

$$\mathfrak{J}_\rho(\mathbf{x}_g, \mathbf{x}', t) = \sum_{\mathbf{x}_s} \int G_{n'i'}(\mathbf{x}_g, \mathbf{x}', t-t'')\, v_\rho(\mathbf{x}')\, \partial_{t''}^2 u_{i'}(\mathbf{x}', \mathbf{x}_s, t'')\, dt'', \tag{D.8}$$

where interaction of the indicate wavefield $u_{i'}(\mathbf{x}', \mathbf{x}_s, t'')$ with the perturbation vector $v_\rho(\mathbf{x}')$ serves as the "secondary scattered source" $f_i'(\mathbf{x}', \mathbf{x}_s, t'')$:

$$f_i'(\mathbf{x}', \mathbf{x}_s, t'') = v_\rho(\mathbf{x}')\, \partial_{t''}^2 u_{i'}(\mathbf{x}', \mathbf{x}_s, t'')\,. \tag{D.9}$$

Convolution of this scattered source with the Green's function $G_{n'i'}(\mathbf{x}_g, \mathbf{x}', t-t'')$ gives the first-order scattered wavefield:

$$\Delta u_{n'}(\mathbf{x}_g, \mathbf{x}', t) = \sum_{\mathbf{x}_s} \int G_{n'i'}(\mathbf{x}_g, \mathbf{x}', t-t'')\, f_i'(\mathbf{x}', \mathbf{x}_s, t'')\, dt''. \tag{D.10}$$

Recorded scattered wavefield at the receiver locations can be considered as adjoint source $\tilde{f}_n'(\mathbf{x}, \mathbf{x}', t)$:

$$\tilde{f}_n'(\mathbf{x}, \mathbf{x}', t) = \sum_{\mathbf{x}_g} \Delta u_n(\mathbf{x}_g, \mathbf{x}', T-t)\, \delta(\mathbf{x} - \mathbf{x}_g)\,. \tag{D.11}$$

Inserting equation (D.11) into equation (D.7) gives the Hessian-vector product as:

$$\mathfrak{H}_{\mathbf{c}\rho}\left(\mathbf{x}\right)=\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_{\Omega(\mathbf{x}')}\int\partial_k u_l\left(\mathbf{x},\mathbf{x}_s,t'\right)\partial_j\tilde{u}_i\left(\mathbf{x}_g,\mathbf{x},\mathbf{x}',T-t'\right)dt'd\mathbf{x}', \tag{D.12}$$

where $\tilde{u}_i\left(\mathbf{x}_g,\mathbf{x},\mathbf{x}',T-t'\right)$ is the adjoint wavefield:

$$\tilde{u}_i\left(\mathbf{x}_g,\mathbf{x},\mathbf{x}',T-t'\right)=\int_0^{T-t'}G_{ni}\left(\mathbf{x}_g,\mathbf{x},T-t'-t'''\right)\tilde{f}'_n\left(\mathbf{x},\mathbf{x}',t'''\right)dt'''. \tag{D.13}$$

Calculating Gauss-Newton Hessian-vector product with the adjoint-state approach needs to construct forward wavefield $u_{i'}\left(\mathbf{x}',\mathbf{x}_s,t''\right)$, Born modelling wavefield $\Delta u_{n'}\left(\mathbf{x}_g,\mathbf{x}',t\right)$ and adjoint wavefield $\tilde{u}_i\left(\mathbf{x}_g,\mathbf{x},\mathbf{x}',T-t'\right)$, The computational cost is 1.5 times than that of calculating gradient (Métivier et al., 2013).

Another approach for Hessian-vector calculation is finite-difference method. Recalling that Hessian operator represents the Fréchet derivative of the gradient vector, with Taylor series expansion:

$$\nabla_{\mathbf{m}}\Phi\left(\mathbf{m}^0+\Delta\mathbf{m}\right)\approx\nabla_{\mathbf{m}}\Phi\left(\mathbf{m}^0\right)+\mathbf{H}\Delta\mathbf{m}, \tag{D.14}$$

where $\mathbf{m}^0$ denote current model. Replacing the model perturbation vector $\Delta\mathbf{m}$ with an arbitrary vector $\mathbf{v}$ scaled by a small constant value $\bar{\epsilon}$ gives:

$$\nabla_{\mathbf{m}}\Phi\left(\mathbf{m}^0+\bar{\epsilon}\mathbf{v}\right)\approx\nabla_{\mathbf{m}}\Phi\left(\mathbf{m}^0\right)+\bar{\epsilon}\mathbf{H}\mathbf{v}. \tag{D.15}$$

An approximate Hessian-vector product solution can be obtained by:

$$\mathfrak{H}\approx\frac{\nabla_{\mathbf{m}}\Phi\left(\mathbf{m}^0+\bar{\epsilon}\mathbf{v}\right)-\nabla_{\mathbf{m}^0}\Phi\left(\mathbf{m}\right)}{\bar{\epsilon}}. \tag{D.16}$$

Two additional pairs of forward and adjoint simulations are required for calculating this Hessian-vector product approximation, which is affordable for large-scale inverse problems. Even this approximation may suffer from rounding errors, the accuracy can be improved with high-order finite-difference approaches at the cost of more computation requirements and for very small coefficient $\bar{\epsilon}$, its accuracy will be very high. For example, if the multiparameter Hessian is applied to vector $\mathbf{v}=[\mathbf{v}_{\mathbf{c}}=0\ \ \mathbf{v}_\rho\neq 0]^\dagger$, the Hessian-vector product $\mathfrak{H}_{\mathbf{c}\rho}=\mathbf{H}_{\mathbf{c}\rho}\mathbf{v}_\rho$ can be obtained by:

$$\mathfrak{H}_{\mathbf{c}\rho}\approx\frac{\nabla_{\mathbf{c}}\Phi\left(\rho_0+\bar{\epsilon}\mathbf{v}_\rho\right)-\nabla_{\mathbf{c}}\Phi\left(\rho_0\right)}{\bar{\epsilon}}. \tag{D.17}$$

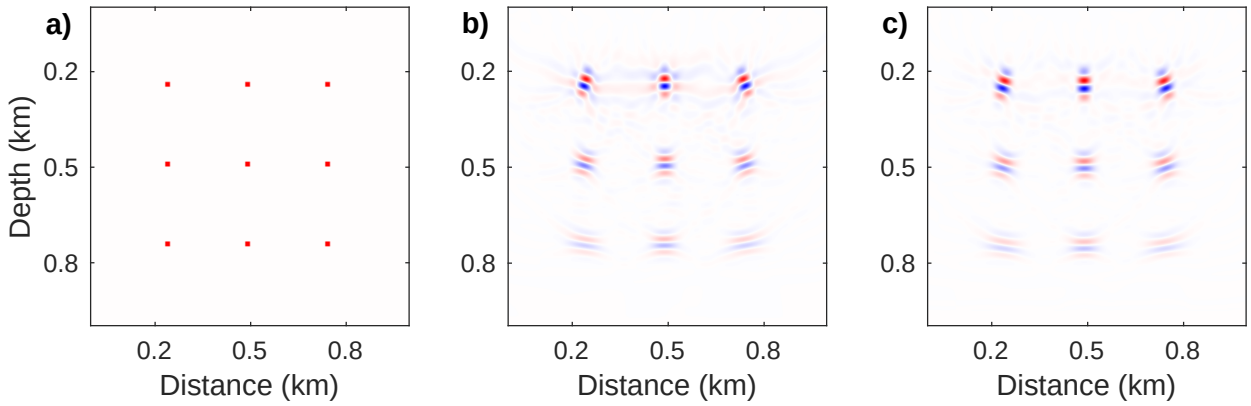

Figure D.1: Panel (a) shows vector $\mathbf{v}_\rho$ with 9 isolated spikes; Figures (b) and (c) are the multiparameter Hessian-vector products $\mathfrak{H}_{\kappa\rho} = \mathbf{H}_{\kappa\rho}\mathbf{v}_\rho$ calculated with second-order adjoint-state and finite-difference methods respectively.

Next, the two approaches with a multiparameter acoustic example are examined for comparison. Figure D.1a shows the vector $\mathbf{v}_\rho$ with 9 isolated spikes. A homogeneous model with bulk modulus $\kappa = 13.5$ GPa and density $\rho = 1500$ kg/m^3 is used as the background model. A set of sources and receivers are distributed regularly along the top boundary of the model. Multiparameter Hessian-vector product $\mathfrak{H}_{\kappa\rho} = \mathbf{H}_{\kappa\rho}\mathbf{v}_\rho$ is calculated with adjoint-state and finite-difference methods, as shown in Figures D.1b and D.1c respectively. The Hessian-vector products by these two methods match very well. In this chapter, I adopt the adjoint-state approach for calculating the multiparameter Hessian-vector products in isotropic-elastic media.

Appendix E

Stochastically estimating diagonals of multiparameter Hessian off-diagonal blocks

This Appendix explains how to efficiently estimate the diagonals of multiparameter Hessian off-diagonal blocks with stochastic probing approach. Zero-mean random vector $\mathbf{v}$ is divided into N_p independent subvectors corresponding to N_p different physical parameters: $\mathbf{v} = \left[\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{N_p}\right]$. $\mathbf{v}_p \sim \mathcal{N}\left(\mathbb{E}\left[\mathbf{v}_p\right], \mathbf{\Sigma}_{\mathbf{v}_p\mathbf{v}_p}\right)$ and $\mathbf{v}_q \sim \mathcal{N}\left(\mathbb{E}\left[\mathbf{v}_q\right], \mathbf{\Sigma}_{\mathbf{v}_q\mathbf{v}_q}\right)$ are independent subvectors within $\mathbf{v}$. Expectation values and variance-covariance matrices of the subvectors satisfy:

$$\mathbb{E}\left[v_p\left(\mathbf{x}\right)\right] = 0, \mathbb{E}\left[v_q\left(\mathbf{x}\right)\right] = 0, \tag{E.1}$$

$$\Sigma_{\mathbf{v}_p\mathbf{v}_p}\left(v_p\left(\mathbf{x}\right), v_p\left(\mathbf{x}'\right)\right) = \mathbb{E}\left[v_p\left(\mathbf{x}\right) v_p\left(\mathbf{x}'\right)\right] = \delta\left(\mathbf{x} - \mathbf{x}'\right), \tag{E.2}$$

$$\Sigma_{\mathbf{v}_q\mathbf{v}_q}\left(v_q\left(\mathbf{x}\right), v_q\left(\mathbf{x}'\right)\right) = \mathbb{E}\left[v_q\left(\mathbf{x}\right) v_q\left(\mathbf{x}'\right)\right] = \delta\left(\mathbf{x} - \mathbf{x}'\right). \tag{E.3}$$

Cross-covariance between $\mathbf{v}_p$ and $\mathbf{v}_p$ can be obtained as:

$$\begin{aligned}
\Sigma_{\mathbf{v}_p\mathbf{v}_q}\left(v_p\left(\mathbf{x}\right), v_q\left(\mathbf{x}'\right)\right) &= \mathbb{E}\left[\left(v_p\left(\mathbf{x}\right) - \mathbb{E}\left[v_p\left(\mathbf{x}\right)\right]\right)\left(v_q\left(\mathbf{x}'\right) - \mathbb{E}\left[v_q\left(\mathbf{x}'\right)\right]\right)^{\dagger}\right] \\
&= \mathbb{E}\left[v_p\left(\mathbf{x}\right) v_q\left(\mathbf{x}'\right)\right] - \mathbb{E}\left[v_p\left(\mathbf{x}\right)\right]\left(\mathbb{E}\left[v_q\left(\mathbf{x}'\right)\right]\right)^{\dagger} \\
&= 0.
\end{aligned} \tag{E.4}$$

Thus, expectation value of the correlation result between subvector $\mathbf{v}_p$ and the sub-Hessian-vector product $\mathfrak{H}_p$ is given by:

$$\begin{aligned}
\mathbb{E}\left[\mathbf{v}_p \odot \mathfrak{H}_p\right] &= \sum_{q=1}^{N_p} \mathbf{v}_p \odot \mathbf{H}_{pq}\mathbf{v}_q \\
&= \int_{\Omega(\mathbf{x}')} H_{pq}\left(\mathbf{x}, \mathbf{x}'\right) \mathbb{E}\left[v_p\left(\mathbf{x}\right) v_p\left(\mathbf{x}'\right)\right] d\mathbf{x}' \\
&= \mathbf{H}_{pq}\left(\mathbf{x}, \mathbf{x}\right).
\end{aligned} \tag{E.5}$$

Chapter 5

Gauss-Newton and full-Newton methods for estimating elastic constants in 2D HTI media

5.1 Summary

This chapter focuses on the problem of estimating the elastic constants of a fractured medium using multiparameter FWI and modeling naturally fractured reservoirs as equivalent anisotropic media. Multiparameter FWI, although promising, remains exposed to the challenge of parameter crosstalk problem, which is strongly influenced by the form of the scattering pattern for each parameter. I have derived 3D radiation patterns associated with scattering from a range of elastic constants in general anisotropic media. Then, scattering patterns specific to a horizontal transverse isotropic (HTI) medium are derived to draw conclusions about parameter crosstalk in FWI. The role of the multiparameter Gauss-Newton (GN) Hessian in suppressing parameter crosstalk is revealed. I found that the second-order term in the multiparameter Hessian can be constructed with the adjoint-state technique. The analytic scattering patterns for HTI media are examined with a 2D numerical example. The roles played by the first- and second-order terms in multiparameter Hessian to suppress parameter crosstalk and second-order scattering artifacts are also verified numerically. The multiparameter GN and FN methods are applied for determining the elastic constants in HTI media with a two-block-layer model.

5.2 Introduction

Naturally fractured reservoirs play an important role in current hydrocarbon production (Nelson, 1985). Fracture properties are amongst the most valuable data for reservoir characterization. The influence of fractures/cracks in a geological medium on the seismic response can be modelled via an equivalent anisotropic solid and the associated elastic stiffness coefficients (Hudson, 1981; Schoenberg, 1983). For transverse isotropy with a horizontal symmetry

axis (HTI media), the simplest azimuthal anisotropic model for describing vertical cracks, reflection seismic signatures can be described by five independent elastic constants (Rüger, 1997; Tsvankin, 1997b,a). In reflection seismology, most current methods for estimating fracture properties focus on amplitude and traveltime methods (Thomsen, 1988; Tsvankin, 1997b). This chapter considers the problem of estimating elastic constants in anisotropic media using multiparameter full-waveform inversion (FWI).

Inverting multiple parameters using multiparameter FWI has also, however, received increased attention in recent years, though it is a very challenging task. As discussed in chapter 4, involving several parameters increases the nonlinearity of the inversion process and furthermore introduces the interparameter tradeoff issue (or parameter crosstalk), the conflation of the influence of one physical property on the data with another (Operto et al., 2013; Prieux et al., 2013a; Innanen, 2014a; Oh et al., 2015; Métivier et al., 2015). Parameter crosstalk artifacts (or interparameter contaminations) are strongly present in steepest-descent method, wherein update in each parameter proceed with no accounting for the multiparameter character of the problem. However, even when the multiparameter character of a system is properly included in a FWI update, crosstalk persists if the wavefield variation caused by one parameter is similar to that caused by another physical parameter. Hence, the Fréchet derivative wavefields associated with different physical parameters are crucial to understand the parameter crosstalk problem in multiparameter FWI. Scattering patterns represent the analytic solutions of Fréchet derivative wavefields. In chapter 4, I used the scattering patterns to analyze the coupling effects of isotropic-elastic parameters within various parameterizations for isotropic-elastic FWI.

Multiparameter Hessian in FWI is a square and symmetric matrix with a block structure. It carries more information than a single-parameter Hessian. Within the approximate Hessian associated with a multiparameter Gauss-Newton update, off-diagonal blocks measure correlation of Fréchet derivative wavefields with respect to different physical parameters, and they act to mitigate the coupling effects between these parameters (Operto et al., 2013). Innanen (2014a) showed that the diagonal elements internal to the off-diagonal blocks suppress crosstalk, in precritical reflection FWI, in a manner consistent with AVO inversion and linearized inverse scattering. Detailed introduction of multiparameter Hessian has been

given in chapter 4 of this thesis.

The gradient vector is also known to be contaminated by second-order scattered energy in the data residuals. Pratt et al. (1998b) discussed and analyzed the second-order term in the single-parameter Hessian, which accounts for the second-order scattering effects. In previous chapters, this term is ignored for sake of compactness. However, this term becomes important when the data residuals or the second-order scattered energy are very strong (Margrave et al., 2011b; Métivier et al., 2014). Incorporating this second-order term can eliminate the second-order scattering effects in the gradient vector (Pratt et al., 1998b). This term in a multiparameter Hessian becomes more complex and the second-order partial derivative wavefields can be caused by the perturbations of different physical parameters. It predicts the change in the gradient due to the multiparameter second-order non-linear effects. In this chapter, this second-order term in multiparameter Hessian is calculated with adjoint-state method following Pratt et al. (1998b). Involving both of the first-order and second-order terms in the multiparameter Hessian for preconditioning the gradient is known as the full-Newton method.

In this chapter, analytic expressions for the 3D scattering patterns of the elastic constants in general anisotropic media are derived. Then, the scattering patterns of elastic constants in a specific HTI medium are illustrated for interparameter tradeoffs (or parameter crosstalk) analysis. Furthermore, the ability of the multiparameter approximate Hessian to suppress parameter crosstalk is examined for HTI elastic constants inversion. Pratt et al. (1998b) calculated the second-order term in single-parameter Hessian using an adjoint-state method. This chapter shows that the second-order term in multiparameter Hessian associated with multiparameter second-order scattering effects can also be constructed with the adjoint-state technique.

This chapter is organized as follows. First, the basic theories for the forward modeling problem in anisotropic media are reviewed. Then, the parameter crosstalk difficulty in multiparameter FWI is discussed. The analytic expressions of 3D scattering patterns for the elastic constants in general anisotropic media are derived. The physical interpretations of the first- and second-order terms in multiparameter Hessian are explained. Their roles in suppressing parameter crosstalk and second-order scattering effects are revealed. How to

construct the second-order term using the adjoint-state method is described. In the numerical modeling section, the analytic and numerical scattering patterns of the elastic constants for parameter crosstalk analysis are examined. Several numerical examples are presented to illustrate the role of multiparameter Hessian in mitigating parameter crosstalk and second-order scattering artifacts. The GN and FN multiparameter FWI are finally applied on a two-block-layer model for comparison.

5.3 Theory and Methods

This section first presents the basic principles of forward modelling in anisotropic media. Introduction for Newton-based optimization methods (Gauss-Newton and full-Newton methods) can be found in chapter 3. The issue of parameter crosstalk for inverting the elastic constants in HTI media using multiparameter FWI is discussed. The roles of the multiparameter Hessian in suppressing parameter crosstalk and second-order scattering effects are revealed.

5.3.1 Forward modelling problem in anisotropic media

In reflection seismology, the wavelengths of seismic waves are typically much larger than the fracture size. When considering that the fractures are closely spaced and parallel, the finite fracture spacings and their detailed spatial distributions can be neglected and the fractured medium can be replaced by effective anisotropic solids. The reflection seismic signatures are associated with the elastic stiffness coefficients c_{IJ} through the equation of motion in general anisotropic media (Hudson, 1981; Schoenberg, 1983), which can be expressed as:

$$\frac{\partial \sigma_{ij}}{\partial x_j} + f_i = \rho \frac{\partial^2 u_i}{\partial t^2}, \tag{5.1}$$

where $u_i(\mathbf{x}, t)$ indicates i component of the particle displacement at Cartesian coordinate position $\mathbf{x} = (x, y, z)$ and time t, $f_i(\mathbf{x}_s)$ is the source term at position $\mathbf{x}_s$, ρ is the density and σ_{ij} denotes the stress tensor, which can be defined using Hooke's law:

$$\sigma_{ij} = c_{ijkl} e_{kl}, \tag{5.2}$$

where c_{ijkl} indicates the elastic modulus tensor, $e_{kl} = \frac{1}{2}\left(\frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k}\right)$ is the strain tensor. Thus, 81 elastic constants are required to characterize the elasticity of the medium. Because the symmetry of the stress and strain tensors, only 21 elastic stiffness coefficients are independent and the $3 \times 3 \times 3 \times 3$ tensor c_{ijkl} can be represented more compactly using 6×6 symmetric matrix c_{IJ} following the Voigt recipe for indexes, where I and J range from 1 to 6 (Crampin, 1984; Tsvankin and Grechka, 2011).

Models containing parallel vertical fractures are equivalent to HTI media (transverse isotropy with horizontal symmetry axis), which can be characterized by five independent elastic constants c_{33}, c_{55}, c_{11}, c_{13}, c_{44}. I extract the x-z plane with zero azimuth angle from 3D geometry, which forms the simplified 2D HTI model described by four elastic constants (c_{33}, c_{55}, c_{11} and c_{13}). Numerical solutions of the wavefields are calculated using an explicit finite-difference method with fourth-order accuracy in space and second-order accuracy in time (Virieux, 1986; Levander, 1988).

5.3.2 Parameter crosstalk (or interparameter tradeoff) problem and the role of scattering pattern

To be separately constrained by seismic observations, the perturbation of each medium parameter type cause a unique variation in its Fréchet derivative wavefields over the observed range of scattering and azimuthal angles. Identical or nearly identical variations are one of the key mechanisms of parameter crosstalk in multiparameter FWI (Operto et al., 2013). The interaction of the incident wavefield with the model perturbation serves as the "virtual source" or "secondary scattered source". The scattering pattern of the "virtual source" governs the amplitude variation of Fréchet derivative wavefield as a function of scattering angle and azimuthal angle. An inversion sensitivity analysis taking these patterns into account is important (Gholami et al., 2013a). Coupling effects between different elastic constants to be examined, and proper parameterization and optimal acquisition geometry should be calculated (Tarantola, 1986; Gholami et al., 2013b). Forgues and Lambaré (1997) studied different parameter classes in acoustic and elastic ray+Born inversion. In this chapter, the 3D scattering patterns for elastic constants in general anisotropic media are derived in Appendix F.

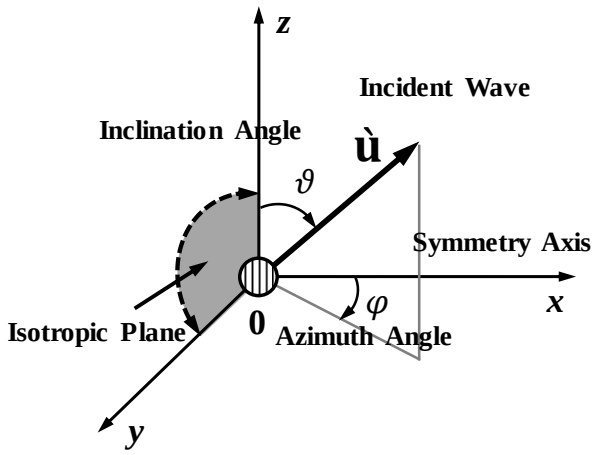

Figure 5.1: The 3D geometry. $\mathbf{\grave{u}}$ indicates normal to the wavefront of the incident wave. ϑ and φ are the inclination angle departing from z axis and azimuth angle departing from x axis for describing the incident wave. The HTI inclusion is at the original point 0. x axis parallel to the symmetry axis of the HTI inclusion and the y-z plane is perpendicular to the axis of symmetry.

The 3D geometry for describing the scattering problem due to the local anisotropic inclusion is presented in Figure 5.1. The inclination angle ϑ and azimuth angle φ are defined for describing the incident wave and inclination angle θ and azimuth angle ϕ for describing the scattered wave. For an HTI inclusion, I also define the symmetry axis to parallel to the x axis and the isotropic plane to be consistent with the y-z plane. Neglecting the contribution from density and considering the perturbation of the elastic constant matrix $\Delta\check{\mathbf{c}}$ for HTI media, the equivalent moment tensor source (equation (F.11)) caused by the perturbations of the elastic constants can be obtained as (Ben-Menahem and Singh, 1981; Chapman, 2004):

$$\Delta\check{\mathbf{M}} = \begin{bmatrix} \Delta c_{11}\tilde{e}_{11} + \Delta c_{13}\left(\tilde{e}_{22} + \tilde{e}_{33}\right) & 2\Delta c_{55}\tilde{e}_{12} & 2\Delta c_{55}\tilde{e}_{13} \\ 2\Delta c_{55}\tilde{e}_{12} & \Delta c_{13}\tilde{e}_{11} + \Delta c_{33}\tilde{e}_{22} + \Delta\nu\tilde{e}_{33} & 2\Delta c_{44}\tilde{e}_{23} \\ 2\Delta c_{55}\tilde{e}_{13} & 2\Delta c_{44}\tilde{e}_{23} & \Delta c_{13}\tilde{e}_{11} + \Delta\nu\tilde{e}_{22} + \Delta c_{33}\tilde{e}_{33} \end{bmatrix}. \tag{5.3}$$

where the symbol "$\check{}$" means specification for HTI media, and $\tilde{e}_{ij}$ indicate the strain components of the incident wave. Considering an incident plane P-wave (equation (F.12)), inserting the reduced equivalent moment tensor source $\Delta\check{\mathbb{M}}$ (see Appendix F) for HTI media

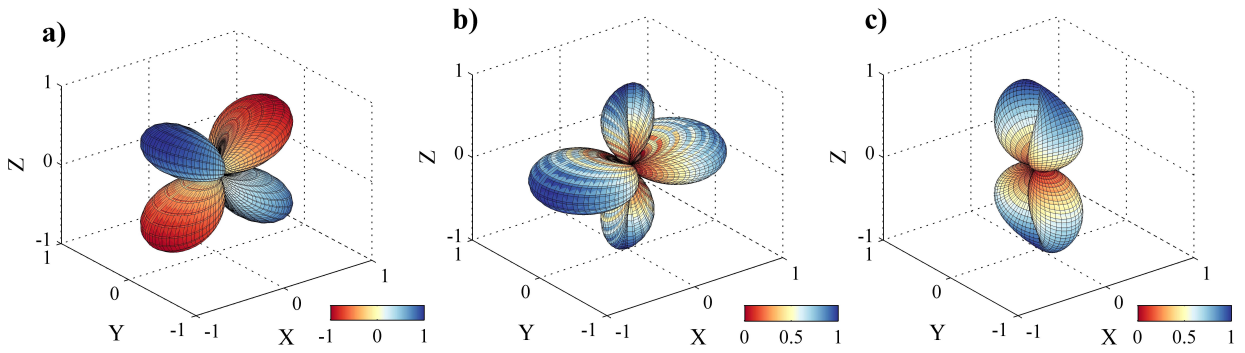

Figure 5.2: (a), (b) and (c) show the P-P (equation (5.5)), P-SV (equation (5.6)) and P-SH (equation (5.7)) scattering patterns due to the perturbation of c_{55}.

into equation (F.18) yields the 3D scattering patterns of the elastic constants:

$$\check{\mathbb{R}}_{\mathrm{P}}\left(\vartheta,\varphi,\theta,\phi\right)=\hat{\mathbf{g}}^{\dagger}\nabla_{\mathbf{m}}\check{\mathbb{M}}\hat{\mathbf{r}}, \tag{5.4}$$

where vectors $\hat{\mathbf{g}}$ and $\hat{\mathbf{r}}$ are defined in equation (F.17). Applying perturbation to one elastic constant and keeping the perturbations of other elastic constants as zero, the scattering pattern for the specified elastic constant can be obtained. For example, considering the perturbation of elastic constant Δc_{55}, the P-P, P-SV and P-SH scattering coefficients due to Δc_{55} can be obtained as:

$$\begin{aligned}\check{\mathbb{R}}_{\text{P-P}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{55}\right)&=4\hat{p}_1\hat{q}_2\hat{r}_1\hat{r}_2+4\hat{p}_1\hat{q}_3\hat{r}_1\hat{r}_3\\&=\sin^2\vartheta\sin 2\varphi\sin^2\theta\sin 2\phi+\sin 2\vartheta\cos\varphi\sin 2\theta\cos\phi,\end{aligned} \tag{5.5}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{P-SV}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{55}\right)&=2\hat{p}_1\hat{q}_2\hat{\theta}_2\hat{r}_1+2\hat{p}_1\hat{q}_3\hat{\theta}_3\hat{r}_1+2\hat{p}_1\hat{q}_2\hat{\theta}_1\hat{r}_2+2\hat{p}_1\hat{q}_3\hat{\theta}_1\hat{r}_3\\&=\sin^2\vartheta\sin 2\varphi\sin\theta\cos\theta\sin 2\phi+\sin 2\vartheta\cos\varphi\cos 2\theta\cos\phi,\end{aligned} \tag{5.6}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{P-SH}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{55}\right)&=2\hat{p}_1\hat{q}_2\hat{\phi}_2\hat{r}_1+2\hat{p}_1\hat{q}_3\hat{\phi}_3\hat{r}_1+2\hat{p}_1\hat{q}_2\hat{\phi}_1\hat{r}_2+2\hat{p}_1\hat{q}_3\hat{\phi}_1\hat{r}_3\\&=\sin^2\vartheta\sin 2\varphi\sin\theta\cos 2\phi-\sin 2\vartheta\cos\varphi\sin\phi\cos\theta.\end{aligned} \tag{5.7}$$

Figures 5.2a, 5.2b and 5.2c show the 3D P-P (equation (5.5)), P-SV (equation (5.6)) and P-SH (equation (5.7)) scattering patterns due to Δc_{55} ($\vartheta=135^{\circ}$ and $\varphi=0^{\circ}$). Figures 5.3a, 5.3b and 5.3c show the P-P scattering patterns due to perturbations of c_{33}, c_{11} and c_{13} respectively. The scattered P-wave, SV-wave and SH-wave with incident plane SV-wave and incident plane SH-wave can be obtained following the equations (F.23) and (F.24).

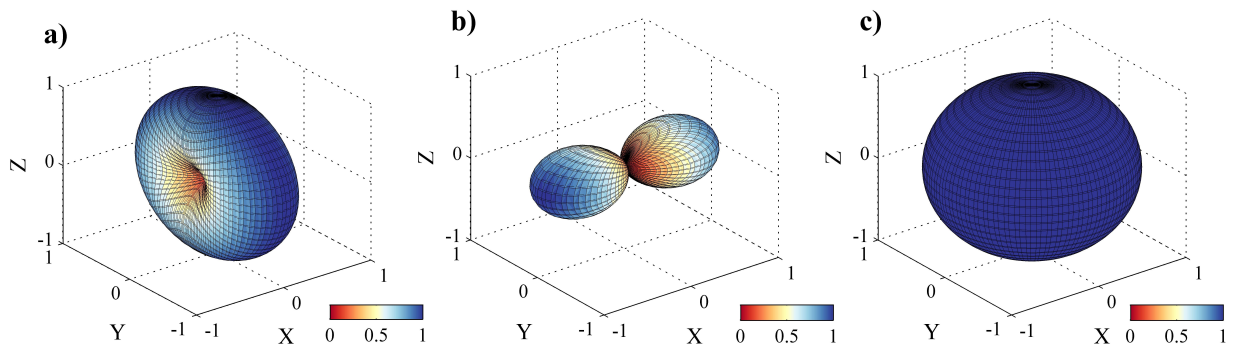

Figure 5.3: (a), (b) and (d) are the P-P scattering patterns due to the perturbations of c_{33}, c_{13} and c_{11} respectively.

The explicit expressions of the 3D scattering patterns for the elastic stiffness coefficients in HTI media with incident P-wave, S-wave, and SH-wave are given in Appendix F. If the scattering patterns associated with different elastic constants significantly overlap over a range of scattering angle or azimuth angle, the parameter crosstalk between these physical parameters will contaminate the update. If the update is not properly preconditioned, the inversion process will be impacted negatively. This difficulty also raises the parameterization issue for managing parameter crosstalk. In chapter 4, the parameterization issue in isotropic-elastic FWI has been discussed. Determining a more proper parameterization for inverting fracture properties using FWI is beyond the scope of this chapter. The 3D scattering patterns of the elastic constants in general anisotropic media given in this chapter can be transformed to the scattering patterns for any parameter class using the chain rule.

5.3.3 Gauss-Newton and full-Newton multiparameter full-waveform inversion

Chapter 3 has introduced the basic principles of Newton-based optimization methods (Gauss-Newton and full-Newton methods) for FWI, which are able to provide quadratic convergence rate but are limited due to extensive computation requirements. In this chapter, the Newton-based optimization methods are applied on a small scale inverse problem for inverting elastic stiffness coefficients in 2D HTI media, which helps reveal the roles of multiparameter Hessian in suppressing parameter crosstalk artifacts and multiparameter second-order scattering effects. The role of the single-parameter Hessian has been discussed and analyzed in chapter 2. The structure of multiparameter Hessian in multiparameter FWI has also been introduced

in chapter 4. For inverting the four elastic constants in 2D HTI media, the multiparameter Hessian $\mathbf{H}$ has 16 block sub-matrices ($N_p = 4$):

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{3333} & \mathbf{H}_{3355} & \mathbf{H}_{3311} & \mathbf{H}_{3313} \\ \mathbf{H}_{5533} & \mathbf{H}_{5555} & \mathbf{H}_{5511} & \mathbf{H}_{5513} \\ \mathbf{H}_{1133} & \mathbf{H}_{1155} & \mathbf{H}_{1111} & \mathbf{H}_{1113} \\ \mathbf{H}_{1333} & \mathbf{H}_{1355} & \mathbf{H}_{1311} & \mathbf{H}_{1313} \end{bmatrix}, \tag{5.8}$$

where the subscripts of the block matrices in $\mathbf{H}$ are consistent with the subscripts of two elastic constants. The multiparameter Hessian $\mathbf{H}$ can be written as the summation of the first-order term $\mathbf{H}_a$ and second-order term $\bar{\mathbf{H}}$. The first-order term $\mathbf{H}_a$ is also known as multiparameter approximate Hessian used in Gauss-Newton method. The elements in $\mathbf{H}_a$ measure the correlations of two Fréchet derivative wavefields. For example, the element $H_{a,\mathbf{m}_1\mathbf{m}_2}(\mathbf{x}, \mathbf{x}')$ can be expressed as:

$$H_{a,\mathbf{m}_1\mathbf{m}_2}(\mathbf{x}, \mathbf{x}') = \sum_{\mathbf{x}_s} \sum_{\mathbf{x}_g} \sum_{\omega} \Re \left(\nabla_{\mathbf{m}_1(\mathbf{x})} \mathbf{u}^{\dagger}(\mathbf{x}_g, \mathbf{x}_s, \omega) \nabla_{\mathbf{m}_2(\mathbf{x}')} \mathbf{u}^{*}(\mathbf{x}_g, \mathbf{x}_s, \omega) \right), \tag{5.9}$$

where when $\mathbf{m}_1 = \mathbf{m}_2$, it indicates the element in diagonal block, and when $\mathbf{m}_1 \neq \mathbf{m}_2$, it indicates the element in off-diagonal block. Note that in this chapter the receiver sampling operator $\mathcal{P}$ is ignored for sake of compactness. The multiparameter approximate Hessian $\mathbf{H}_a$ is essential in overcoming the crosstalk difficulty in multiparameter FWI (Operto et al., 2013). As discussed in chapter 4, the similarity of the Fréchet derivative wavefields with respect to different physical parameters gives rise to the crosstalk problem. The off-diagonal blocks in multiparameter approximate Hessian, as indicated by the grey boxes in Figure 5.4, predict the coupling effects and applying its inverse to the gradient can remove the parameter crosstalk artifacts (or interparameter contaminations).

The space-type multiparameter Hessian approximation $\mathbf{H}_{a,\mathbf{s}}$ given by Innanen (2014a) neglects the contributions of the off-diagonal blocks and stresses the correlation of Fréchet derivative wavefields with respect to the same physical parameter, as indicated by the 4 black diagonal boxes in Figure 5.4. This approximation can scale the amplitudes of the gradient and de-blur the gradient, but can not suppress parameter crosstalk. The parameter-type multiparameter Hessian approximation $\mathbf{H}_{a,\mathbf{p}}$ (as indicated by the white-dash lines in Figure 5.4) only keeps the diagonal elements of the blocks, which is also capable of mitigating

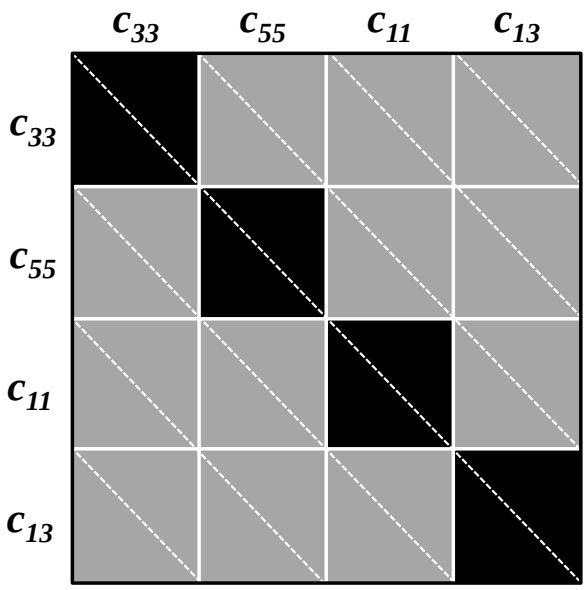

Figure 5.4: The schematic diagram of the multiparameter approximate Hessian $\mathbf{H}_a$ associated with the elastic constants c_{33}, c_{55}, c_{11} and c_{13}.

parameter crosstalk but limited in resolving the gradient spatially. When correlating the Fréchet derivative wavefields with the data residuals, the doubly-scattered energy in the data residuals will result in artifacts or spurious correlations in the gradient. The second-order term in the Hessian matrix is the correlation of the second-order partial derivative wavefields with the complex conjugate of the data residuals and it works as a de-multiple operator for suppressing the second-order scattering effects (Pratt et al., 1998b). The second-order partial derivative wavefields for multiparameter Hessian become more complex. Figure 5.5 shows a schematic diagram for multiparameter second-order scattering effects. Fréchet derivative wavefield $\nabla_{\mathbf{m}_1(\mathbf{x})}\mathbf{u}(\mathbf{x}_g, \mathbf{x}_s, \omega)$ due to the perturbation of model parameter $\mathbf{m}_1(\mathbf{x})$, is scattered secondly due to the perturbation of physical parameter $\mathbf{m}_2(\mathbf{x}')$, which yields second-order partial derivative wavefields $\nabla_{\mathbf{m}_1(\mathbf{x})}\nabla_{\mathbf{m}_2(\mathbf{x}')}\mathbf{u}(\mathbf{x}_g, \mathbf{x}_s, \omega)$. Hence, the gradient will be contaminated by artifacts due to the multiparameter doubly-scattered energy. These artifacts can be suppressed by the second-order term in multiparameter Hessian. While it is quite expensive to calculate this second-order preconditioner explicitly, this chapter will show that it can be constructed using the adjoint-state method more efficiently.

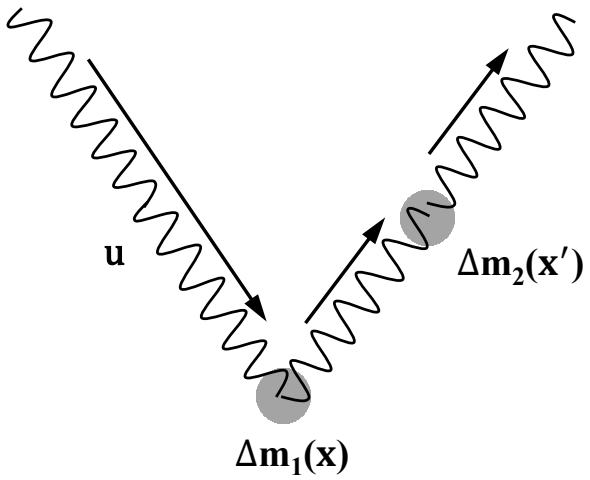

Figure 5.5: Schematic diagram of multiparameter second-order scattering. The indicate wavefields $\mathbf{u}$ is firstly scattered by the model perturbation $\Delta\mathbf{m}_1$ at $\mathbf{x}$ and then scattered again by model perturbation $\Delta\mathbf{m}_2$ at $\mathbf{x}'$.

5.3.4 Constructing multiparameter second-order preconditioner with the adjoint-state method

To calculate the second-order preconditioner for monoparameter FWI explicitly, $\left(N_x N_z\right)^2/2$ forward modelling problems need to be solved, which is extremely expensive (Pratt et al., 1998b). Considering the first-order partial derivative wavefield with respect to $\mathbf{m}$ at position $\mathbf{x}$:

$$\nabla_{\mathbf{m}(\mathbf{x})}\mathbf{u} = -\mathbf{L}\left(\mathbf{m},\omega\right)^{-1}\nabla_{\mathbf{m}(\mathbf{x})}\mathbf{L}\left(\mathbf{m},\omega\right)\mathbf{u}. \tag{5.10}$$

Taking partial derivative with respect to model parameter $\mathbf{m}\left(\mathbf{x}'\right)$ on both sides of equation (5.10) gives the equation describing the propagation of second-order partial derivative wavefield:

$$\mathbf{L}\left(\mathbf{m},\omega\right)\nabla_{\mathbf{m}(\mathbf{x})}\nabla_{\mathbf{m}(\mathbf{x}')}\mathbf{u}\left(\mathbf{x}_g,\mathbf{x}_s,\omega\right) = \tilde{\tilde{\mathbf{f}}}_s\left(\mathbf{x},\mathbf{x}',\omega\right), \tag{5.11}$$

where $\tilde{\tilde{\mathbf{f}}}_s\left(\mathbf{x},\mathbf{x}',\omega\right)$ indicates the second-order virtual source:

$$\begin{aligned}\tilde{\tilde{\mathbf{f}}}_s\left(\mathbf{x},\mathbf{x}',\omega\right) = -\sum_{\mathbf{x}_g}\sum_{\mathbf{x}_s}\Re\big(&\nabla_{\mathbf{m}(\mathbf{x})}\mathbf{L}\left(\mathbf{m},\omega\right)\nabla_{\mathbf{m}(\mathbf{x}')}\mathbf{u}\left(\mathbf{x}_g,\mathbf{x}_s,\omega\right)\\ &-\nabla_{\mathbf{m}(\mathbf{x}')}\mathbf{L}\left(\mathbf{m},\omega\right)\nabla_{\mathbf{m}(\mathbf{x})}\mathbf{u}\left(\mathbf{x}_g,\mathbf{x}_s,\omega\right)\\ &-\nabla_{\mathbf{m}(\mathbf{x})}\nabla_{\mathbf{m}(\mathbf{x}')}\mathbf{L}\left(\mathbf{m},\omega\right)\mathbf{u}\left(\mathbf{x}_g,\mathbf{x}_s,\omega\right)\big),\end{aligned} \tag{5.12}$$

where the first term indicates the second-order virtual source constructed by the interaction of Fréchet derivative wavefield $\nabla_{\mathbf{m}(\mathbf{x}')}\mathbf{u}\left(\mathbf{x}_g,\mathbf{x}_s,\omega\right)$ with model perturbation $\Delta\mathbf{m}\left(\mathbf{x}\right)$. Its

second term is the interaction of Fréchet derivative wavefield $\nabla_{\mathbf{m}(\mathbf{x})}\mathbf{u}\left(\mathbf{x}_g,\mathbf{x}_s,\omega\right)$ with model perturbation $\Delta\mathbf{m}\left(\mathbf{x}'\right)$. In monoparameter FWI, the third term in equation (5.12) is zero when $\mathbf{x}\neq\mathbf{x}'$. Isolating the second-order partial derivative wavefield in equation (5.11) and inserting it into the second term of full Hessian yields:

$$\bar{H}\left(\mathbf{x},\mathbf{x}'\right)=\sum_{\mathbf{x}_g}\sum_{\mathbf{x}_s}\sum_{\omega}\Re\left(\tilde{\tilde{\mathbf{f}}}_s^{\dagger}\left(\mathbf{x},\mathbf{x}',\omega\right)\mathbf{L}\left(\mathbf{m},\omega\right)^{-1}\Delta\mathbf{d}^{*}\left(\mathbf{x}_g,\mathbf{x}_s,\omega\right)\right). \tag{5.13}$$

It can be observed that equation (5.13) is similar to gradient calculation using the adjoint-state technique and the term $\mathbf{L}\left(\mathbf{m},\omega\right)^{-1}\Delta\mathbf{d}^{*}\left(\mathbf{x}_g,\mathbf{x}_s,\omega\right)$ serves as the backward propagated residual wavefield. Thus, the second-order preconditioner can be constructed by multiplying the backpropagated wavefield with the second-order virtual source using the adjoint-state method, which only needs N_xN_z additional forward modelling problems (Pratt et al., 1998b).

It is more complex to construct the second-order preconditioner in multiparameter Hessian because the second-order partial derivative wavefield can be caused by perturbations of different physical parameters. Considering two different physical parameters $\mathbf{m}_1$ and $\mathbf{m}_2$ and following equation (5.11), the wave equation describing the propagation of multiparameter second-order scattered wavefield is given by:

$$\mathbf{L}\left(\mathbf{m},\omega\right)\nabla_{\mathbf{m}_1(\mathbf{x})}\nabla_{\mathbf{m}_2(\mathbf{x}')}\mathbf{u}\left(\mathbf{x}_g,\mathbf{x}_s,\omega\right)=\tilde{\tilde{\mathbf{f}}}_{\mathbf{m}_1\mathbf{m}_2}\left(\mathbf{x},\mathbf{x}',\omega\right). \tag{5.14}$$

To construct the second-order partial derivative wavefield explicitly, $\left(N_pN_xN_z\right)^2/2$ forward modelling problems need to be solved. $\tilde{\tilde{\mathbf{f}}}_{\mathbf{m}_1\mathbf{m}_2}\left(\mathbf{x},\mathbf{x}',\omega\right)$ in equation (5.14) is the multiparameter second-order virtual source:

$$\begin{aligned}\tilde{\tilde{\mathbf{f}}}_{\mathbf{m}_1\mathbf{m}_2}\left(\mathbf{x},\mathbf{x}',\omega\right)=-\sum_{\mathbf{x}_g}\sum_{\mathbf{x}_s}\Re\big(&\nabla_{\mathbf{m}_1(\mathbf{x})}\mathbf{L}\left(\tilde{\mathbf{m}},\omega\right)\nabla_{\mathbf{m}_2(\mathbf{x}')}\mathbf{u}\left(\mathbf{x}_g,\mathbf{x}_s,\omega\right)\\&-\nabla_{\mathbf{m}_1(\mathbf{x}')}\mathbf{L}\left(\tilde{\mathbf{m}},\omega\right)\nabla_{\mathbf{m}_2(\mathbf{x})}\mathbf{u}\left(\mathbf{x}_g,\mathbf{x}_s,\omega\right)\\&-\nabla_{\mathbf{m}_1(\mathbf{x})}\nabla_{\mathbf{m}_2(\mathbf{x}')}\mathbf{L}\left(\tilde{\mathbf{m}},\omega\right)\mathbf{u}\left(\mathbf{x}_g,\mathbf{x}_s,\omega\right)\big),\end{aligned} \tag{5.15}$$

where $\tilde{\mathbf{m}}$ denotes all of the physical parameters considered. The first term in equation (5.15) indicates the multiparameter second-order virtual source caused by the interaction of the Fréchet derivative wavefield $\nabla_{\mathbf{m}_1(\mathbf{x}')}\mathbf{u}\left(\mathbf{x}_g,\mathbf{x}_s,\omega\right)$ with $\Delta\mathbf{m}_2\left(\mathbf{x}\right)$. The second term is formed by the interaction of the Fréchet derivative wavefield $\nabla_{\mathbf{m}_2(\mathbf{x}')}\mathbf{u}\left(\mathbf{x}_g,\mathbf{x}_s,\omega\right)$ with $\Delta\mathbf{m}_2\left(\mathbf{x}'\right)$. The value of the third term in equation (5.15) is determined by the parameterization for

describing the subsurface media. If $\mathbf{m}_1$ and $\mathbf{m}_2$ are independent physical parameters, the third term in equation (5.15) is zero. However, if $\mathbf{m}_1$ and $\mathbf{m}_2$ are not independent physical parameters (e.g., P-wave velocity α and density ρ'), the third term is not zero (Fichtner and Trampert, 2011b). In this research, this term is ignored because the four elastic constants used to describe the 2D HTI media are independent.

It is also possible to calculate the second-order preconditioner for multiparameter FWI using the adjoint-state technique. Similar to equation (5.13), the multiparameter second-order preconditioner can be expressed as:

$$\bar{H}_{\mathbf{m}_1\mathbf{m}_2}(\mathbf{x},\mathbf{x}') = \sum_{\mathbf{x}_g}\sum_{\mathbf{x}_s}\sum_{\omega} \Re\left(\tilde{\mathbf{f}}^{\dagger}_{\mathbf{m}_1\mathbf{m}_2}(\mathbf{x},\mathbf{x}',\omega)\,\mathbf{L}(\tilde{\mathbf{m}},\omega)^{-1}\,\Delta\mathbf{d}^{*}(\mathbf{x}_g,\mathbf{x}_s,\omega)\right). \tag{5.16}$$

Thus, additional $N_pN_xN_z$ forward modelling simulations are required for constructing the multiparameter second-order preconditioner. For example, to inverse the four elastic constants in HTI media, the off-diagonal block $\bar{\mathbf{H}}_{c_{33}c_{55}}$ in the second-order preconditioner is expressible as:

$$\bar{H}_{c_{33}c_{55}}(\mathbf{x},\mathbf{x}') = \sum_{\mathbf{x}_g}\sum_{\mathbf{x}_s}\sum_{\omega} \Re\left(\tilde{\mathbf{f}}^{\dagger}_{c_{33}c_{55}}(\mathbf{x},\mathbf{x}',\omega)\,\mathbf{L}(\tilde{\mathbf{m}},\omega)^{-1}\,\Delta\mathbf{d}^{*}(\mathbf{x}_g,\mathbf{x}_s,\omega)\right), \tag{5.17}$$

where $\tilde{\mathbf{m}}$ indicate all of the four elastic constants c_{33}, c_{55}, c_{11} and c_{13}.

5.4 Numerical Examples

In this section, several numerical examples are provided to test the proposed methods. The scattering patterns of the elastic constants in HTI media are first examined by comparing analytic results with numerical results. The multiparameter Hessian using a 2D HTI model is presented. The effectiveness of multiparameter Hessian in suppressing parameter crosstalk and second-order scattering effects are verified. Finally, the GN and FN multiparameter FWI are enacted on a two-block-layer model.

5.4.1 Scattering patterns of the elastic constants: analytic vs. numerical results

In this numerical example, the analytic and numerical scattering patterns of the elastic constants are examined for parameter crosstalk analysis. The x-z plane with zero azimuth

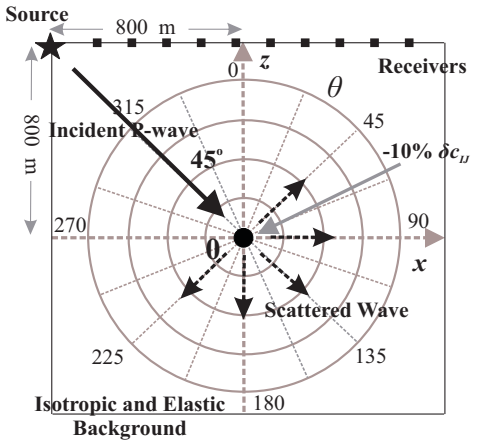

Figure 5.6: The 2D numerical model for examining the scattering patterns of elastic constants. The black point at the center of the model indicates the HTI anomaly at position $\mathbf{x}_0$ =(0 m, 0 m, 0 m). The source is located at $\mathbf{x}_s$ =(−800 m, 0 m, 800 m). The receivers (the black squares) are arranged along the top surface for a reflection survey.

angle in 3D geometry (Figure 5.1) is extracted, which forms the specified 2D HTI model. The acquisition geometry is shown in Figure 5.6. The model consists of 320×320 grid cells with grid sizes $\Delta x = \Delta z = 5$ m. The background model is isotropic-elastic with elastic constants $c_{33} = 14.06$ GPa, $c_{55} = 6.32$ GPa, $c_{11} = 14.06$ GPa and $c_{13} = 1.42$ GPa (P-wave velocity $\alpha = 2651.4$ m/s, S-wave velocity $\beta = 1777.6$ m/s, density $\rho' = 2.0$ g/cm^3). -10% perturbations are applied to the elastic constants of the node located at the center of the model, which forms the anisotropic anomaly (the black circle point in Figure 5.6). One source is located at top-left corner of the 2D model, as indicated by the black star in Figure 5.6. When the incident P-wave ($\vartheta = 135^o$ and $\varphi = 0^o$) interacts with the HTI anomaly, the scattered wave will propagate at all directions and its amplitudes change with varying angle θ. The receivers are deployed along the top surface of the model for a reflection survey, which means that only the scattered wave at the range of $\theta \in [315^o, 360^o]$ and $\theta \in [0^o, 45^o]$ are recorded.

The x-z plane is extracted from the analytic 3D scattering patterns for these elastic constants. The bold-black curves in Figures 5.7a, 5.7b, 5.7c and 5.7d show the P-P scattering patterns due to Δc_{33}, Δc_{55}, Δc_{11} and Δc_{13} with incident P-wave ($\vartheta = 135^o$ and $\varphi = 0^o$). Perturbations of different elastic constants serve as different types of secondary sources as-

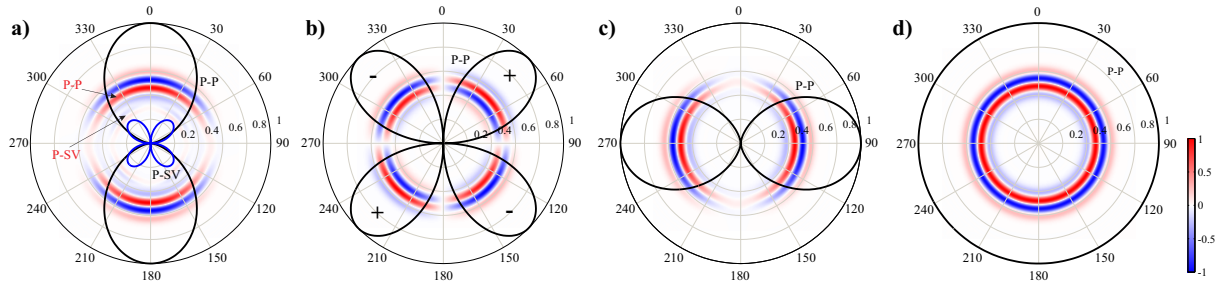

Figure 5.7: Analytic vs. numerical results of the scattering patterns for the elastic constants in 2D HTI media. (a), (b), (c) and (d) show the scattering patterns due to elastic constants Δc_{33}, Δc_{55}, Δc_{11} and Δc_{13} respectively. The symbols "+" and "-" in (b) mean positive and negative polarities of the scattered wave. The amplitudes have been normalized.

sociated with different scattering patterns. The P-SV scattering pattern due to Δc_{33} is also plotted, as indicated by the blue curve in Figure 5.7a. The analytic scattering patterns are overlain by numerical modelling results for comparison. The amplitude variations of the analytic scattering patterns are consistent with those of the numerical results.

In angle regimes where the scattering pattern of one parameter is indistinguishable from that of another, the influences of the two parameters are not separable, and crosstalk appears. Comparing the scattering pattern in Figure 5.7a with that in Figure 5.7d, for instance, it is observed that the P-P scattering patterns due to Δc_{33} and Δc_{13} are significantly overlapped at near offset, indicating strong crosstalk between c_{33} and c_{13} for this reflection survey. Furthermore, parameter crosstalk between c_{55} and c_{13} is very strong at mid offset. In Figure 5.7c, it is seen that strong scattered wavefields response due to Δc_{11} can only be recorded at large offset. For inversing the elastic constants using FWI, the parameter crosstalk among these parameters are strong and complex for this reflection acquisition survey, which will undermine the inversion process without proper preconditioning.

5.4.2 Suppressing parameter crosstalk with multiparameter approximate Hessian

To examine the ability of the multiparameter approximate Hessian to suppress parameter crosstalk, the Gauss-Newton update is enacted on a 2D HTI point scatterer model. The 2D HTI model consists of 900 nodes ($N_x = N_z = 30$) with grid size of 5 m in both horizontal and vertical dimensions and four elastic constants (c_{33}, c_{55}, c_{11} and c_{13}) are used to describe each node. The initial model is elastic and isotropic with elastic constants $c_{33} = 14.06$

GPa, $c_{55} = 6.32$ GPa, $c_{11} = 14.06$ GPa and $c_{13} = 1.42$ GPa. The true model is built by embedding one HTI point anomaly at the center position of the background model and four elastic constants are all perturbed by -10% at this point scatterer. A 50 Hz Ricker wavelet is used for forward modelling and the absorbing boundary condition is applied on all of the boundaries of the model. In these numerical examples presented in this research, multi-component data is used for inversion.

First, the multiparameter approximate Hessian $\mathbf{H}_a$ is constructed explicitly with one source located at $\mathbf{x}_{s1}$ =(75 m, 0 m, 0 m). Thirty receivers are arranged along the top surface of the model with a spacing of 5 m. As shown in Figure 5.8, the multiparameter approximate Hessian is a 3600×3600 square and symmetric matrix with 4 diagonal blocks and 12 off-diagonal blocks, which are consistent with the schematic diagram shown in Figure 5.4. Each block matrix is a 900×900 square matrix.

It can be seen that the sub-blocks in multiparameter approximate Hessian are banded due to finite-frequency effects. Because elastic constant c_{33} directly relates to P-wave velocity α ($c_{33} = \rho'\alpha^2$) and the Fréchet derivative wavefield caused by Δc_{33} recorded at the receivers are much stronger than those due to other elastic constants, the diagonal block $\mathbf{H}_{a,3333}$ dominates the whole matrix. The 4 diagonal blocks $\mathbf{H}_{a,3333}$, $\mathbf{H}_{a,5555}$, $\mathbf{H}_{a,1111}$ and $\mathbf{H}_{a,1313}$ are extracted, as shown in Figures 5.9a, 5.9b, 5.9c and 5.9d respectively (the amplitudes have been re-normalized). These 4 diagonal blocks form the space-type multiparameter Hessian approximation $\mathbf{H}_{a,\mathrm{s}}$ given by Innanen (2014b). Furthermore, the energy distribution in the 4 diagonal blocks are quite different, which are determined by the scattering patterns of these elastic constants.

The diagonal elements in the diagonal blocks $\mathbf{H}_{a,3333}$, $\mathbf{H}_{a,5555}$, $\mathbf{H}_{a,1111}$ and $\mathbf{H}_{a,1313}$ are extracted and plotted in model space (as shown in Figures 5.10a, 5.10b, 5.10c and 5.10d) and they mainly account for illumination compensation and removing the geometrical spreading effects. The parameter crosstalk between different physical parameters are measured by the 12 off-diagonal blocks of $\mathbf{H}_a$, as shown in Figure 5.8. Stronger amplitudes in the off-diagonal blocks means stronger parameter crosstalk. Figures 5.10e, 5.10f, 5.10g, 5.10h, 5.10i, and 5.10j show the diagonal elements of the off-diagonal blocks $\mathbf{H}_{a,3355}$, $\mathbf{H}_{a,3311}$, $\mathbf{H}_{a,3313}$, $\mathbf{H}_{a,5511}$, $\mathbf{H}_{a,5513}$ and $\mathbf{H}_{a,1113}$ respectively, which mainly account for removing the parameter crosstalk

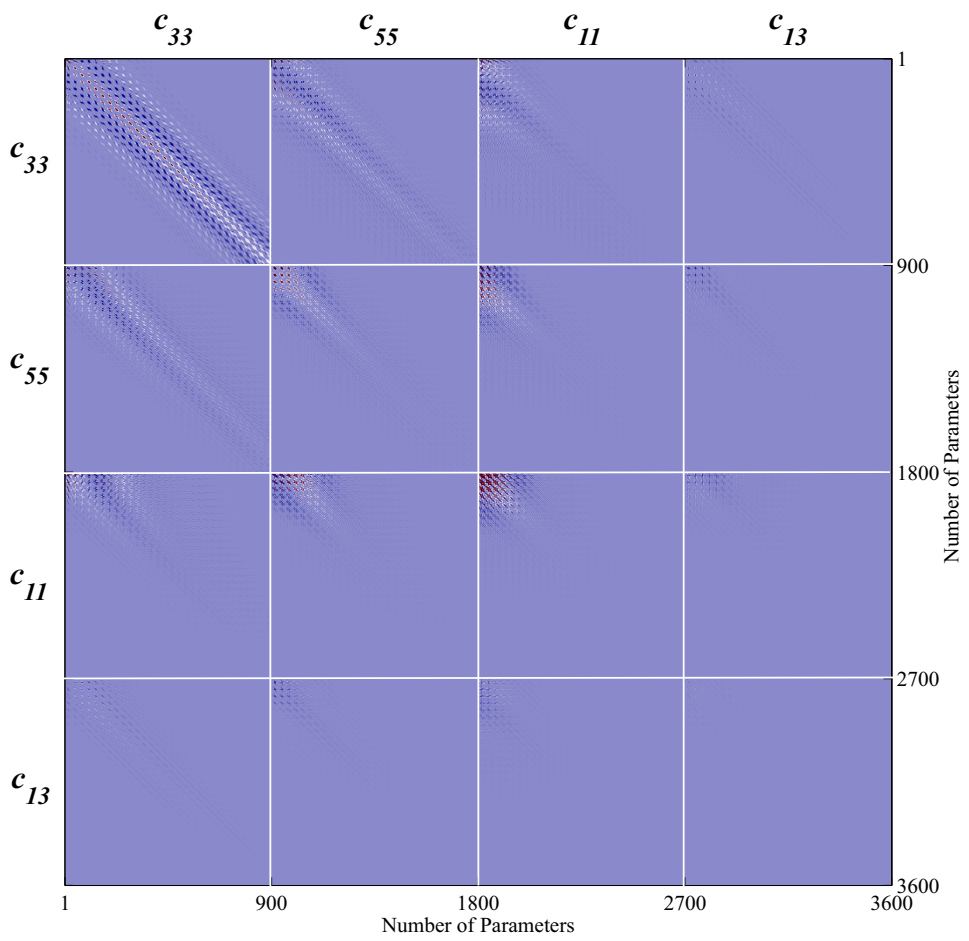

Figure 5.8: The multiparameter approximate Hessian $\mathbf{H}_a$ for elastic constants c_{33}, c_{55}, c_{11} and c_{13} with the 2D HTI model. The multiparameter approximate Hessian is a 3600×3600 square and symmetric matrix ($N_p = 4$ and $N_x = N_z = 30$).

(Innanen, 2014a).

A numerical example is presented to show that preconditioning the gradient with the multiparameter approximate Hessian $\mathbf{H}_a$ can suppress parameter crosstalk and resolve the gradient. The search direction $\Delta\mathbf{m}_k$ associated with the Gauss-Newton update can be obtained by solving the Newton linear system (equation (2.12)) approximately using a conjugate-gradient (CG) algorithm, which is known as truncated-Newton method (Métivier et al., 2013; Pan et al., 2017a). In this chapter, the gradient is preconditioned by the pseudo-inverse of the multiparameter approximate Hessian $\mathbf{H}_a^{-1}$, which is calculated using Singular Value Decomposition (SVD).

First, the data residual vector $\Delta\mathbf{d}_{33}$ caused by perturbation of c_{33} is used to construct

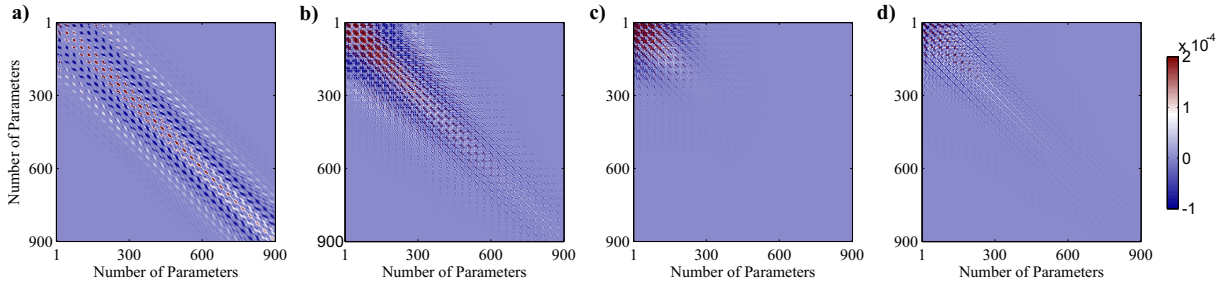

Figure 5.9: The diagonal blocks of the multiparameter approximate Hessian shown in Figure 5.8. (a), (b), (c) and (d) show the diagonal blocks $\mathbf{H}_{a,3333}$, $\mathbf{H}_{a,5555}$, $\mathbf{H}_{a,1111}$ and $\mathbf{H}_{a,1313}$ of the multiparameter approximate Hessian respectively.

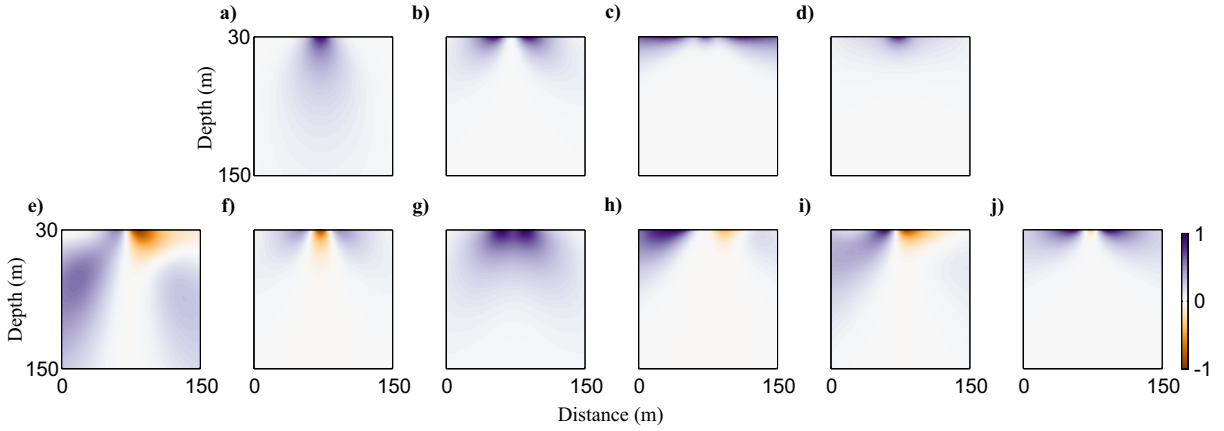

Figure 5.10: The diagonal elements of the block matrices plotted in model space. (a), (b), (c) and (d) show the diagonal elements of the diagonal block matrices $\mathbf{H}_{a,3333}$, $\mathbf{H}_{a,5555}$, $\mathbf{H}_{a,1111}$ and $\mathbf{H}_{a,1313}$ respectively. (e), (f), (g), (h), (i) and (j) show the diagonal elements of the off-diagonal block matrices $\mathbf{H}_{a,3355}$, $\mathbf{H}_{a,3311}$, $\mathbf{H}_{a,3313}$, $\mathbf{H}_{a,5511}$, $\mathbf{H}_{a,5513}$ and $\mathbf{H}_{a,1113}$ respectively. Because the symmetry of the multiparameter approximate Hessian $\mathbf{H}_a$, only the diagonal elements of 6 off-diagonal blocks in $\mathbf{H}_a$ are plotted.

the gradients of all elastic constants c_{33}, c_{55}, c_{11} and c_{13}. Figures 5.11a, 5.11b, 5.11c and 5.11d show the gradient vectors $\nabla_{c_{33}}\Phi$, $\nabla_{c_{55}}\Phi$, $\nabla_{c_{11}}\Phi$ and $\nabla_{c_{13}}\Phi$ without multiparameter approximate Hessian preconditioning respectively. Only the gradient vector $\nabla_{c_{33}}\Phi$ is real and the gradient vectors for other elastic constants are all artifacts caused by parameter crosstalk. The multiparameter approximate Hessian $\mathbf{H}_a$ is then applied to precondition the gradient vectors and the estimated model perturbations for the elastic constants c_{33}, c_{55}, c_{11} and c_{13} are illustrated in Figures 5.11e, 5.11f, 5.11g and 5.11h respectively. It can be observed that the artifacts in Figures 5.11b, 5.11c and 5.11d have been obviously removed and the estimated model perturbation for c_{33} in Figure 5.11a is resolved and de-blurred. Figures 5.11i, 5.11j, 5.11k and 5.11l show the gradient vectors calculated using the data residual vector $\Delta\mathbf{d}_{55}$ due to Δc_{55}. Similarly, only the gradient vector $\nabla_{c_{55}}\Phi$ in Figure 5.11j is real and other gradient vectors in Figures 5.11i, 5.11k and 5.11l are all spurious correlations. Figures 5.11m, 5.11n, 5.11o and 5.11p show the model perturbation estimations with multiparameter approximate Hessian preconditioning for the four elastic constants respectively. It can be seen that the artifacts in Figures 5.11i, 5.11k and 5.11l are suppressed and the gradient vector $\nabla_{c_{55}}\Phi$ in Figure 5.11j is resolved obviously. These two numerical examples show the ability of the multiparameter approximate Hessian in suppressing parameter crosstalk artifacts and resolving the gradient vectors.

Figures 5.12a, 5.12b, 5.12c and 5.12d show the gradient vectors obtained using the data residuals $\Delta\mathbf{d}$ due to the perturbations of four elastic constants with 3 sources. The three sources are located at $\mathbf{x}_{s1}$ =(75 m, 0 m, 0 m), $\mathbf{x}_{s2}$ =(0 m, 0 m, 0 m) and $\mathbf{x}_{s3}$ =(150 m, 0 m, 0 m) respectively. Figures 5.12e, 5.12f, 5.12g and 5.12h show the estimated model perturbations with multiparameter approximate Hessian preconditioning. Figures 5.12i, 5.12j, 5.12k and 5.12l show the estimated model perturbations of the elastic constants using Gauss-Newton multiparameter FWI after 3 iterations. It can be seen that the four elastic constants can be inverted simultaneously very well with multiparameter approximate Hessian preconditioning.

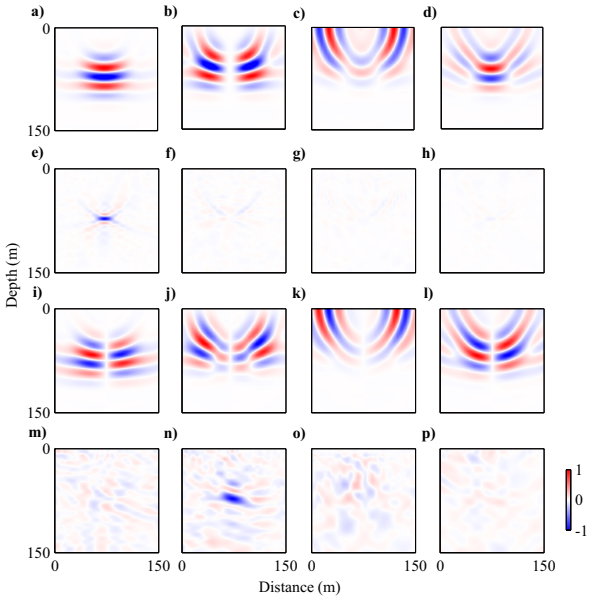

Figure 5.11: (a), (b), (c), and (d) show the gradient vectors $\nabla_{c_{33}}\Phi$, $\nabla_{c_{55}}\Phi$, $\nabla_{c_{11}}\Phi$ and $\nabla_{c_{13}}\Phi$ constructed by the data residual vector $\Delta\mathbf{d}_{33}$. (e), (f), (g) and (h) show the perturbation estimations after multiparameter approximate Hessian preconditioning for the corresponding elastic constants. (i), (j), (k) and (l) show the gradient vectors constructed by data residual vector $\Delta\mathbf{d}_{55}$ for the corresponding elastic constants. (m), (n), (o) and (p) show the perturbation estimations for the corresponding elastic constants with multiparameter approximate Hessian preconditioning.

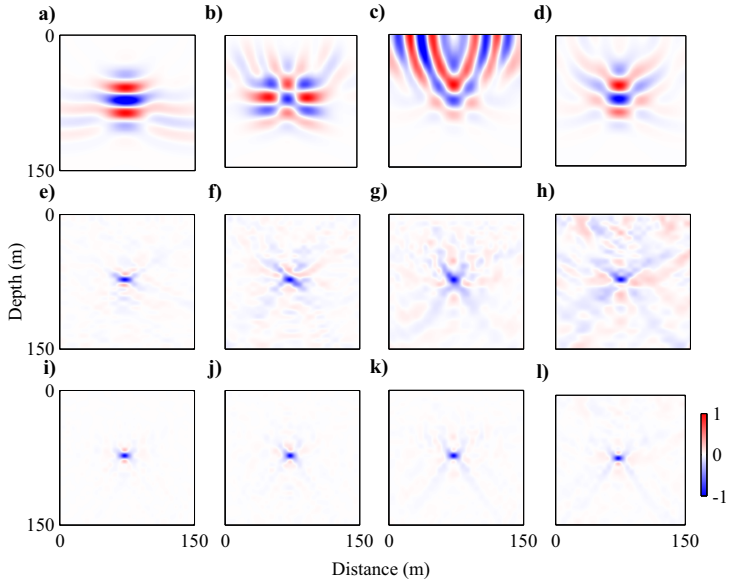

Figure 5.12: (a), (b), (c), and (d) show the gradient vectors $\nabla_{c_{33}}\Phi$, $\nabla_{c_{55}}\Phi$, $\nabla_{c_{11}}\Phi$ and $\nabla_{c_{13}}\Phi$ constructed by $\Delta\mathbf{d}$ using 3 sources. (e), (f), (g) and (h) show the perturbation estimations with multiparameter approximate Hessian preconditioning for the corresponding elastic constants. (i), (j), (k) and (l) show estimated perturbations for elastic constants after 3 iterations. The amplitudes have been normalized.

5.4.3 Suppressing multiparameter second-order scattering effects

In this numerical example, the effectiveness of second-order preconditioner in suppressing multiparameter second-order scattering effects is shown. The second-order term $\bar{\mathbf{H}}$ is constructed by correlating the second-order partial derivative wavefields with the data residuals. When considering multiple physical parameters, the second-order partial derivative wavefields can be caused by different physical parameters. Furthermore, it is quite expensive to calculate the second-order term directly. In this research, the adjoint-state method is used to calculate the second-order term $\bar{\mathbf{H}}$ in the multiparameter Hessian $\mathbf{H}$.

Considering the elastic and isotropic background model used in previous example, two HTI point anomalies are embedded in the background model, as shown in Figure 5.13. The two HTI points anomalies are located at $\mathbf{x}_1 = (80\text{ m}, 0\text{ m}, 65\text{ m})$ and $\mathbf{x}_2 = (90\text{ m}, 0\text{ m}, 75\text{ m})$. At position $\mathbf{x}_1$, the elastic constants c_{33}, c_{55}, c_{11} and c_{13} are perturbed by $+10\%$, $+10\%$, 0% and $+10\%$ respectively. At position $\mathbf{x}_2$, the four elastic constants are perturbed by -10%, -10%, -10% and 0% respectively. The normalized true model perturbations for

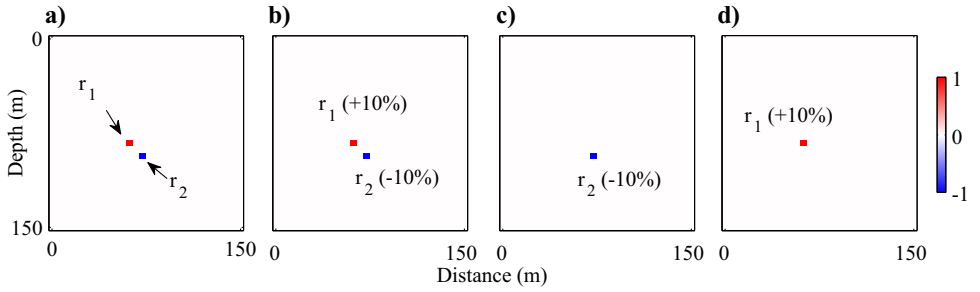

Figure 5.13: True model perturbations of the two scatterers model. (a), (b), (c) and (d) show the true model perturbations for elastic constants c_{33}, c_{55}, c_{11} and c_{13} respectively. The amplitudes have been normalized.

elastic constants c_{33}, c_{55}, c_{11} and c_{13} are illustrated in Figures 5.13a, 5.13b, 5.13c and 5.13d respectively.

Figure 5.14 shows the elements of first-order and second-order terms of the multiparameter Hessian plotted in model space. Considering the model parameter position $\mathbf{x}_2$, the correlation of the Fréchet derivative wavefield due to $\Delta c_{33}(\mathbf{x}_2)$ with the Fréchet derivative wavefields due to $\Delta c_{33}(\tilde{\mathbf{x}})$ ($\tilde{\mathbf{x}}$ indicate all positions in the model) forms the 555th row in diagonal block $\mathbf{H}_{a,3333}$, as shown in Figure 5.14a. Figures 5.14b, 5.14c and 5.14d show the 555th rows in the off-diagonal blocks $\mathbf{H}_{a,3355}$, $\mathbf{H}_{a,3311}$ and $\mathbf{H}_{a,3313}$ respectively. Stronger amplitudes mean stronger correlations of the Fréchet derivative wavefields.

The Fréchet derivative wavefield due to $\Delta c_{33}(\mathbf{x}_2)$ can be further scattered due to $\Delta c_{33}(\tilde{\mathbf{x}})$ or $\Delta c_{55}(\tilde{\mathbf{x}})$, $\Delta c_{11}(\tilde{\mathbf{x}})$ and $\Delta c_{13}(\tilde{\mathbf{x}})$. Correlating the multiparameter second-order scattered wavefield with the data residuals forms the 555th rows of diagonal block $\bar{\mathbf{H}}_{3333}$, off-diagonal blocks $\bar{\mathbf{H}}_{3355}$, $\bar{\mathbf{H}}_{3311}$ and $\bar{\mathbf{H}}_{3313}$, as shown in Figures 5.14e, 5.14f, 5.14g and 5.14h, which are obtained using explicit perturbation method with additional 900 forward modelling simulations (Pratt et al., 1998b). Stronger amplitudes mean stronger correlations between the second-order scattered wavefields with the data residuals. Figures 5.14i, 5.14j, 5.14k and 5.14l show the 555th rows in blocks $\bar{\mathbf{H}}_{3333}$, $\bar{\mathbf{H}}_{3355}$, $\bar{\mathbf{H}}_{3311}$ and $\bar{\mathbf{H}}_{3313}$ calculated using the adjoint-state method following equation (5.17) with additional 1 forward modelling simulation. Constructing the multiparameter second-order preconditioner with the adjoint-state method, an additional 3600 forward modelling simulations are required.

A numerical example is given to show the artifacts caused by the second-order scattering

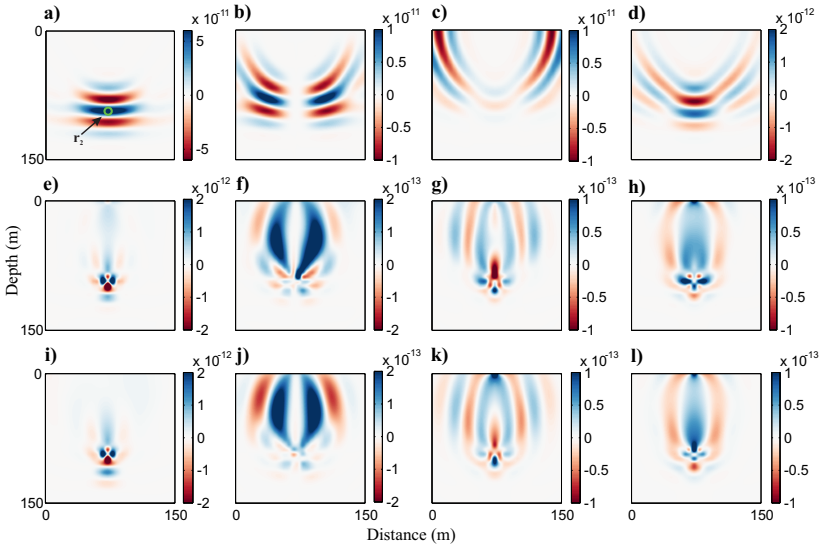

Figure 5.14: The multiparameter Hessian $\mathbf{H}$ plotted in model space. The green circle in (a) shows parameter position $\mathbf{x}_2$. (a), (b), (c) and (d) show the 555th rows in the diagonal block $\mathbf{H}_{a,3333}$ and off-diagonal blocks $\mathbf{H}_{a,3355}$, $\mathbf{H}_{a,3311}$ and $\mathbf{H}_{a,3313}$. (e), (f), (g) and (h) show the 555th rows in the diagonal block $\bar{\mathbf{H}}_{3333}$ and off-diagonal blocks $\bar{\mathbf{H}}_{3355}$, $\bar{\mathbf{H}}_{3311}$ and $\bar{\mathbf{H}}_{3313}$. (i), (j), (k) and (l) show the 555th rows in the blocks $\bar{\mathbf{H}}_{3333}$, $\bar{\mathbf{H}}_{3355}$, $\bar{\mathbf{H}}_{3311}$ and $\bar{\mathbf{H}}_{3313}$ using adjoint-state method.

effects. Figures 5.15a, 5.15b, 5.15c and 5.15d show the Gauss-Newton updates for Δc_{33} when the true model perturbation Δc_{33} was increased from 10% to 20%, 30% and 40% respectively. Larger model perturbation means stronger second-order scattered energy in the data residuals. It can be seen that the artifacts become stronger with increasing the model perturbation. Figures 5.15e, 5.15f, 5.15g and 5.15h show the Gauss-Newton updates for c_{55} when increasing the model perturbation Δc_{33}. It can be seen that the artifacts become very strong in Figures 5.15g and 5.15h and it is difficult to recognize the anomalies at positions $\mathbf{x}_1$ and $\mathbf{x}_2$. Figures 5.15i, 5.15j, 5.15k and 5.15l show the inverted model perturbations for elastic constants of c_{33}, c_{55}, c_{11} and c_{13} using GN method and Figures 5.15m, 5.15n, 5.15o and 5.15p show the inverted model perturbations for the elastic constants using FN method after 5 iterations when model perturbation Δc_{33} is 30%. For GN method, the elastic constants c_{33} and c_{55} can be determined very well (Figures 5.15i and 5.15j). However, for c_{11} and c_{13} (Figures 5.15k and 5.15l), the artifacts are still very strong. While, FN method can suppress the artifacts in estimated model perturbations Δc_{11} and Δc_{13} (Figures 5.15o and 5.15p) obviously for incorporating the multiparameter second-order preconditioner.

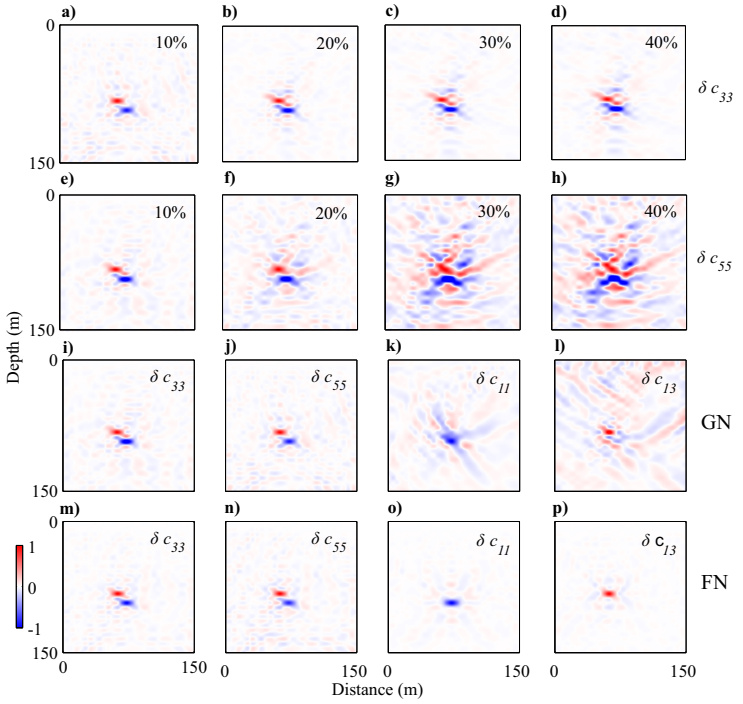

Figure 5.15: (a), (b), (c) and (d) show the estimated model perturbation Δc_{33} with multiparameter approximate Hessian $\mathbf{H}_a$ preconditioning when c_{33} is perturbed by 10%, 20%, 30% and 40%; (e), (f), (g) and (h) show the estimated model perturbation Δc_{55} with increasing model perturbations of c_{33}; (i), (j), (k) and (l) show the estimated model perturbations Δc_{33}, Δc_{55}, Δc_{11} and Δc_{11} after 5 iterations with GN method; (m), (n), (o) and (p) show the estimated model perturbations for the corresponding elastic constants after 5 iterations with FN method.

5.4.4 Applying GN and FN multiparameter FWI on a two-block-layer model

Finally, the GN and FN multiparameter FWI are enacted on a two-block-layer model for comparison. The model consists of 50×50 grid points with $\Delta x = \Delta z = 10$ m and a 20 Hz Ricker source wavelet is used for forward modelling. The initial model used in this numerical example is elastic and isotropic and its properties are consistent with those used in previous numerical examples. Two anisotropic block layers are embedded in the isotropic background and the true perturbations for elastic constants c_{33}, c_{55}, c_{11} and c_{13} are shown in Figures 5.16a, 5.16b, 5.16c and 5.16d. For the first block layer, the perturbations for elastic constants c_{33}, c_{55}, c_{11} and c_{13} are -4.218 GPa (-30%), -0.632 GPa (-10%), -1.406 GPa (-10%) and 0 GPa (0%). For the second block layer, the perturbations for these elastic constants are $+4.218$ GPa ($+30\%$), $+0.632$ GPa ($+10\%$), 0 GPa (0%) and $+0.142$ GPa ($+10\%$). The doubly-scattered energy between the two block layers can cause artifacts in the estimated model perturbations. A total of 10 iterations are applied for inversion using GN and FN methods. A multi-scale approach is employed by increasing the frequency band from [1 Hz, 10 Hz] to [1 Hz, 19 Hz] by 1 Hz every iteration (Pratt and Worthington, 1990; Sirgue and Pratt, 2004). To evaluate the quality of the inversion results, the relative least-squares error (RLSE) (equation (2.42)) is used.

Figures 5.16e, 5.16f, 5.16g and 5.16h show the inverted model perturbations for elastic constants c_{33}, c_{55}, c_{11} and c_{13} using GN multiparameter FWI. It can be seen that for elastic constants c_{33} and c_{55}, GN method can get acceptable results even though the two block layers are not de-blurred very well. While for elastic constants c_{11} and c_{13}, the estimated model perturbations are contaminated by strong artifacts. Figures 5.16i, 5.16j, 5.16k and 5.16l show the inverted model perturbations for these elastic constants using FN multiparameter FWI. It can be observed that the two block layers for c_{33} are de-blurred better and the artifacts for elastic constants c_{11} and c_{13} have been suppressed. Figures 5.17a, 5.17b, 5.17c and 5.17d show the RLSE (equation (2.42)) $\tilde{\epsilon}_{33}$, $\tilde{\epsilon}_{55}$, $\tilde{\epsilon}_{11}$ and $\tilde{\epsilon}_{13}$ for elastic constants c_{33}, c_{55}, c_{11} and c_{13} as the iteration proceeds. The solid lines and dash lines indicate the RLSE obtained using GN and FN methods. FN method can estimate the model perturbations more efficiently than GN method by incorporating the second-order term in multiparameter Hessian. Furthermore, Figures 5.17c and 5.17d reveal that the effectiveness of FN method

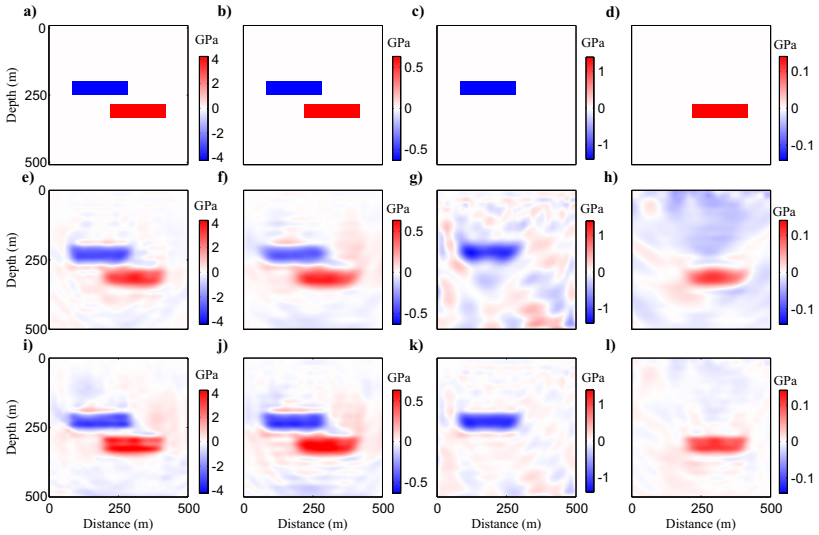

Figure 5.16: Inversion results comparison. (a), (b), (c) and (d) show the true model perturbations for the elastic constants c_{33}, c_{55}, c_{11} and c_{13} respectively. (e), (f), (g) and (h) show the inversion results for the elastic constants c_{33}, c_{55}, c_{11} and c_{13} using GN FWI after 10 iterations. (i), (j), (k) and (l) show the inversion results for the corresponding elastic constants using FN FWI after 10 iterations.

is more obvious in estimating elastic constants c_{11} and c_{13}.

5.5 Discussion

This chapter develops FWI techniques for inverting the properties of naturally fractured reservoirs. Robust technology of this kind is expected to have significant impact in areas like reservoir characterization. Fractured reservoirs can be described using many different parameters, such as fracture spacing, fracture density, fracture orientation, weakness, compliance, etc. Our current study focuses on inverting for the elastic constants of equivalent HTI media. Different parameterizations impact the inversion process greatly, in large part because of parameter crosstalk. Comparison of the stability and efficiency given different parameterizations is an important area of future research. The 3D scattering patterns for elastic constants in general anisotropic media given in this paper can be used to analyze the parameter crosstalk problem when inverting for the elastic constants. These can subsequently be transformed to any parameterization following the chain rule. The analytic 3D scattering patterns have been examined using a 2D HTI numerical example. Three di-

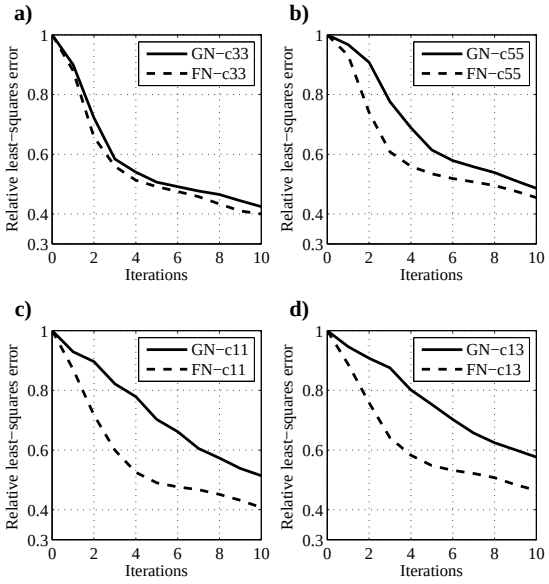

Figure 5.17: Comparison of the RLSE for different elastic constants using GN and FN methods. (a), (b), (c) and (d) show the RLSE $\tilde{\epsilon}_{33}$, $\tilde{\epsilon}_{55}$, $\tilde{\epsilon}_{11}$ and $\tilde{\epsilon}_{13}$ for elastic constants c_{33}, c_{55}, c_{11} and c_{13} as the iteration proceeds.

mensional examples for more complex media (such as orthorhombic media) can be carried out in future studies. The detailed spatial features of the fractured reservoirs by using the long wavelength approximation and describe the fractured media using anisotropic elastic constants are ignored. It will be valuable to consider the fracture size when using FWI for fractured reservoir characterization (Fang et al., 2013; Zheng et al., 2013; Pan and Innanen, 2013).

Full-waveform inversion is an ill-posed problem, which means that an infinite number of models matches the data (Virieux and Operto, 2009). Regularization technique can alleviate the non-uniqueness of the ill-posed inverse problem and make FWI better posed (Menke, 1984). In this paper, no regularization technique is employed. Hence, for further research, introducing regularization technique, such as Tikhonov regularization (Asnaashari et al., 2013) and Total-variation regularization (Lin, 2015), in the objective function is necessary for improving the performance of the proposed strategies.

Most current inversion strategies for multiparameter FWI are hierarchical methods, with parameterization and acquisition geometry having been selected to mitigate or avoid parameter crosstalk problem. Prieux et al. (2013b) considered visco-acoustic multiparameter

FWI using the l-BFGS optimization method with simultaneous strategy, compared against a hierarchical strategy. In this thesis, the GN and FN methods are used to invert for the elastic constants simultaneously, with the multiparameter Hessian acting to suppress parameter crosstalk. Métivier et al. (2014) applied a truncated-Newton method for single-parameter FWI with matrix-free method, which can also be employed for simultaneous inversion strategy of multiparameter FWI.

In the numerical section, I have restricted our selection to 2D numerical examples, for examining the possibilities of inverting for the elastic constants. For large 2D or 3D practical applications, it will be much more computationally expensive to carry out Gauss-Newton and full-Newton optimization methods for multiparameter FWI. One possible recourse is target-oriented FWI, in which only one portion of the multiparameter Hessian need to be calculated aiming at the target area. Phase-encoding methods have been widely studied for calculating the gradient (Vigh and Starr, 2008; Tang, 2009; Anagaw and Sacchi, 2014; Pan et al., 2014a) or Hessian approximations (Castellanos et al., 2015). The phase-encoding methods can also be used in multiparameter FWI to calculate the gradient and multiparameter Hessian for reducing the computational burden.

5.6 Conclusions

In this chapter, I have applied the Gauss-Newton and full-Newton multiparameter FWI to invert for the elastic constants of a HTI media. The parameter crosstalk difficulty in multiparameter FWI is introduced and I also derive the 3D scattering patterns for the elastic constants in general anisotropic media for parameter crosstalk analysis. The roles of the multiparameter Hessian in mitigating parameter crosstalk and reducing second-order scattering effects have been revealed. I also explain how to construct the multiparameter second-order preconditioner using the adjoint-state method. In the numerical section, I give examples to testify the effectiveness of the multiparameter Hessian in suppressing parameter crosstalk and second-order scattering effects. The Gauss-Newton and full-Newton FWI are finally applied on a two-block-layer model for comparison. The full-Newton method gave better inversion results for incorporating the multiparameter second-order preconditioner.

Appendix F

3D Scattering Patterns of Elastic Constants in General Anisotropic Media

In this Appendix, I first derive the 3D scattering patterns for the elastic stiffness coefficients in general anisotropic media and then give the explicit expressions of the scattering patterns of the elastic stiffness coefficients in HTI media. The solution of equation (5.1) can be expressed using the integral form of the Green's tensor vector in the frequency domain (Ben-Menahem and Singh, 1981; Aki and Richards, 2002; Kamath and Tsvankin, 2014):

$$\bar{u}_i\left(\mathbf{x},\omega\right)=\int_{\Omega(\mathbf{x}_s)}\int_{\omega_s} f_j(\mathbf{x}_s,\omega_s)G_{ij}(\mathbf{x},\omega;\mathbf{x}_s,\omega_s)d\Omega(\mathbf{x}_s)d\omega_s, \tag{F.1}$$

where $G_{ij}(\mathbf{x},\omega;\mathbf{x}_s,\omega_s)$ indicates the i component of the Green's tensor vector at position $\mathbf{x}$ due to a point source $f_j(\mathbf{x}_s,\omega_s)$ in j direction at position $\mathbf{x}_s$. $\Omega\left(\mathbf{x}_s\right)$ indicates the volume including all of the sources.

Considering that a general anisotropic inclusion with density ρ and elastic constants c_{ijkl} is embedded in an infinite isotropic-elastic background with properties $\tilde{\rho}$ and $\tilde{c}_{ijkl}$, the differences between the perturbed and unperturbed model properties are defined as (Stolt and Weglein, 2012):

$$\begin{aligned}\Delta\rho&=\rho-\tilde{\rho},\\ \Delta c_{ijkl}&=c_{ijkl}-\tilde{c}_{ijkl},\end{aligned} \tag{F.2}$$

where $\Delta\rho$ and Δc_{ijkl} denote the density and elastic constant perturbations. Assuming that the size of the anisotropic obstacle is rather small compared to the wavelength of the incident wave, the perturbed wavefields corresponding to these model variations can be written as:

$$\Delta\mathbf{u}=\mathbf{u}-\tilde{\mathbf{u}}, \tag{F.3}$$

where $\tilde{\mathbf{u}}$ and $\Delta\mathbf{u}$ indicate the unperturbed wavefields and scattered wavefields respectively. Plugging equations (F.2) and (F.3) into equation (5.1) and ignoring the high order terms

based on the Born approximation, the equation of motion splits into two equations:

$$\frac{\partial}{\partial x_j}\left(\tilde{c}_{ijkl}\frac{\partial \tilde{u}_k}{\partial x_l}\right) - \tilde{\rho}\frac{\partial^2 \tilde{u}_i}{\partial t^2} = -f_i, \tag{F.4}$$

$$\frac{\partial}{\partial x_j}\left(\tilde{c}_{ijkl}\frac{\partial \Delta u_k}{\partial x_l}\right) - \tilde{\rho}\frac{\partial^2 \Delta u_i}{\partial t^2} = \Delta\rho\frac{\partial^2 \tilde{u}_i}{\partial t^2} - \frac{\partial \Delta M_{ij}}{\partial x_j}, \tag{F.5}$$

where $\Delta\mathbf{M}$ in equation (F.5) is the equivalent moment tensor source (Ben-Menahem and Singh, 1981; Chapman, 2004) and it indicates the perturbations of the elastic constants:

$$\Delta M_{ij} = \Delta c_{ijkl}\tilde{e}_{kl}, \tag{F.6}$$

where $\tilde{e}_{kl}$ are the strain components of the incident wave. First it is noticed that equation (F.4) is equivalent to equation (5.1), meaning that the unperturbed wavefield $\tilde{\mathbf{u}}$ propagates in the isotropic background media. Further examination reveals that equation (F.5) describes the propagation of the scattered wavefield $\Delta\mathbf{u}$ in the isotropic background media. The right-hand side of the equation (F.5) is referred to as "scattered sources". It underlines the fact that the scattered wavefields due to the perturbations of the model parameters (e.g $\Delta\rho$ or Δc_{ijkl}), can be interpreted as the wavefields generated by a set of secondary body forces, which propagate in the current (or unperturbed) medium (Dietrich and Kormendi, 1990).

According to equation (F.1), the solution of equation (F.5) can be written as an integral formulation in the frequency domain:

$$\begin{aligned}\Delta\bar{u}_n(\mathbf{x},\omega) = &\int_{\Omega(\mathbf{x}')}\int_{\omega'} \Delta\rho\omega^2\tilde{u}_i\tilde{G}_{ni}(\mathbf{x},\omega;\mathbf{x}',\omega')d\Omega(\mathbf{x}')d\omega' \\ &+ \int_{\Omega(\mathbf{x}')}\int_{\omega'} \frac{\partial \Delta M_{ij}}{\partial x'_j}\tilde{G}_{ni}(\mathbf{x},\omega;\mathbf{x}',\omega')d\Omega(\mathbf{x}')d\omega',\end{aligned} \tag{F.7}$$

where $\tilde{G}_{ij}(\mathbf{x},\omega;\mathbf{x}',\omega')$ indicates the Green's tensor in the unperturbed background medium due to the scattered source at position $\mathbf{x}' = (x', y', z')$. Ignoring the contribution from density (Here, only the perturbations of the elastic constants are considered) and applying integration by parts, the scattered wavefields can be obtained as:

$$\Delta\bar{u}_n(\mathbf{x},\omega) \approx -\int_{\Omega(\mathbf{x}')}\int_{\omega'} \delta M_{ij}\frac{\partial \tilde{G}_{ni}(\mathbf{x},\omega;\mathbf{x}',\omega')}{\partial x'_j}d\Omega(\mathbf{x}')d\omega'. \tag{F.8}$$

or a more compact form:

$$\Delta\bar{u}_n \approx -\Delta M_{ij}\tilde{G}_{ni,j}. \tag{F.9}$$

Taking the partial derivative of the scattered wavefields with respect to the variations of the model parameters yields the Fréchet derivative wavefield:

$$\nabla_{\mathbf{m}}\bar{u}_n\left(\mathbf{x},\omega\right)=-\int_{\Omega(\mathbf{x}')}\int_{\omega'}\nabla_{\mathbf{m}}M_{ij}\frac{\partial\tilde{G}_{ni}(\mathbf{x},\omega;\mathbf{x}',\omega')}{\partial x'_j}d\Omega(\mathbf{x}')d\omega', \tag{F.10}$$

where $\mathbf{m}$ denotes elastic constants c_{ijkl} in general anisotropic medium. Equation (F.10) is known as the Fréchet derivative (or inversion sensitivity kernel) which is widely analyzed and utilized in the linearized inversion framework (Tarantola, 1984, 1986; Pratt et al., 1998b; Virieux and Operto, 2009).

Applying Voigt recipe of indexes to the elastic constants perturbation matrix $\Delta\mathbf{c}$, equation (F.6) can be written in a matrix form:

$$\begin{bmatrix}\Delta M_{11}\\ \Delta M_{22}\\ \Delta M_{33}\\ \Delta M_{23}\\ \Delta M_{13}\\ \Delta M_{12}\end{bmatrix}=\begin{bmatrix}\Delta c_{11} & \Delta c_{12} & \Delta c_{13} & \Delta c_{14} & \Delta c_{15} & \Delta c_{16}\\ & \Delta c_{22} & \Delta c_{23} & \Delta c_{24} & \Delta c_{25} & \Delta c_{26}\\ & & \Delta c_{33} & \Delta c_{34} & \Delta c_{35} & \Delta c_{36}\\ & & & \Delta c_{44} & \Delta c_{45} & \Delta c_{46}\\ & & & & \Delta c_{55} & \Delta c_{56}\\ & & & & & \Delta c_{66}\end{bmatrix}\begin{bmatrix}\tilde{e}_{11}\\ \tilde{e}_{22}\\ \tilde{e}_{33}\\ 2\tilde{e}_{23}\\ 2\tilde{e}_{13}\\ 2\tilde{e}_{12}\end{bmatrix}. \tag{F.11}$$

Thus, the information of incident wave is encoded in the equivalent moment tensor source $\Delta\mathbf{M}$. First, an incident plane P-wave is considered:

$$\grave{\mathbf{u}}_{\mathrm{P}}=U\mathbf{exp}\left[i(\omega t-\mathbf{k}_\alpha\cdot\mathbf{n})\right]\hat{\mathbf{p}}, \tag{F.12}$$

where U is the amplitude of the incident P-wave and $\mathbf{n}$ indicates the unit vector in Cartesian coordinates. $\mathbf{k}_\alpha$ is the P-wave wavenumber vector in Spherical coordinates and $\hat{\mathbf{p}}$ is the polarization vector indicating the positive direction of the particle motion:

$$\mathbf{k}_\alpha=k_\alpha\hat{\mathbf{q}}=k_\alpha\left(\sin\vartheta\cos\varphi\mathbf{x}+\sin\vartheta\sin\varphi\mathbf{y}+\cos\vartheta\mathbf{z}\right), \tag{F.13}$$

$$\hat{\mathbf{p}}=\sin\vartheta\cos\varphi\mathbf{x}+\sin\vartheta\sin\varphi\mathbf{y}+\cos\vartheta\mathbf{z}, \tag{F.14}$$

where ϑ is the inclination angle of incident wave, which departs from z axis and φ departing from x axis indicates the azimuth angle of the incident wave. $\hat{\mathbf{q}}$ is the unit vector within Spherical coordinates. Thus, the strain components can be obtained as:

$$\tilde{e}_{ij}=-ik_\alpha U\hat{p}_i\hat{q}_j\mathbf{exp}\left[i(\omega t-\mathbf{k}_\alpha\cdot\mathbf{n})\right]. \tag{F.15}$$

Inserting equation (F.15) and the far-field approximation of the Green's function tensor in 3D isotropic-elastic media (Aki and Richards, 2002; Chapman, 2004) into equation (F.10), the analytic expressions of the 3D Fréchet derivative wavefield with incident plane P-wave can be obtained:

$$\nabla_{\mathbf{m}}\mathbf{u}_{\mathrm{P}} = -\frac{\omega^2 U \mathbf{exp}(-ik_\xi r)}{4\pi\rho\alpha\xi^3 r}\left(\hat{\mathbf{g}}^\dagger \nabla_{\mathbf{m}}\mathbb{M}\hat{\mathbf{r}}\right), \tag{F.16}$$

where ξ can be α or β for P-wave velocity or S-wave velocity respectively. k_ξ can be k_α or k_β for P-wave wavenumber or S-wave wavenumber. The vector $\hat{\mathbf{g}}$ can be $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$ for scattered P-wave, SV-wave and SH-wave:

$$\begin{aligned}
\hat{\mathbf{r}} &= [\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta]^\dagger, \\
\hat{\boldsymbol{\theta}} &= [\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta]^\dagger, \\
\hat{\boldsymbol{\phi}} &= [-\sin\phi, \cos\phi, 0]^\dagger,
\end{aligned} \tag{F.17}$$

where θ indicates the inclination angle departing from z axis and ϕ indicates the azimuth angle departing from x axis for describing the scattered wave. $\Delta\mathbb{M}$ in equation (F.16) indicates reduced moment tensor source by taking the terms of $ik_\alpha U$ and $\mathbf{exp}\left[i\left(\omega t - \mathbf{k}_\alpha \cdot \mathbf{n}\right)\right]$ out of $\Delta\mathbf{M}$. In this research, the scattering coefficient (or scattering pattern) due to perturbation of model parameter $\mathbf{m}$ is defined as (Chapman, 2004):

$$\mathbb{R}_{\mathrm{P}}\left(\vartheta, \varphi, \theta, \phi\right) = \hat{\mathbf{g}}^\dagger \nabla_{\mathbf{m}}\mathbb{M}\hat{\mathbf{r}}, \tag{F.18}$$

where $\mathbb{R}_{\mathrm{P}}$ is associated with 4 angles, which are used to describe the incident wave and scattered wave. Similarly, for incident plane SV-wave:

$$\grave{\mathbf{u}}_{\mathrm{SV}} = U_{\mathrm{SV}}\mathbf{exp}\left[i(\omega t - \mathbf{k}_\beta \cdot \mathbf{n})\right]\hat{\mathbf{p}}^{\mathrm{SV}}, \tag{F.19}$$

where U_{SV} is the amplitude of the incident SV-wave, $\mathbf{k}_\beta = k_\beta\hat{\mathbf{q}}$ and $\hat{\mathbf{p}}^{\mathrm{SV}} = \cos\vartheta\cos\varphi\mathbf{x} + \cos\vartheta\sin\varphi\mathbf{y} - \sin\vartheta\mathbf{z}$. Its strain components can be expressed as:

$$\tilde{e}_{ij}^{\mathrm{SV}} = -\tfrac{1}{2}ik_\beta U_{\mathrm{SV}}\left(\hat{p}_i^{\mathrm{SV}}\hat{q}_j + \hat{p}_j^{\mathrm{SV}}\hat{q}_i\right)\mathbf{exp}\left[i(\omega t - \mathbf{k}_\beta \cdot \mathbf{n})\right]. \tag{F.20}$$

For incident plane SH-wave:

$$\grave{\mathbf{u}}_{\mathrm{SH}} = U_{\mathrm{SH}}\mathbf{exp}\left[i(\omega t - \mathbf{k}_\beta \cdot \mathbf{n})\right]\hat{\mathbf{p}}^{\mathrm{SH}}, \tag{F.21}$$

where U_{SH} is the amplitude of incident SH-wave, $\hat{\mathbf{p}}^{\mathrm{SH}} = -\sin\varphi\mathbf{x} + \cos\varphi\mathbf{y}$ and its strain components can be expressed as:

$$\tilde{e}_{ij}^{\mathrm{SH}} = -\tfrac{1}{2}ik_\beta U\left(\hat{p}_i^{\mathrm{SH}}\hat{q}_j + \hat{p}_j^{\mathrm{SH}}\hat{q}_i\right)\mathbf{exp}\left[i(\omega t - \mathbf{k}_\beta\cdot\mathbf{n})\right]. \tag{F.22}$$

Inserting the strain components in equations (F.20) and (F.22) into equation (F.11) and then equation (F.10), the scattered wavefields with incident plane SV-wave and SH-wave are given by:

$$\nabla_{\mathbf{m}}\mathbf{u}_{\mathrm{SV}} = -\frac{\omega^2 U_{\mathrm{SV}}\mathbf{exp}(-ik_\xi r)}{4\pi\rho\beta\xi^3 r}\left(\hat{\mathbf{g}}^\dagger\nabla_{\mathbf{m}}\mathbb{M}^{\mathrm{SV}}\hat{\mathbf{r}}\right), \tag{F.23}$$

$$\nabla_{\mathbf{m}}\mathbf{u}_{\mathrm{SH}} = -\frac{\omega^2 U_{\mathrm{SH}}\mathbf{exp}(-ik_\xi r)}{4\pi\rho\beta\xi^3 r}\left(\hat{\mathbf{g}}^\dagger\nabla_{\mathbf{m}}\mathbb{M}^{\mathrm{SH}}\hat{\mathbf{r}}\right). \tag{F.24}$$

The scattering patterns given in this research are consistent with the 3D scattering patterns of isotropic parameters by Wu and Aki (1985).

In equations (5.5), (5.6) and (5.7), I give the explicit expressions of P-P, P-SV and P-SH scattering patterns due to Δc_{55} in HTI media. In this appendix, the explicit expressions of the 3D scattering patterns of the 5 elastic constants in HTI media with incident P-wave, SV-wave and SH-wave will be given. The P-P, P-SV and P-SH scattering patterns due to Δc_{33} are given in equations (F.25), (F.26) and (F.27) and plotted in Figure F.1.

$$\begin{aligned}\check{\mathbb{R}}_{\text{P-P}}(\vartheta,\varphi,\theta,\phi,\Delta c_{33}) &= \hat{r}_2^2\left(\hat{p}_2\hat{q}_2 + \hat{p}_3\hat{q}_3\right) + \hat{r}_3^2\left(\hat{p}_2\hat{q}_2 + \hat{p}_3\hat{q}_3\right)\\ &= \left(\sin^2\theta\sin^2\phi + \cos^2\theta\right)\left(\sin^2\vartheta\sin^2\varphi + \cos^2\varphi\right),\end{aligned} \tag{F.25}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{P-SV}}(\vartheta,\varphi,\theta,\phi,\Delta c_{33}) &= \hat{r}_2\hat{\theta}_2\left(\hat{p}_2\hat{q}_2 + \hat{p}_3\hat{q}_3\right) + \hat{r}_3\hat{\theta}_3\left(\hat{p}_2\hat{q}_2 + \hat{p}_3\hat{q}_3\right)\\ &= -0.5\sin 2\theta\cos^2\phi\left(\sin^2\vartheta\sin^2\varphi + \cos^2\varphi\right),\end{aligned} \tag{F.26}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{P-SH}}(\vartheta,\varphi,\theta,\phi,\Delta c_{33}) &= \hat{r}_2\hat{\phi}_2\left(\hat{p}_2\hat{q}_2 + \hat{p}_3\hat{q}_3\right) + \hat{r}_3\hat{\phi}_3\left(\hat{p}_2\hat{q}_2 + \hat{p}_3\hat{q}_3\right)\\ &= 0.5\sin\theta\sin 2\phi\left(\sin^2\vartheta\sin^2\varphi + \cos^2\varphi\right),\end{aligned} \tag{F.27}$$

The P-P, P-SV and P-SH scattering patterns due to Δc_{44} are given in equations (F.28), (F.29) and (F.30) and plotted in Figure F.2.

$$\begin{aligned}\check{\mathbb{R}}_{\text{P-P}}(\vartheta,\varphi,\theta,\phi,\Delta c_{44}) &= -2\hat{r}_2^2\hat{p}_3\hat{q}_3 + 4\hat{r}_2\hat{r}_3\hat{p}_2\hat{q}_3 - 2\hat{r}_3^2\hat{p}_3\hat{q}_3\\ &= -2\left(\sin^2\theta\sin^2\phi + \cos^2\theta\right)\cos^2\vartheta + \sin 2\theta\sin\phi\sin\vartheta\sin 2\varphi,\end{aligned} \tag{F.28}$$

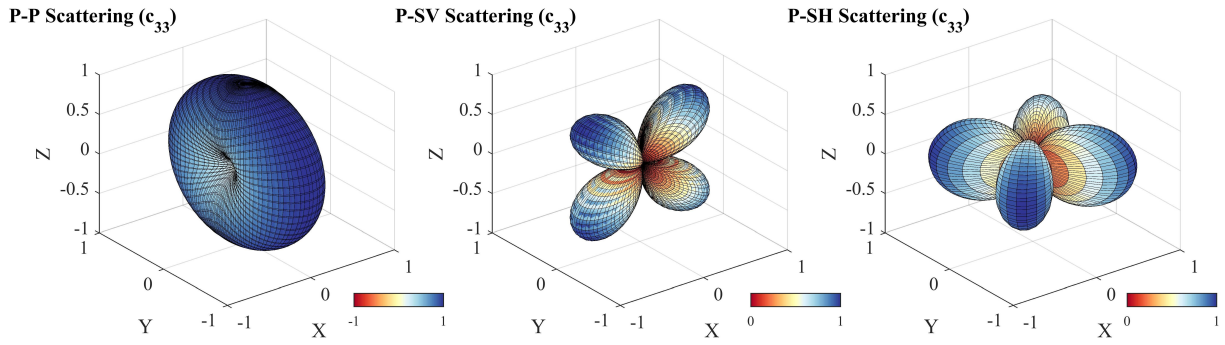

Figure F.1: The left, middle and right figures show the P-P (equation (F.25)), P-SV (equation (F.26)) and P-SH (equation (F.27)) scattering patterns due to the perturbation of c_{33} ($\vartheta = 135^{\circ}$ and $\varphi = 0^{\circ}$).

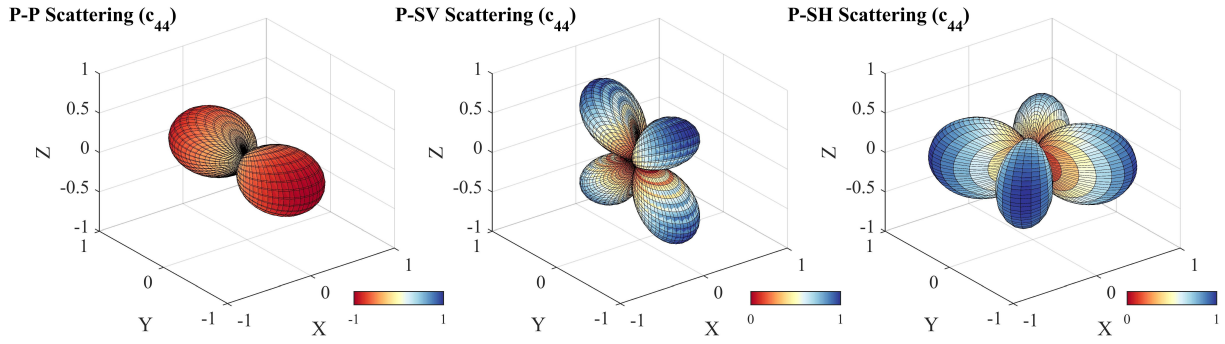

Figure F.2: The left, middle and right figures show the P-P (equation (F.28)), P-SV (equation (F.29)) and P-SH (equation (F.30)) scattering patterns due to the perturbation of c_{44} ($\vartheta = 135^{\circ}$ and $\varphi = 0^{\circ}$).

$$\begin{aligned}\check{\mathbb{R}}_{\text{P-SV}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{44}\right) &= -2\hat{r}_2\hat{\theta}_2\hat{p}_3\hat{q}_3 + 2\hat{r}_2\hat{\theta}_3\hat{p}_2\hat{q}_3 + 2\hat{r}_3\hat{\theta}_2\hat{p}_2\hat{q}_3 - 2\hat{r}_3\hat{\theta}_3\hat{p}_3\hat{q}_3 \\ &= \sin 2\theta\cos^2\varphi\cos^2\vartheta + \sin\phi\cos 2\theta\sin 2\vartheta\sin\varphi.\end{aligned} \tag{F.29}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{P-SH}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{44}\right) &= -2\hat{r}_2\hat{\phi}_2\hat{p}_3\hat{q}_3 + 2\hat{r}_2\hat{\phi}_3\hat{p}_2\hat{q}_3 + 2\hat{r}_3\hat{\phi}_2\hat{p}_2\hat{q}_3 - 2\hat{r}_3\hat{\phi}_3\hat{p}_3\hat{q}_3 \\ &= -\sin\theta\sin 2\phi\cos^2\vartheta + \cos\theta\cos\phi\sin 2\vartheta\sin\varphi.\end{aligned} \tag{F.30}$$

The P-P, P-SV and P-SH scattering patterns due to Δc_{11} are given in equations (F.31), (F.32) and (F.33) and plotted in Figure F.3.

$$\check{\mathbb{R}}_{\text{P-P}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{11}\right) = \hat{r}_1^2\hat{p}_1\hat{q}_1 = \sin^2\theta\cos^2\phi\sin^2\vartheta\cos^2\varphi, \tag{F.31}$$

$$\check{\mathbb{R}}_{\text{P-SV}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{11}\right) = \hat{r}_1\hat{\theta}_1\hat{p}_1\hat{q}_1 = 0.5\sin 2\theta\cos^2\phi\sin^2\vartheta\cos^2\varphi, \tag{F.32}$$

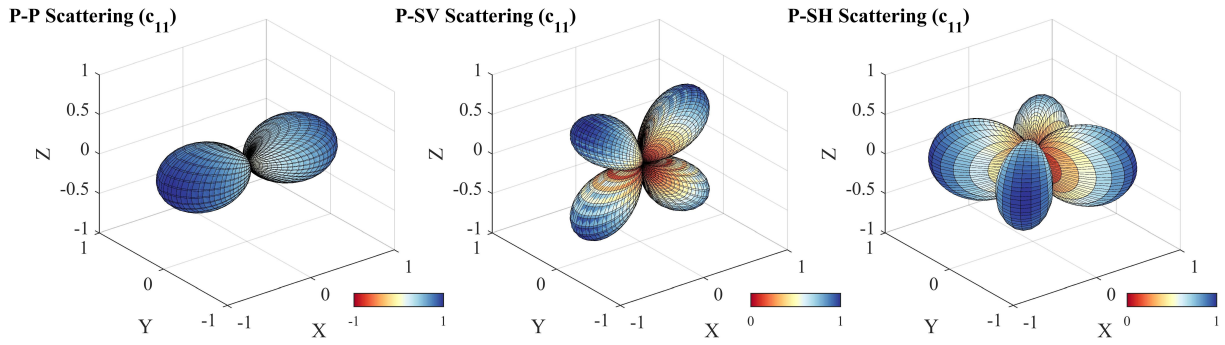

Figure F.3: The left, middle and right figures show the P-P (equation (F.31)), P-SV (equation (F.32)) and P-SH (equation (F.33)) scattering patterns due to the perturbation of c_{11} ($\vartheta = 135^{\circ}$ and $\varphi = 0^{\circ}$).

$$\check{\mathbb{R}}_{\text{P-SH}}\left(\vartheta, \varphi, \theta, \phi, \Delta c_{11}\right) = \hat{r}_1 \hat{\phi}_1 \hat{p}_1 \hat{q}_1 = -\sin^2\theta \cos\phi \sin^2\vartheta \cos^2\varphi, \tag{F.33}$$

The P-P, P-SV and P-SH scattering patterns due to Δc_{13} are given in equations (F.34), (F.35) and (F.36) and plotted in Figure F.4.

$$\begin{aligned}\check{\mathbb{R}}_{\text{P-P}}\left(\vartheta, \varphi, \theta, \phi, \Delta c_{13}\right) &= \hat{r}_1^2\left(\hat{p}_2\hat{q}_2 + \hat{p}_3\hat{q}_3\right) + \hat{r}_2^2 \hat{p}_1\hat{q}_1 + \hat{r}_3^2\hat{p}_1\hat{q}_1 \\ &= \sin^2\theta\cos^2\phi\left(\sin^2\vartheta\sin^2\varphi + \cos^2\vartheta\right) \\ &+ \left(\sin^2\theta\sin^2\phi + \cos^2\theta\right)\sin^2\vartheta\cos^2\varphi,\end{aligned} \tag{F.34}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{P-SV}}\left(\vartheta, \varphi, \theta, \phi, \Delta c_{13}\right) &= \hat{r}_1\hat{\theta}_1\left(\hat{p}_2\hat{q}_2 + \hat{p}_3\hat{q}_3\right) + \hat{r}_2\hat{\theta}_2\hat{p}_1\hat{q}_1 + \hat{r}_3\hat{\theta}_3\hat{p}_1\hat{q}_1 \\ &= 0.5\sin 2\theta\cos^2\phi\left(\sin^2\vartheta\sin^2\varphi + \cos^2\vartheta\right) \\ &- 0.5\sin 2\theta\cos^2\phi\sin^2\vartheta\cos^2\varphi,\end{aligned} \tag{F.35}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{P-SH}}\left(\vartheta, \varphi, \theta, \phi, \Delta c_{13}\right) &= \hat{r}_1\hat{\phi}_1\left(\hat{p}_2\hat{q}_2 + \hat{p}_3\hat{q}_3\right) + \hat{r}_2\hat{\phi}_2\hat{p}_1\hat{q}_1 + \hat{r}_3\hat{\phi}_3\hat{p}_1\hat{q}_1 \\ &= 0.5\sin\theta\sin 2\phi\left(\sin^2\vartheta\cos 2\varphi - \cos^2\vartheta\right),\end{aligned} \tag{F.36}$$

The SV-P, SV-SV and SV-SH scattering patterns due to Δc_{33} are given in equations (F.37), (F.38) and (F.39) and plotted in Figure F.5.

$$\begin{aligned}\check{\mathbb{R}}_{\text{SV-P}}\left(\vartheta, \varphi, \theta, \phi, \Delta c_{33}\right) &= \hat{r}_2^2\left(\hat{p}_2^{\text{SH}}\hat{q}_2 + \hat{p}_3^{\text{SV}}\hat{q}_3\right) + \hat{r}_3^2\left(\hat{p}_2^{\text{SV}}\hat{q}_2 + \hat{p}_3^{\text{SV}}\hat{q}_3\right) \\ &= \left(\sin^2\theta\sin^2\phi + \cos^2\theta\right)\left(0.5\sin 2\vartheta\sin^2\varphi + \sin^2\vartheta\right),\end{aligned} \tag{F.37}$$

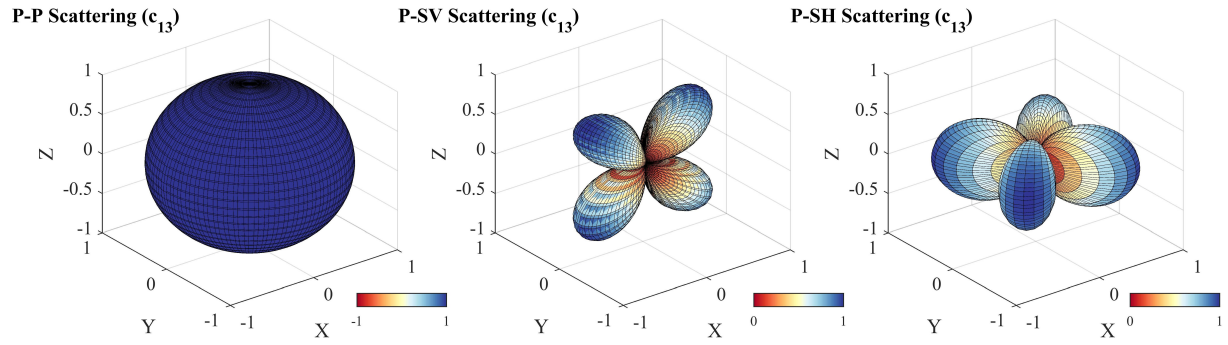

Figure F.4: The left, middle and right figures show the P-P (equation (F.34)), P-SV (equation (F.35)) and P-SH (equation (F.36)) scattering patterns due to the perturbation of c_{13} ($\vartheta = 135^{\circ}$ and $\varphi = 0^{\circ}$).

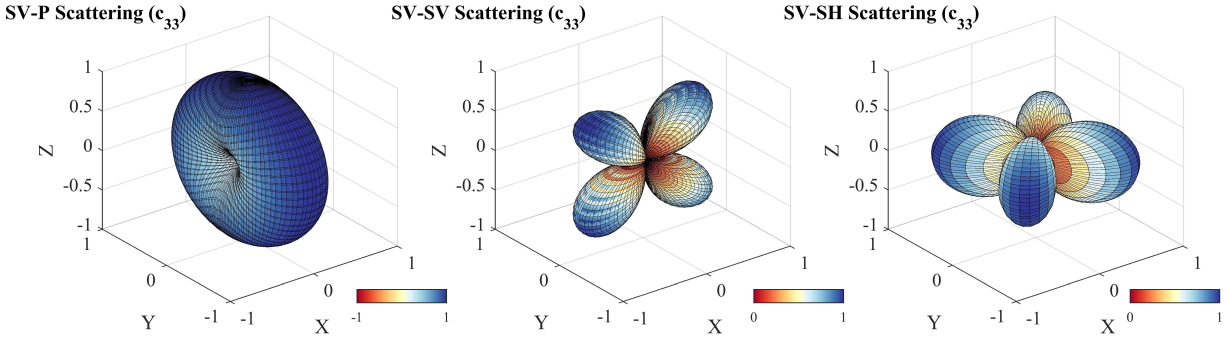

Figure F.5: The left, middle and right figures show the SV-P (equation (F.37)), SV-SV (equation (F.38)) and SV-SH (equation (F.39)) scattering patterns due to the perturbation of c_{33} ($\vartheta = 135^{\circ}$ and $\varphi = 0^{\circ}$).

$$\begin{aligned}\check{\mathbb{R}}_{\text{SV-SV}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{33}\right) &= \hat{r}_2\hat{\theta}_2\left(\hat{p}_2^{\text{SV}}\hat{q}_2 + \hat{p}_3^{\text{SV}}\hat{q}_3\right) + \hat{r}_3\hat{\theta}_3\left(\hat{p}_2^{\text{SV}}\hat{q}_2 + \hat{p}_3^{\text{SV}}\hat{q}_3\right) \\ &= -0.5\sin 2\theta\cos^2\phi\left(0.5\sin 2\vartheta\sin^2\varphi + \sin^2\vartheta\right),\end{aligned} \tag{F.38}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{SV-SH}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{33}\right) &= \hat{r}_2\hat{\phi}_2\left(\hat{p}_2^{\text{SV}}\hat{q}_2 + \hat{p}_3^{\text{SV}}\hat{q}_3\right) + \hat{r}_3\hat{\phi}_3\left(\hat{p}_2^{\text{SV}}\hat{q}_2 + \hat{p}_3^{\text{SV}}\hat{q}_3\right) \\ &= 0.5\sin\theta\sin 2\phi\left(0.5\sin 2\vartheta\sin^2\varphi + \sin^2\vartheta\right),\end{aligned} \tag{F.39}$$

The SV-P, SV-SV and SV-SH scattering patterns due to Δc_{55} are given in equations (F.40), (F.41) and (F.42) and plotted in Figure F.6.

$$\begin{aligned}\check{\mathbb{R}}_{\text{SV-P}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{55}\right) &= 2\hat{r}_1\hat{r}_2\left(\hat{p}_1^{\text{SV}}\hat{q}_2 + \hat{p}_2^{\text{SV}}\hat{q}_1\right) + 2\hat{r}_1\hat{r}_3\left(\hat{p}_1^{\text{SV}}\hat{q}_3 + \hat{p}_3^{\text{SV}}\hat{q}_1\right) \\ &= 0.5\sin^2\theta\sin 2\phi\sin 2\varphi\sin 2\theta + \sin 2\theta\cos\phi\sin 2\vartheta\cos\varphi,\end{aligned} \tag{F.40}$$

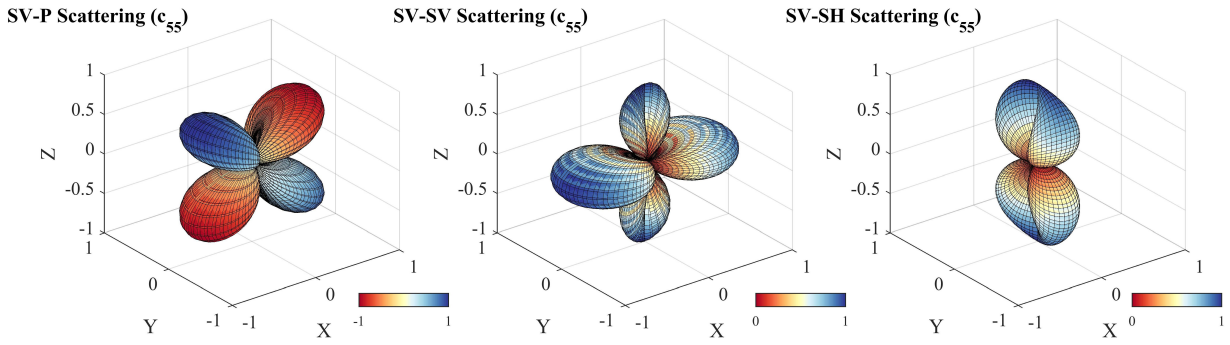

Figure F.6: The left, middle and right figures show the SV-P (equation (F.40)), SV-SV (equation (F.41)) and SV-SH (equation (F.42)) scattering patterns due to the perturbation of c_{55} ($\vartheta = 135^{\circ}$ and $\varphi = 0^{\circ}$).

$$\begin{aligned}\check{\mathbb{R}}_{\text{SV-SV}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{55}\right) &= \left(\hat{r}_1\hat{\theta}_2+\hat{r}_2\hat{\theta}_1\right)\left(\hat{p}_1^{\text{SV}}\hat{q}_2+\hat{p}_2^{\text{SV}}\hat{q}_1\right)+\left(\hat{r}_1\hat{\theta}_3+\hat{r}_3\hat{\theta}_1\right)\left(\hat{p}_1^{\text{SV}}\hat{q}_3+\hat{p}_3^{\text{SV}}\hat{q}_1\right)\\ &= 0.25\sin 2\theta\sin 2\phi\sin 2\varphi\sin 2\vartheta+\sin 2\theta\cos\phi\sin 2\vartheta\cos\varphi,\end{aligned} \tag{F.41}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{SV-SH}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{55}\right) &= \left(\hat{r}_1\hat{\phi}_2+\hat{r}_2\hat{\phi}_1\right)\left(\hat{p}_1^{\text{SV}}\hat{q}_2+\hat{p}_2^{\text{SV}}\hat{q}_1\right)+\left(\hat{r}_1\hat{\phi}_3+\hat{r}_3\hat{\phi}_1\right)\left(\hat{p}_1^{\text{SV}}\hat{q}_3+\hat{p}_3^{\text{SV}}\hat{q}_1\right)\\ &= 0.5\sin\theta\sin 2\phi\sin 2\varphi\sin 2\vartheta-\cos\theta\sin\phi\sin 2\vartheta\cos\varphi,\end{aligned} \tag{F.42}$$

The SV-P, SV-SV and SV-SH scattering patterns due to Δc_{44} are given in equations (F.43), (F.44) and (F.45) and plotted in Figure F.7.

$$\begin{aligned}\check{\mathbb{R}}_{\text{SV-P}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{44}\right) &= -2\hat{r}_2^2\hat{p}_3^{\text{SV}}\hat{q}_3+2\hat{r}_2\hat{r}_3\left(\hat{p}_2^{\text{SV}}\hat{q}_3+\hat{p}_3^{\text{SV}}\hat{q}_2\right)-2\hat{r}_3^2\hat{p}_3^{\text{SV}}\hat{q}_3\\ &= \left(\sin^2\theta\sin^2\phi+\cos^2\theta\right)\sin 2\vartheta+\sin 2\theta\sin\phi\sin 2\vartheta\sin\varphi,\end{aligned} \tag{F.43}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{SV-SV}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{44}\right) &= -2\left(\hat{r}_2\hat{\theta}_2+\hat{r}_3\hat{\theta}_3\right)\hat{p}_3^{\text{SV}}\hat{q}_3+\left(\hat{r}_2\hat{\theta}_3+\hat{r}_3\hat{\theta}_2\right)\left(\hat{p}_2^{\text{SV}}\hat{q}_3+\hat{p}_3^{\text{SV}}\hat{q}_2\right)\\ &= \left(0.25\sin 2\theta\sin 2\phi-0.5\sin 2\theta\right)\sin 2\vartheta+\sin 2\theta\sin\phi\sin 2\vartheta\sin\varphi,\end{aligned} \tag{F.44}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{SV-SH}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{44}\right) &= -2\left(\hat{r}_2\hat{\phi}_2+\hat{r}_3\hat{\phi}_3\right)\hat{p}_3^{\text{SV}}\hat{q}_3+\left(\hat{r}_2\hat{\phi}_3+\hat{r}_3\hat{\phi}_2\right)\left(\hat{p}_2^{\text{SV}}\hat{q}_3+\hat{p}_3^{\text{SV}}\hat{q}_2\right)\\ &= 0.5\sin\theta\sin 2\phi\sin 2\theta-\sin\theta\cos\phi\sin 2\vartheta\sin\varphi.\end{aligned} \tag{F.45}$$

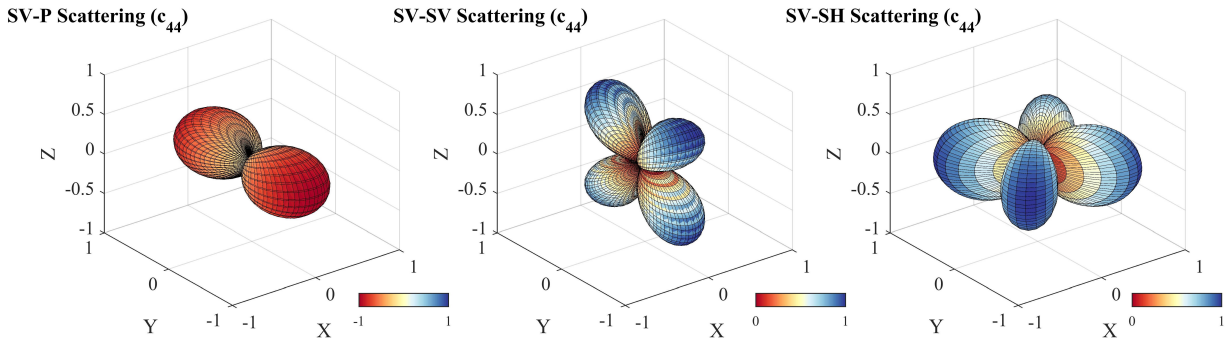

Figure F.7: The left, middle and right figures show the SV-P (equation (F.43)), SV-SV (equation (F.44)) and SV-SH (equation (F.45)) scattering patterns due to the perturbation of c_{44} ($\vartheta = 135^{\circ}$ and $\varphi = 0^{\circ}$).

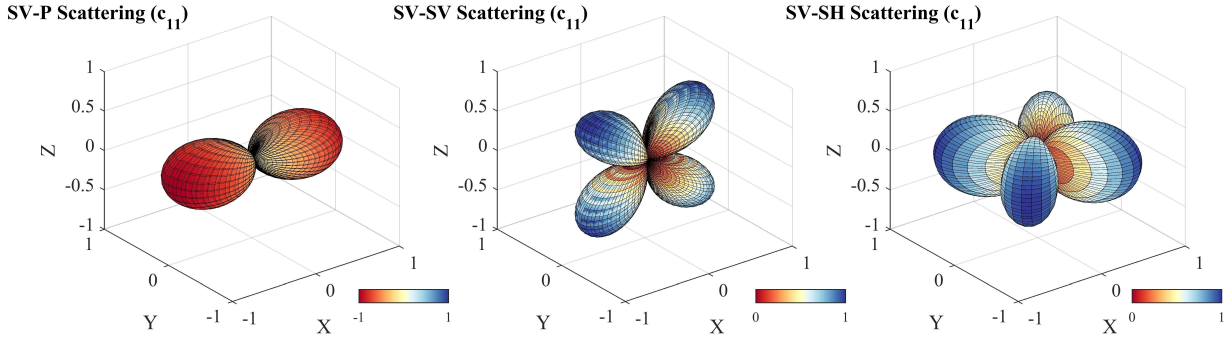

Figure F.8: The left, middle and right figures show the SV-P (equation (F.46)), SV-SV (equation (F.47)) and SV-SH (equation (F.48)) scattering patterns due to the perturbation of c_{11} ($\vartheta = 135^{\circ}$ and $\varphi = 0^{\circ}$).

The SV-P, SV-SV and SV-SH scattering patterns due to Δc_{11} are given in equations (F.46), (F.47) and (F.48) and plotted in Figure F.8.

$$\check{\mathbb{R}}_{\text{SV-P}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{11}\right) = \hat{r}_1^2 \hat{p}_1^{\text{SV}} \hat{q}_1 = 0.5 \sin^2\theta \cos^2\phi \sin 2\vartheta \cos^2\varphi, \tag{F.46}$$

$$\check{\mathbb{R}}_{\text{SV-SV}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{11}\right) = \hat{r}_1 \hat{\theta}_1 \hat{p}_1^{\text{SV}} \hat{q}_1 = 0.25 \sin 2\theta \cos^2\phi \sin 2\vartheta \cos^2\varphi, \tag{F.47}$$

$$\check{\mathbb{R}}_{\text{SV-SH}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{11}\right) = \hat{r}_1 \hat{\phi}_1 \hat{p}_1^{\text{SV}} \hat{q}_1 = -0.25 \sin\theta \sin 2\phi \sin 2\vartheta \cos^2\varphi, \tag{F.48}$$

The SV-P, SV-SV and SV-SH scattering patterns due to Δc_{13} are given in equations

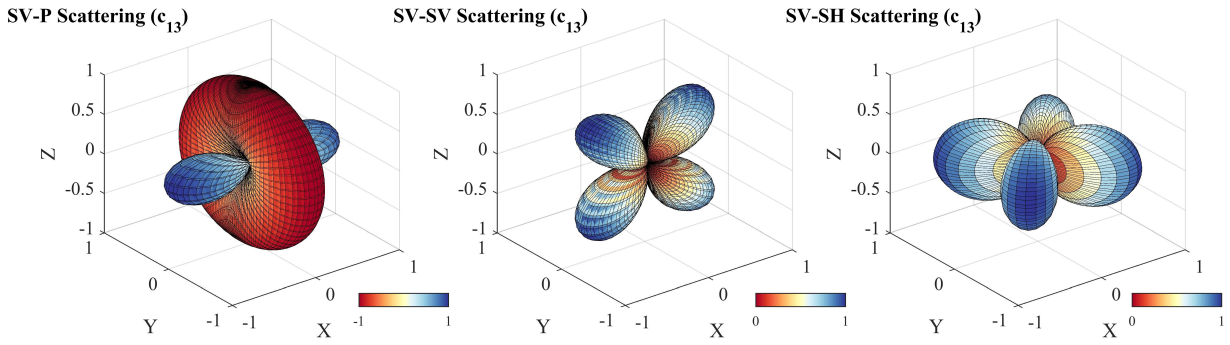

Figure F.9: The left, middle and right figures show the SV-P (equation (F.49)), SV-SV (equation (F.50)) and SV-SH (equation (F.51)) scattering patterns due to the perturbation of c_{13} ($\vartheta = 135^{\circ}$ and $\varphi = 0^{\circ}$).

(F.49), (F.50) and (F.51) and plotted in Figure F.9.

$$\begin{aligned}\check{\mathbb{R}}_{\text{SV-P}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{13}\right) &= \hat{r}_1^2\left(\hat{p}_2^{\text{SV}}\hat{q}_2+\hat{p}_3^{\text{SV}}\hat{q}_3\right)+\hat{r}_2^2\hat{p}_1^{\text{SV}}\hat{q}_1+\hat{r}_3^2\hat{p}_1^{\text{SV}}\hat{q}_1\\ &= -0.5\sin^2\theta\cos^2\phi\sin 2\phi\cos^2\phi\\ &\quad +0.5\left(\sin^2\theta\sin^2\phi+\cos^2\theta\right)\sin 2\vartheta\cos^2\varphi,\end{aligned} \tag{F.49}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{SV-SV}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{13}\right) &= \hat{r}_1\hat{\theta}_1\left(\hat{p}_2^{\text{SV}}\hat{q}_2+\hat{p}_3^{\text{SV}}\hat{q}_3\right)+\hat{r}_2\hat{\theta}_2\hat{p}_1^{\text{SV}}\hat{q}_1+\hat{r}_3\hat{\theta}_3\hat{p}_1^{\text{SV}}\hat{q}_1\\ &= -0.25\sin 2\theta\cos^2\phi\sin 2\phi\cos^2\phi\\ &\quad -0.5\cos^2\phi\sin 2\theta\sin 2\vartheta\cos^2\varphi,\end{aligned} \tag{F.50}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{SV-SH}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{13}\right) &= \hat{r}_1\hat{\phi}_1\left(\hat{p}_2^{\text{SV}}\hat{q}_2+\hat{p}_3^{\text{SV}}\hat{q}_3\right)+\hat{r}_2\hat{\phi}_2\hat{p}_1^{\text{SV}}\hat{q}_1+\hat{r}_3\hat{\phi}_3\hat{p}_1^{\text{SV}}\hat{q}_1\\ &= 0.25\sin\theta\sin 2\phi\cos^2\varphi\left(\sin 2\varphi-\sin 2\vartheta\right),\end{aligned} \tag{F.51}$$

The SH-P, SH-SV and SH-SH scattering patterns due to Δc_{33} are given in equations (F.52), (F.53) and (F.54) and plotted in Figure F.10.

$$\begin{aligned}\check{\mathbb{R}}_{\text{SH-P}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{33}\right) &= \hat{r}_2^2\left(\hat{p}_2^{\text{SH}}\hat{q}_2+\hat{p}_3^{\text{SH}}\hat{q}_3\right)+\hat{r}_3^2\left(\hat{p}_2^{\text{SH}}\hat{q}_2+\hat{p}_3^{\text{SH}}\hat{q}_3\right)\\ &= -\left(\sin^2\theta\sin^2\phi+\cos^2\theta\right)\sin^2\varphi\sin\vartheta,\end{aligned} \tag{F.52}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{SH-SV}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{33}\right) &= \hat{r}_2\hat{\theta}_2\left(\hat{p}_2^{\text{SH}}\hat{q}_2+\hat{p}_3^{\text{SH}}\hat{q}_3\right)+\hat{r}_3\hat{\theta}_3\left(\hat{p}_2^{\text{SH}}\hat{q}_2+\hat{p}_3^{\text{SH}}\hat{q}_3\right)\\ &= -0.25\sin 2\theta\cos^2\phi\sin 2\varphi\sin\vartheta,\end{aligned} \tag{F.53}$$

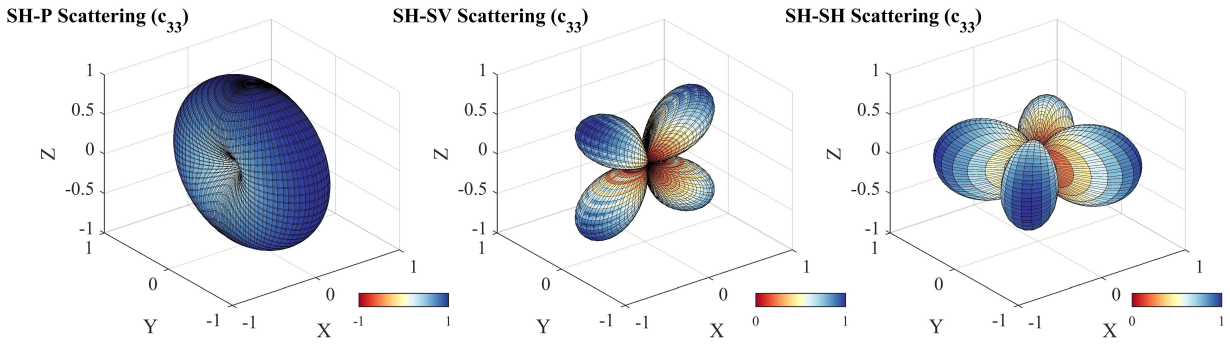

Figure F.10: The left, middle and right figures show the SH-P (equation (F.52)), SH-SV (equation (F.53)) and SV-SH (equation (F.54)) scattering patterns due to the perturbation of c_{33} ($\vartheta = 135^{\circ}$ and $\varphi = 30^{\circ}$).

$$\begin{aligned}\check{\mathbb{R}}_{\text{SH-SH}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{33}\right) &= \hat{r}_2\hat{\phi}_2\left(\hat{p}_2^{\text{SH}}\hat{q}_2 + \hat{p}_3^{\text{SH}}\hat{q}_3\right) + \hat{r}_3\hat{\phi}_3\left(\hat{p}_2^{\text{SH}}\hat{q}_2 + \hat{p}_3^{\text{SH}}\hat{q}_3\right) \\ &= 0.25\sin\theta\sin 2\phi\sin 2\varphi\sin\vartheta,\end{aligned} \tag{F.54}$$

The SH-P, SH-SV and SH-SH scattering patterns due to Δc_{55} are given in equations (F.55), (F.56) and (F.57) and plotted in Figure F.11.

$$\begin{aligned}\check{\mathbb{R}}_{\text{SH-P}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{55}\right) &= 2\hat{r}_1\hat{r}_2\left(\hat{p}_1^{\text{SH}}\hat{q}_2 + \hat{p}_2^{\text{SH}}\hat{q}_1\right) + 2\hat{r}_1\hat{r}_3\left(\hat{p}_1^{\text{SH}}\hat{q}_3 + \hat{p}_3^{\text{SH}}\hat{q}_1\right) \\ &= \sin^2\theta\sin 2\phi\sin 2\varphi\sin\vartheta - \sin 2\theta\cos\phi\sin\varphi\cos\vartheta,\end{aligned} \tag{F.55}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{SH-SV}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{55}\right) &= \left(\hat{r}_1\hat{\theta}_2 + \hat{r}_2\hat{\theta}_1\right)\left(\hat{p}_1^{\text{SH}}\hat{q}_2 + \hat{p}_2^{\text{SH}}\hat{q}_1\right) + \left(\hat{r}_1\hat{\theta}_3 + \hat{r}_3\hat{\theta}_1\right)\left(\hat{p}_1^{\text{SH}}\hat{q}_3 + \hat{p}_3^{\text{SH}}\hat{q}_1\right) \\ &= 0.5\sin 2\theta\sin 2\phi\sin 2\varphi\sin\vartheta - \sin 2\theta\cos\phi\sin\varphi\cos\vartheta,\end{aligned} \tag{F.56}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{SH-SH}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{55}\right) &= \left(\hat{r}_1\hat{\phi}_2 + \hat{r}_2\hat{\phi}_1\right)\left(\hat{p}_1^{\text{SH}}\hat{q}_2 + \hat{p}_2^{\text{SH}}\hat{q}_1\right) + \left(\hat{r}_1\hat{\phi}_3 + \hat{r}_3\hat{\phi}_1\right)\left(\hat{p}_1^{\text{SH}}\hat{q}_3 + \hat{p}_3^{\text{SH}}\hat{q}_1\right) \\ &= \sin\theta\sin 2\phi\sin 2\varphi\sin\vartheta + \cos\theta\sin\phi\cos\vartheta\cos\varphi,\end{aligned} \tag{F.57}$$

The SH-P, SH-SV and SH-SH scattering patterns due to Δc_{44} are given in equations (F.58), (F.59) and (F.60) and plotted in Figure F.12.

$$\begin{aligned}\check{\mathbb{R}}_{\text{SH-P}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{44}\right) &= -2\hat{r}_2^2\hat{p}_3^{\text{SH}}\hat{q}_3 + 2\hat{r}_2\hat{r}_3\left(\hat{p}_2^{\text{SH}}\hat{q}_3 + \hat{p}_3^{\text{SH}}\hat{q}_2\right) - 2\hat{r}_3^2\hat{p}_3^{\text{SH}}\hat{q}_3 \\ &= \sin\phi\sin 2\theta\cos\varphi\cos\vartheta,\end{aligned} \tag{F.58}$$

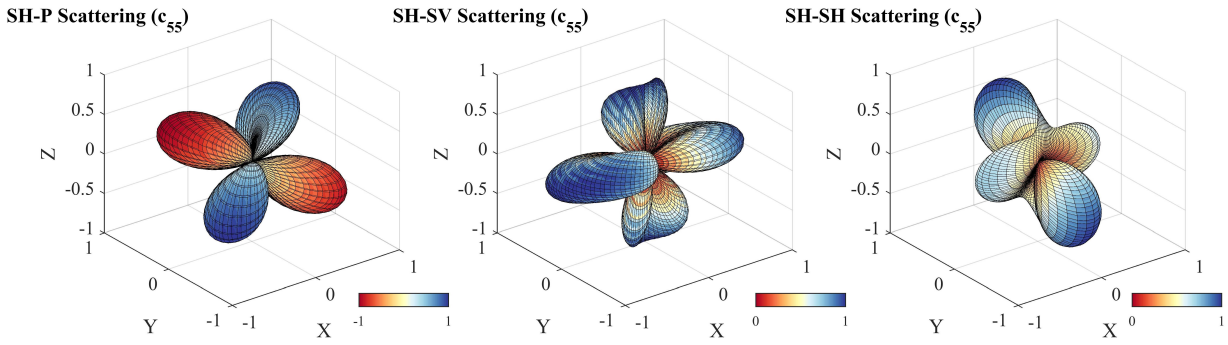

Figure F.11: The left, middle and right figures show the SH-P (equation (F.55)), SH-SV (equation (F.56)) and SV-SH (equation (F.57)) scattering patterns due to the perturbation of c_{55} ($\vartheta = 135^{\circ}$ and $\varphi = 30^{\circ}$).

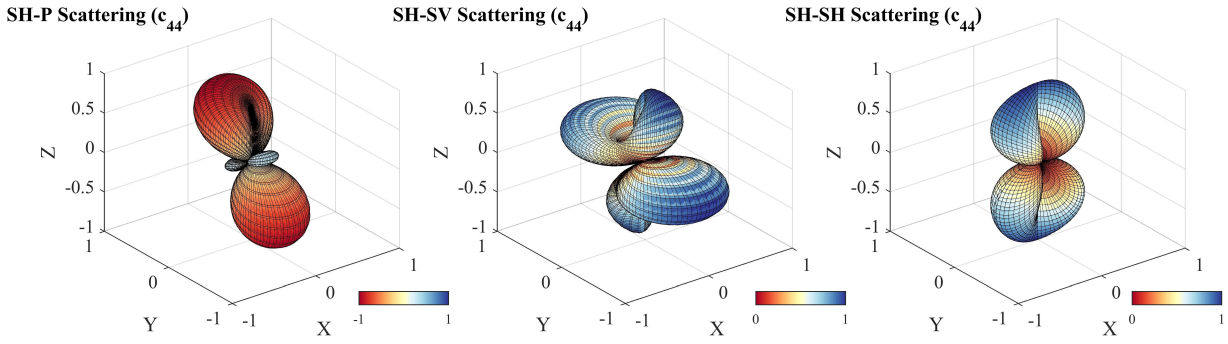

Figure F.12: The left, middle and right figures show the SH-P (equation (F.58)), SH-SV (equation (F.59)) and SV-SH (equation (F.60)) scattering patterns due to the perturbation of c_{44} ($\vartheta = 135^{\circ}$ and $\varphi = 30^{\circ}$).

$$\begin{aligned}\check{\mathbb{R}}_{\text{SH-SV}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{44}\right) &= -2\left(\hat{r}_2\hat{\theta}_2+\hat{r}_3\hat{\theta}_3\right)\hat{p}_3^{\text{SH}}\hat{q}_3+\left(\hat{r}_2\hat{\theta}_3+\hat{r}_3\hat{\theta}_2\right)\left(\hat{p}_2^{\text{SH}}\hat{q}_3+\hat{p}_3^{\text{SH}}\hat{q}_2\right)\\ &= 0.5\sin 2\theta\sin\phi\sin 2\varphi\sin\vartheta,\end{aligned} \tag{F.59}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{SH-SH}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{44}\right) &= -2\left(\hat{r}_2\hat{\phi}_2+\hat{r}_3\hat{\phi}_3\right)\hat{p}_3^{\text{SH}}\hat{q}_3+\left(\hat{r}_2\hat{\phi}_3+\hat{r}_3\hat{\phi}_2\right)\left(\hat{p}_2^{\text{SH}}\hat{q}_3+\hat{p}_3^{\text{SH}}\hat{q}_2\right)\\ &= 0.5\cos\theta\cos\phi\sin 2\varphi\sin\vartheta,\end{aligned} \tag{F.60}$$

The SH-P, SH-SV and SH-SH scattering patterns due to Δc_{11} are given in equations (F.61), (F.62) and (F.63) and plotted in Figure F.13.

$$\check{\mathbb{R}}_{\text{SH-P}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{11}\right) = \hat{r}_1^2\hat{p}_1^{\text{SH}}\hat{q}_1 = -0.5\sin^2\theta\cos^2\phi\sin 2\vartheta\sin\varphi, \tag{F.61}$$

$$\check{\mathbb{R}}_{\text{SH-SV}}\left(\vartheta,\varphi,\theta,\phi,\Delta c_{11}\right) = \hat{r}_1\hat{\theta}_1\hat{p}_1^{\text{SH}}\hat{q}_1 = -0.25\sin 2\theta\cos^2\phi\sin 2\vartheta\sin\varphi, \tag{F.62}$$

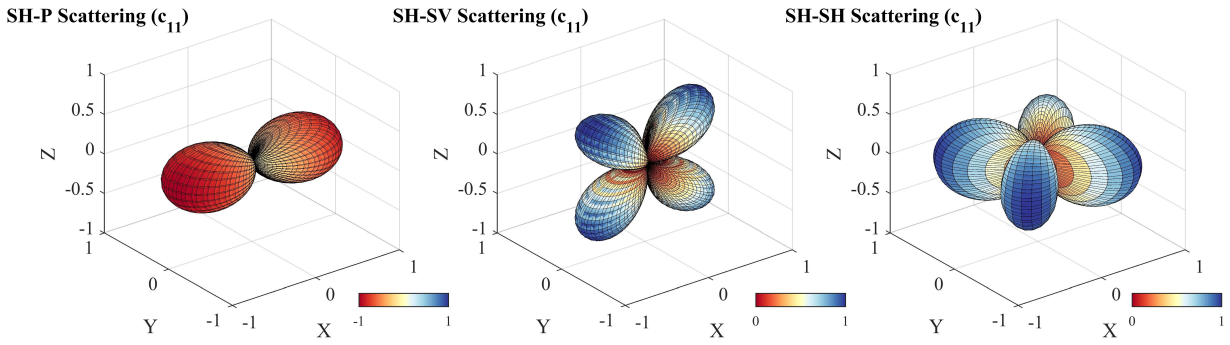

Figure F.13: The left, middle and right figures show the SH-P (equation (F.61)), SH-SV (equation (F.62)) and SV-SH (equation (F.63)) scattering patterns due to the perturbation of c_{11} ($\vartheta = 135^{\circ}$ and $\varphi = 30^{\circ}$).

$$\check{\mathbb{R}}_{\text{SH-SH}}\left(\vartheta, \varphi, \theta, \phi, \Delta c_{11}\right) = \hat{r}_1 \hat{\phi}_1 \hat{p}_1^{\text{SH}} \hat{q}_1 = 0.25 \sin\vartheta \sin 2\varphi, \tag{F.63}$$

The SH-P, SH-SV and SH-SH scattering patterns due to Δc_{13} are given in equations (F.64), (F.65) and (F.66) and plotted in Figure F.14.

$$\begin{aligned}\check{\mathbb{R}}_{\text{SH-P}}\left(\vartheta, \varphi, \theta, \phi, \Delta c_{13}\right) &= \hat{r}_1^2 \left(\hat{p}_2^{\text{SH}} \hat{q}_2 + \hat{p}_3^{\text{SH}} \hat{q}_3\right) + \hat{r}_2^2 \hat{p}_1^{\text{SH}} \hat{q}_1 + \hat{r}_3^2 \hat{p}_1^{\text{SH}} \hat{q}_1 \\ &= \sin^2\theta \cos^2\phi \sin^2\varphi \cos\varphi - 0.5 \left(\sin^2\theta \sin^2\phi + \cos^2\theta\right) \sin 2\vartheta \cos^2\varphi,\end{aligned} \tag{F.64}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{SH-SV}}\left(\vartheta, \varphi, \theta, \phi, \Delta c_{13}\right) &= \hat{r}_1 \hat{\theta}_1 \left(\hat{p}_2^{\text{SH}} \hat{q}_2 + \hat{p}_3^{\text{SH}} \hat{q}_3\right) + \hat{r}_2 \hat{\theta}_2 \hat{p}_1^{\text{SH}} \hat{q}_1 + \hat{r}_3 \hat{\theta}_3 \hat{p}_1^{\text{SH}} \hat{q}_1 \\ &= 0.5 \sin 2\theta \cos^2\phi \cos\varphi \sin^2\vartheta + 0.25 \cos^2\phi \sin 2\theta \sin 2\varphi \sin\vartheta,\end{aligned} \tag{F.65}$$

$$\begin{aligned}\check{\mathbb{R}}_{\text{SH-SH}}\left(\vartheta, \varphi, \theta, \phi, \Delta c_{13}\right) &= \hat{r}_1 \hat{\phi}_1 \left(\hat{p}_2^{\text{SH}} \hat{q}_2 + \hat{p}_3^{\text{SH}} \hat{q}_3\right) + \hat{r}_2 \hat{\phi}_2 \hat{p}_1^{\text{SH}} \hat{q}_1 + \hat{r}_3 \hat{\phi}_3 \hat{p}_1^{\text{SH}} \hat{q}_1 \\ &= -0.5 \sin\theta \sin 2\phi \cos\varphi \sin^2\vartheta - 0.25 \sin\theta \sin 2\phi \sin 2\varphi \sin\vartheta,\end{aligned} \tag{F.66}$$

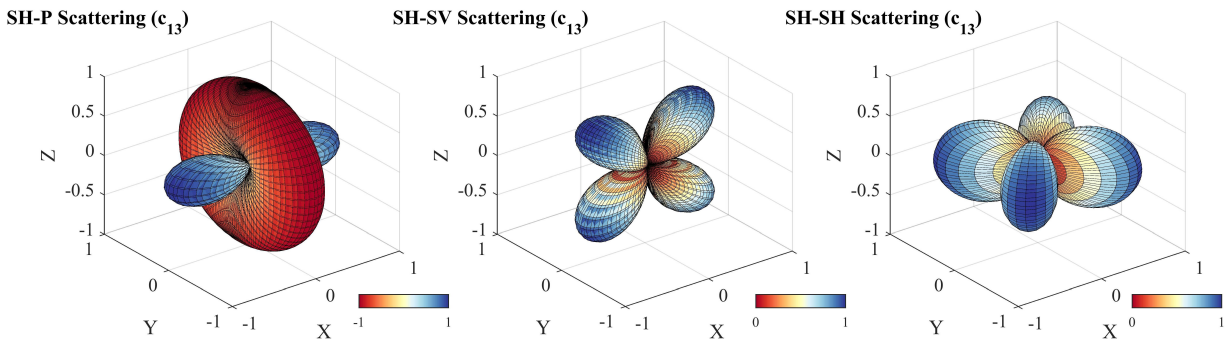

Figure F.14: The left, middle and right figures show the SH-P (equation (F.64)), SH-SV (equation (F.65)) and SV-SH (equation (F.66)) scattering patterns due to the perturbation of c_{13} ($\vartheta = 135^{\circ}$ and $\varphi = 30^{\circ}$).

Chapter 6

Conclusions and Future Studies

6.1 Conclusions

As described in the introduction section, full-waveform inversion (FWI) techniques are powerful and hold the promise of providing high-resolution estimations of subsurface properties. However, FWI applications often fail in practice because of a series of difficulties including extensive computation requirements, slow convergence rate, cycle-skipping, interparameter tradeoffs (or parameter crosstalk), etc. This thesis aims at overcoming these problems in FWI with phase-encoding strategies, advanced non-linear optimization methods and novel inversion strategies to reduce the parameter crosstalk artifacts and inverting subsurface elastic and anisotropic properties for reservoir characterization.

An efficient linear phase-encoding strategies in τ-p domain is described in chapter 2. It has been shown that both of the gradient and diagonal Hessian preconditioner can be constructed with this linear phase-encoding approach more efficiently. Numerical examples are given to show that compared to traditional shot-profile method, the phase-encoding approach can provide inverted model with comparable quality and reduce the computational cost at the same time. To improve the convergence rate of FWI, various non-linear optimization methods including steepest-descent, non-linear conjugate-gradient, l-BFGS and (preconditioned) Hessian-free Gauss-Newton methods have been examined. Compared to gradient-based methods, the second-order optimization methods are able to provide faster convergence rate. To accelerate the Hessian-free Gauss-Newton method, various preconditioning approaches for the inner conjugate-gradient algorithm are developed. It has been shown that the l-BFGS inverse Hessian approximation works as best preconditioner for Hessian-free Gauss-Newton FWI. The l-BFGS preconditioned Hessian-free Gauss-Newton method is also applied to reconstruct P-wave velocity and density simultaneously in multiparameter acoustic media. In chapter 4, the coupling effects of isotropic-elastic parameters are analyzed with scattering patterns and multiparameter Hessian-vector products for

isotropic-elastic FWI. I found that S-wave velocity produces strong contaminations into density updates which makes density difficult to be recovered. A novel inversion strategy with approximate contamination kernels is also proposed to reduce the influence of interparameter tradeoffs in the inversion process. This new inversion strategy is applied to Hussar real seismic dataset. Furthermore, coupling effects within various parameterizations are also investigated. In chapter 5, the Gauss-Newton and full-Newton FWI are investigated to invert elastic stiffness coefficients in HTI media. The 3D scattering patterns of elastic stiffness coefficients are first derived to analyze the coupling effects in HTI media. Furthermore, numerical examples are given to show that the multiparameter Hessian is able to remove the parameter crosstalk artifacts. The second-order preconditioner in multiparameter Hessian constructed with adjoint-state approach is also able to suppress the artifacts due to multiparameter second-order scattering effects.

6.2 Future Studies

In this thesis, various strategies have been proposed to overcome the difficulties in single parameter and especially in multiparameter FWI. These strategies can be shown to positively affect issues associated with the computational expense of FWI, and the types and character of model features constructed in both synthetic and field data environments. However, applying FWI techniques to realistic seismic dataset is still a very challenging task. I have designed plans and strategies for FWI applications in future studies.

In the thesis, only synthetic examples are given to show the benefits and advantages of phase-encoding approaches for monoparameter FWI. As further step in future studies, it is expected to reconstruct multiple physical parameters (i.e., S-wave velocity, density, attenuation, etc) using FWI with phase-encoding approaches. In other words, to merge the techniques created in chapters 2 and 4. Furthermore, the effectiveness of phase-encoding approaches should also be examined using real seismic dataset. Advanced second-order optimization methods (l-BFGS and Hessian-free methods) are applied for multiparameter acoustic FWI with velocity-density parameterization. Because different parameterizations have different performances for inverting subsurface parameters. It is necessary to apply ad-

vanced optimization methods to multiparameter acoustic FWI with different parameterizations (i.e., impedance-density parameterization) for comparison. Furthermore, the proposed l-BFGS preconditioned Hessian-free Gauss-Newton FWI can in principle be applied to the problem of inverting for isotropic- and anisotropic-elastic parameters.

Only synthetic examples of various parameterizations in isotropic-elastic FWI with nonlinear conjugate-gradient method are given in chapter 4. In future studies, I plan to apply isotropic-elastic FWI with various parameterizations to the Hussar field seismic dataset and examine the performances of different parameterizations with second-order optimization methods (i.e., preconditioned Hessian-free Gauss-Newton method). In chapter 5, Gauss-Newton and full-Newton methods are applied to invert the elastic constants in 2D HTI media. Because Gauss-Newton and full-Newton methods are both very expensive for constructing the Hessian matrix explicitly. One further research direction is to apply Hessian-free optimization methods for inverting parameters in HTI media, in which only Hessian-vector products are calculated. It is more meaningful to estimate the properties of 3D HTI media. Fractured media can be described with many different parameters (i.e., fracture spacing, fracture density, weakness, compliance, etc). Using full-waveform inversion methods to invert these properties of fractured media will be studied.

References

Akcelik, V., Biros, G., and Ghattas, O., 2002, Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation: Proceedings of the ACM/IEEE Conference on Supercomputing, 2357–2361.

Aki, K., and Richards, P. G., 2002, Quantitative Seismology: University Science Books, 2nd edn.

Alkhalifa, T., and Plessix, R., 2014, A recipe for practical full-waveform inversion in anisotropic media: An analytic parameter resolution study: Geophysics, **79**, R91–R101.

An, M., 2012, A simple method for determing the spatial resolution of a general inverse problem: Geophysical Journal International, **191**, 849–864.

Anagaw, A. Y., and Sacchi, M. D., 2012, Full waveform inversion with simultaneous sources using the full Newton method: SEG Expanded Abstracts, 971–975.

Anagaw, A. Y., and Sacchi, M. D., 2014, Comparison of multifrequency selection strategies for simultaneous-source full-waveform inversion: Geophysics, **79**, R165–R181.

ans S. Treitel, L. R. L., 1984, Tutorial: A review of least-squares inversion and its application to geophysical problems: Geophysical Prospecting, **32**, 159–186.

Armijo, L., 1966, Minimization of functions having Lipschitz continuous first partial derivatives: Pacific J. Math, **16**, 1–3.

Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P., and Virieux, J., 2013, Regularized seismic full-waveform inversion with prior model information: Geophysics, **78**, R25–R36.

Backus, G., and Gilbert, F., 1968, The resolving power of gross earth data: Geophysical Journal International, **16**, 169–205.

Baumstein, A., 2014, Extended subspace method for attenuation of crosstalk in multiparameter Full Wavefield Inversion: SEG Expanded Abstracts, 1121–1125.

Ben-Hadj-Ali, H., Operto, S., and Virieux, J., 2011, An efficient frequency-domain full waveform inversion method using simultaneous encoded sources: Geophysics, **76**, WCC177–WCC188.

Ben-Menahem, A., and Singh, S., 1981, Seismic waves and sources: New York, 2nd edn.

Berenger, J., 1994, A perfectly matched layer for the absorption of electromagnetic waves: Journal of Computational Physics, **114**, 185–200.

Berkhout, A. J., 2012, Blended acquisition with dispersed source arrays: Geophysics, **77**, A19–A23.

Bernauer, M., Fichtner, A., and Igel, H., 2014, Optimal observables for multiparameter seismic tomography: Geophysical Journal International, **198**, 1241–1254.

Boehm, C., and Ulbrich, M., 2014, A Semismooth Newton-CG Method for Constrained Parameter Identification in Seismic Tomography: SIAM J. Sci. Comput, **37**, S334–S364.

Boonyasiriwat, C., Valasek, P., Routh, P., Cao, W., Schuster, G. T., and Macy, B., 2009, An efficient multiscale method for time-domain waveform tomography: Geophysics, **74**, WCC59–WCC68.

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

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

Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N., and Pugmire, D., 2016, Global adjoint tomography: first-generation model: Geophysical Journal International, **207**, 1739–1766.

Brossier, R., Operto, S., and Virieux, J., 2009, Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion: Geophysics, **74**, WCC105–WCC118.

Brossier, R., Operto, S., and Virieux, J., 2010, Which data residual norm for robust elastic frequency-domain full-waveform inversion?: Geophysics, **75**, R37–R46.

Broyden, C. G., 1970, The convergence of a class of double-rank minimization algorithms: IMA Journal of Applied Mathematics, **6**, 222–231.

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

Byrd, R. H., Lu, P., and Nocedal, J., 1995, A limited memory algorithm for bound constrained optimization: SIAM Journal on Scientific and Statistical Computing, **16**, 1190–1208.

Castellanos, C., Métivier, L., Operto, S., Brossier, R., and Virieux, J., 2015, Fast full waveform inversion with source encoding and second-order optimization methods: Geophysical Journal International, **200**, 720–744.

Chapman, C., 2004, Fundamentals of Seismic Wave Propagation: Cambridge University Express.

Chen, B., and Xie, X., 2015, An efficient method for broadband seismic illumination and resolution analysises: SEG Expanded Abstracts, 4227–4231.

Chi, B., Dong, L., and Liu, Y., 2015, Correlation-based reflection full-waveform inversion: Geophysics, **80**, R189–R202.

Choi, Y., and Alkhalifah, T., 2012, Application of multi-source waveform inversion to marine streamer data using the global correlation norm: Geophysical Prospecting, **60**, 748–758.

Choi, Y., Min, D. J., and Shin, C., 2008, 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**, 2402–2415.

Crampin, S., 1984, An introduction to wave propagation in anisotropic media: Geophysical Journal International, **76**, 17–28.

Cui, T., 2015, Improving seismic-to-well ties: M.Sc. thesis, University of Calgary.

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

Davis, T. A., and Duff, I. S., 1997, An unsymmetric pattern multifrontal method for sparse lu factorization: SIAM Journal on Matrix Analysis and Applications, **18**, 140–158.

Demanet, L., Létourneau, P. D., Noumal, N., Calandra, H., and Chiu, J. S., 2012, Matrix probing: a randomized preconditioner for the wave-equation Hessian: Appl. Comput. Harmon. Anal, **32**, 155–168.

Dettmer, J., Dosso, S. E., and Holland, C. W., 2007, Uncertainty estimation in seismo-acoustic reflection travel-time inversion: J. Acoust. Soc. Am., **122**, 161–176.

Dietrich, M., and Kormendi, F., 1990, Perturbation of plane-wave reflectivity of a depth-dependent elastic medium by weak inhomogeneities: Geophysical Journal International, **100**, 203–214.

Eisenstat, S. C., and Walker, H. F., 1996, Choosing the forcing terms in an inexact Newton method: SIAM Journal on Scientific Computing, **17**, 16–32.

Epanomeritakis, I., Akçelik, V., Ghattas, O., and Bielak, J., 2008, A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion: Inverse Problems, **24**, 1–26.

Esmaeili, S., 2016, Influence of low frequencies on seismic impedance inversion: M.Sc. thesis, University of Calgary.

Esser, E., Guasch, L., Herrmann, F. J., and Warner, M., 2016, Constrained waveform inversion for automatic salt flooding: The Leading Edge, **35**, 235–239.

Etgen, J. T., 2005, How many angles do we really need for delayed-shot migration?: SEG Technical Program Expanded Abstracts, 1985–1988.

Fang, X., Fehler, M. C., Chen, T., Burns, D. R., and Zhu, Z., 2013, Sensiticity analysis of fracture scattering: Geophysics, **78**, T1–T10.

Fichtner, A., Bunge, H. P., and Igel, H., 2006, The adjoint method in seismology:I. Theory: Physics of the Earth and Planetary Interiors, **157**, 86–104.

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

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

Fichtner, A., and Trampert, J., 2011c, Resolution analysis in full waveform inversion: Geophysical Journal International, **187**, 1604–1642.

Fichtner, A., and van Leeuwen, T., 2015, Resolution analysis by random probing: Journal of Geophysical Research: Solid Earth, **120**, 5549–5573.

Flath, H. P., Wilcox, L. C., Akcelik, V., Hill, J., Waanders, B. V. B., and Ghattas, O., 2011, Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial Hessian approximations: SIAM Journal on Scientific Computing, **33**, 407–432.

Fletcher, R., 1970, A new approach to variable metric algorithms: The Computer Journal, **13**, 317–322.

Fletcher, R., and Reeves, C. M., 1964, Function minimization by conjugate gradients: Comp. J., **7**, 149–154.

Forgues, E., and Lambaré, G., 1997, Parameterization study for acoustic and elastic ray+Born inversion: Journal of Seismic Exploration, **6**, 253–278.

Gao, F., Atle, A., and Williamson, P., 2010, Full waveform inversion using deterministic source encoding: SEG Technical Program Expanded Abstracts, 1013–1017.

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

Geng, Y., Innanen, K. A., and Pan, W., 2017a, Frequency domain full-waveform inversion with nonlinear descent directions: Geophysical Journal International, *submited*.

Geng, Y., Pan, W., and Innanen, K. A., 2017b, Frequency-domain full-waveform inversion with updates based on nonlinear sensitivities: SEG Expanded Abstracts, 1330–1335.

Gholami, Y., Brossier, R., Operto, S., Prieus, V., Ribodetti, A., and Virieux, J., 2013a, Which parameterization is suitable for acoustic vertical transverse isotropic full waveform inversion? part 2: Synthetic and real data case studies from valhall: Geophysics, **78**, R107–R124.

Gholami, Y., Brossier, R., Operto, S., Ribodetti, A., and Virieux, J., 2013b, Which parametrization for acoustic VTI full waveform inversion? - part 1: sensitivity and trade-off analysis: Geophysics, **78**, R81–R105.

Godwin, J., and Sava, P., 2013, A comparison of shot-encoding schemes for wave-equation migration: Geophysical Prospecting, **61**, 391–408.

Goldfarb, D., 1970, A family of variable-metric methods derived by variational means: Mathematics of Computation, 23–26.

Gouveia, W., and Scales, J., 1998, Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis: Journal of Geophysical Research, **103**, 2759–2779.

Gray, S. H., 2013, Spatial sampling, migration aliasing, and migrated amplitudes: Geophysics, **78**, S157–S164.

Guitton, A., and Díaz, E., 2012, Attenuating crosstalk noise with simultaneous source full waveform inversion: Geophysical Prospecting, **60**, 759–768.

Halko, N., Martinsson, P., and Tropp, J., 2011, Finding structure with randomness: probabilistic algorithms for constructing approximation matrix decomposition: SIAM Review, **53**, 217–281.

Herrmann, F. J., Erlangga, Y. A., and Lin, T. T. Y., 2009, Compressive simultaneous full-waveform simulation: Geophysics, **74**, A35–A40.

Herrmann, F. J., Hanlon, I., Kumar, R., van Leeuwen, T., Li, X., Smithyman, B., Watson, H., Calvert, A. J., and M. Javanmehri, E. T. T., 2013, Frugal full-waveform inversion: From theory to a practical algorithm: The Leading Edge, **32**, 1082–1092.

Hu, J., Schuster, G. T., and A.Valasek, P., 2001, Poststack migration deconvolution: Geophysics, **66**, 939–952.

Hu, W., Abubakar, A., and Habashy, T. M., 2009, Simultaneous multifrequency inversion of full-waveform seismi data: Geophysics, **74**, R1–R14.

Hu, W., Abubakar, A., Habashy, T. M., and Liu, J., 2011, Preconditioned non-linear conjugate gradient method for frequency domain full-waveform seismic inversion: Geophysical Prospecting, **59**, 477–491.

Hudson, J. A., 1981, Wave speeds and attenuation of elastic waves in material containing cracks: Geophys.J.Roy.Astr.Soc, **64**, 133–150.

Hutchinson, M. F., 1990, A stochastic estimator of the trace of the influence matrix for Laplacian smoothing filters: Communications in Statistics -Simulation and Computation, **19**, 433–450.

Innanen, K. A., 2011, Inversion of the seismic AVF/AVA signatures of highly attenuative targets: Geophysics, **76**, R1–R14.

Innanen, K. A., 2012, Potentials for anelastic scattering: CREWES Annual Report, 1–16.

Innanen, K. A., 2013, Coupling in amplitude variation with offset and the wiggins approximation: Geophysics, **78**, N21–N33.

Innanen, K. A., 2014a, Reconciling seismic AVO and precritical reflection FWI-analysis of the inverse Hessian: SEG Technical Program Expanded Abstracts, 1022–1027.

Innanen, K. A., 2014b, Seismic AVO and the inverse Hessian in precritical reflection full waveform inversion: Geophysical Journal International, **199**, 717–734.

Jang, U., Min, D., and Shin, C., 2009, Comparison of scaling methods for waveform inversion: Geophysical Prospecting, **57**, 49–59.

Kamath, N., and Tsvankin, I., 2014, Sensitivity anaysis for elastic full-waveform inversion in VTI media: SEG Expanded Abstracts, 1162–1166.

Kennett, B. L. N., Sambridge, M. S., and Williamson, P. R., 1988, Subspace methods for large inverse problems with multiple parameter classes: Geophysical Journal Interational, **94**, 237–247.

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

Komatitsch, D., and Tromp, J., 2005, Introduction to the spectral-element method for 3-D seismic wave propagation: Geophysical Journal Interational, **139**, 806–822.

Krebes, E. S., 1987, Reflection and transmission at plane boundaries in non-welded contact: Journal of the Canadian Society of Exploration, **23**, 66–72.

Krebes, E. S., and Daley, P. F., 2007, Difficulties with computing anelastic plane-wave reflection and tranmission coefficients: Geophysical Journal International, **170**, 205–216.

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

Kuo, C., and Romanowicz, B., 2002, On the resolution of density anomalies in the Earth's mantle using spectral fitting of normal-mode data: Geophysical Journal International, **150**, 162–179.

Kwon, T., Seol, S. J., and Byun, J., 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 migration: Conference on Inverse Scattering, Theory and Applications, SIAM, Expanded Abstracts, 206–220.

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

Levenberg, K., 1944, A method for the solution of certain non-linear problems in least squares: Quarterly of Applied Mathematics, **2**, 164–168.

Li, Y., and Demanet, L., 2016, Full-waveform inversion with extrapolated low-frequency data: Geophysics, **81**, R339–R348.

Li, Y. E., and Demanet, L., 2015, Phase and amplitude tracking for seismic event seperation: Geophysics, **80**, WD59–WD72.

Lin, Y., 2015, Acoustic- and elastic-waveform inversion using a modified total-variation regularization scheme: Geophysical Journal International, **200**, 489–502.

Lines, L., 1999, Density and AVO: Canadian Society of Exploration Geophysics, **35**, 32–35.

Lines, L. R., Wong, J., Innanen, K. A., Vasheghani, F., Sondergeld, C., Treitel, S., and Ulrych, T., 2014, Research Note: Experimental measurements of Q-contrast reflections: Geophysical Prospecting, **62**, 190–195.

Liu, F., Hanson, D. W., Whitemore, N. D., Day, R. S., and Stolt, R. H., 2006, Toward a unified analysis for source plane-wave migration: Geophysics, **71**, S129–S139.

Liu, Q., and Tromp, J., 2006, Finite-frequency kernels based on adjoint methods: Bulletin of the Seismological Society of America, **96**, 2383–2397.

Liu, Y., Yang, J., Chi, B., and Dong, L., 2015, An improved scattering-integral approach for frequency-domain full waveform inversion: Geophysical Journal International, **202**, 1827–1842.

Lloyd, H., 2013, An investigation of the role of low frequencies in seismic impedance inversion: M.Sc. thesis, University of Calgary.

Luo, Y., 2012, Seismic imaging and inversion based on spectral-element and adjoint methods: Ph.D. thesis, Princeton University.

Luo, Y., and Schuster, G. T., 1991, Wave-equation traveltime inversion: Geophysics, **56**, 645–653.

Luo, Y., Tromp, J., Denel, B., and Calandra, H., 2013, 3D coupled acoustic-elastic migration with topography and bathymetry based on spectral-element and adjoint methods: Geophysics, **78**, S193–S202.

Ma, Y., and Hale, D., 2012, Quasi-Newton full-waveform inversion with a projected Hessian matrix: Geophysics, **77**, R207–R216.

Ma, Y., and Hale, D., 2013, Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion: Geophysics, **78**, R223–R233.

MacCarthy, J. K., Borchers, B., and Aster, R. C., 2011, Efficient stochastic estimation of the model resolution matrix diagonal and generalized cross-validation for large geophysical inverse problems: Journal of Geophysical Research, **116**, B10,304.

Mahmoudian, F., Margrave, G. F., Wong, J., and Henley, D. C., 2014, Azimuthal amplitude variation with offset analysis of physical modelling data acquired over an azimuthally anisotropic medium: Geophysics, **80**, C21–C35.

Marfurt, K., 1984, Accuracy of finite-difference and finite-elements modeling of the scalar and elastic wave equation: Geophysics, **49**, 533–549.

Margrave, G. F., Ferguson, R. J., and Hogan, C. M., 2010, Full waveform inversion with wave equation migration and well control: CREWES Annual Report, 1–20.

Margrave, G. F., Ferguson, R. J., and Hogan, C. M., 2011a, Full waveform inversion using wave-equation depth migration with tying to wells: SEG Technical Program Expanded Abstracts, 2454–2458.

Margrave, G. F., Mewhort, L., Phillips, T., Hall, M., Bertram, M. B., Lawton, D. C., Innanen, K., Hall, K. W., and Bertram, K., 2012, The hussar low-frequency experiment: CSEG Recorder, 25–39.

Margrave, G. F., Yedlin, M., and Innanen, K. A., 2011b, Full waveform inversion and the inverse Hessian: CREWES Annual Report, 1–13.

Marquardt, D., 1963, An algorithm for least-squares estimation of nonlinear parameters: SIAM Journal, **11**, 431–441.

Matson, K. H., 1997, An inverse scattering series method for attenuating elastic multiples from multicomponent land and ocean bottom seismic data: Ph.D. thesis, The University of British Columbia.

Menke, W., 1984, Geophysical data analysis: Discrete inverse theory: Academic Press.

Métivier, L., Bretaudeau, F., Brossier, R., Virieux, J., and Operto, S., 2014, Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures: Geophysical Prospecting, **62**, 1–23.

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

Métivier, L., Brossier, R., Operto, S., and Virieux, J., 2015, Acoustic multi-parameter fwi for the reconstruction of p-wave velocity, density and attenuation: preconditioned truncated newton approach: SEG Expanded Abstracts, 1198–1203.

Métivier, L., Brossier, R., Virieux, J., and Operto, S., 2012, The truncated Newton method for full waveform inversion: SEG Technical Program Expanded Abstracts, 1–5.

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

Modrak, R., and Tromp, J., 2016, Seismic waveform inversion best practices: regional, global, and exploration best cases: Geophysical Journal Interational, **206**, 1864–1889.

Modrak, R., Tromp, J., and Yuan, Y. O., 2016, On the choice of materials parameters for elastic waveform inversion: SEG Expanded Abstracts, 1115–1119.

Moghaddam, P. P., Keers, H., Herrmann, F. J., and Mulder, W. A., 2013, A new optimization approach for source-encoding full-waveform inversion: Geophysics, **78**, R125–R132.

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

Moradi, S., and Innanen, K. A., 2015, Scattering of homogeneous and inhomogeneous seismic waves in low-loss viscoelastic media: Geophysical Journal International, **202**, 1722–1732.

Morton, S. A., and Ober, C. C., 1998, Faster shot-record depth migrations using phase encoding test: SEG Technical Program Expanded Abstracts, 1131–1134.

Nammour, R., and Symes, W., 2009, Approximate constant density acoustic inverse scattering using dip-dependent scaling: SEG Technical Program Expanded Abstracts, 2347–2351.

Nash, S. G., 1985, Preconditioning of truncated-Newton methods: SIAM Journal on Scientific and Statistical Computing, **6**, 599–616.

Nash, S. G., 2000, A survey of truncated-Newton methods: Journal of Computational and Applied Mathematics, **124**, 45–59.

Nelson, R. A., 1985, Geologic analysis of naturally fractured reservoirs: Golf Professional Publishing, 2nd edn.

Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **35**, 773–782.

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

Oh, J., Alkhalifah, T., and Min, D., 2015, Multi-stage full waveform inversion strategy for 2D elastic VTI media: SEG Expanded Abstracts, 1204–1208.

Oh, J. W., and Alkhalifah, T., 2016, The scattering potential of partial derivative wavefields in 3-D elastic orthorhombic media: an inversion prospective: Geophysical Journal International, **206**, 1740–1760.

Oldenborger, G. A., and Routh, P. S., 2009, The point-spread function measure of resolution for the 3-Delectrical resistivity experiment: Geophysical Journal International, **176**, 405–414.

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

Pan, W., Geng, Y., and Innanen, K. A., 2018, Interparameter tradeoff quantification and reduction in isotropic-elastic full-waveform inversion: synthetic experiments and Hussar dataset application, submitted.

Pan, W., and Innanen, K. A., 2013, AVO/AVF analysis of thin beds in elastic media: SEG Expanded Abstracts, 373–377.

Pan, W., and Innanen, K. A., 2016a, Elastic full-waveform inversion: density effects, cycle-skipping, and inter-parameter mapping: CREWES Annual Report, 1–16.

Pan, W., and Innanen, K. A., 2016b, Suppress Parameter Cross-talk for Elastic Full-waveform Inversion: Parameterization and Acquisition Geometry: GeoConvention, 1–5.

Pan, W., and Innanen, K. A., 2016c, A Summary of Several Challenges Facing Multi-Parameter Elastic Full Waveform Inversion: CSEG Recorder, 36–39.

Pan, W., Innanen, K. A., and Liao, W., 2017a, Accelerating Hessian-free Gauss-Newton full-waveform inversion via l-BFGS preconditioned conjugate-gradient algorithm: Geophysics, **82**, R49–R64.

Pan, W., Innanen, K. A., and Margrave, G. F., 2014a, A comparison of different scaling methods for least-squares migration/inversion: EAGE Expanded Abstracts, We G103 14.

Pan, W., Innanen, K. A., Margrave, G. F., and Cao, D., 2015a, Efficient pseudo-Gauss-Newton full-waveform inversion in the τ-p domain: Geophysics, **80**, R225–R14.

Pan, W., Innanen, K. A., Margrave, G. F., Fehler, M. C., Fang, X., and Li, J., 2015b, Estimation of elastic constants in HTI media using Gauss-Newton and Full-Newton multi-parameter full waveform inversion: SEG Technical Program Expanded Abstracts, 1177–1182.

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

Pan, W., Innanen, K. A., and Yuan, Y. O., 2017b, Inter-parameter tradeoff quantification and reduction in isotropic-elastic full-waveform inversion: SEG Expanded Abstracts, 1–5.

Pan, W., Innanen, K. A., and Yuan, Y. O., 2017c, Interparameter tradeoff quantification and reduction in isotropic-elastic full-waveform inversion: SEG Expanded Abstracts, 1539–1544.

Pan, W., Innanen, K. A., Yuan, Y. O., and Simons, F., 2017d, Quantifying parameter tradeoff in elastic full-waveform inversion via multi-parameter Hessian probing: GeoConvention, 1–5.

Pan, W., Margrave, G. F., and Innanen, K. A., 2014b, Iterative modeling migration and inversion (IMMI): Combining full waveform inversion with standard inversion methodology: SEG Technical Program Expanded Abstracts, 938–943.

Perrone, F., and Sava, P., 2012, Wave-equation migration with dithered plane waves: Geophysical Prospecting, **60**, 444–465.

Peters, B., and Herrmann, F. J., 2017, Constraints versus penalties for edge-preserving full-waveform inversion: The Leading Edge, **36**, 94–100.

Pica, A., Diet, J. P., and Tarantola, A., 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.

Plessix, R. E., and Mulder, W. A., 2004, Frequency-domain finite-difference amplitude-preserving migration: Geophysical Journal International, **157**, 975–987.

Plonka, A., Blom, N., and Fichtner, A., 2016, The imprint of crustal density heterogeneities on regional seismic wave propagation: Solid Earth, **7**, 1591–1608.

Podgornova, O., Leaney, S., and Liang, L., 2015, Analysis of resolution limits of VTI anisotropy with full waveform inversion: SEG Expanded Abstracts, 1188–1192.

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

Pratt, R. G., and Chapman, C. H., 1992, Traveltime tomography in anisotropic media-II. application: Geophysical Journal International, **109**, 20–37.

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

Pratt, R. G., and Worthington, M. H., 1990, Inverse theory applied to multi-source cross-hole tomography. Part I: acoustic wave-equation method: Geophysical Prospecting, **38**, 287–310.

Prieux, V., Brossier, R., Operto, S., and Virieux, J., 2013a, Multiparameter full waveform inversion of multicomponent ocean-bottom cable data from the valhall field. part 1: imaging compressional wave speed, density and attenuation: Geophysical Journal International, **194**, 1640–1664.

Prieux, V., Brossier, R., Operto, S., and Virieux, J., 2013b, Multiparameter full waveform inversion of multicomponent ocean-bottom cable data from the valhall field. Part 1: imaging compressional wave speed, density and attenuation: Geophysical Journal International, **194**, 1640–1664.

Raknes, E. B., and Arntsen, B., 2015, A numerical study of 3D elastic time-lapse full-waveform inversion using multicomponent seismic data: Geophysics, **80**, R303–R315.

Rawlinson, N., Fichtner, A., Sambridge, M., and K.Young, M., 2014, Seismic tomography and the assessment of uncertainty: Advances in Geophysics, **55**, 1–76.

Rawlinson, N., and Spakman, W., 2016, On the use of sensitivity tests in seismic tomography: Geophysical Journal International, **205**, 1221–1243.

Ren, H., Wu, R., and Wang, H., 2011, Wave equation least square imaging using the local angular hessian for amplitude correction: Geophysical Prospecting, **59**, 651–661.

Rickers, F., Fichtner, A., and Trampert, J., 2013, The Iceland-Jan Mayen plume system and its impact on mantle dynamics in the North Atlantic region: Evidence from full-waveform inversion: Earth and Planetary Science Letters, **367**, 39–51.

Romero, L. A., Ghiglia, D. C., Ober, C. C., and Morton, S. A., 2000, Phase encoding of shot records in prestack migration: Geophysics, **65**, 426–436.

Routh, P., Krebs, J., Lazaratos, S., Baumstein, A., Lee, S., C, Y. H., Chikichev, I., Downey, N., Hinkley, D., and Anderson, J., 2011, Encoded simultaneous source full-wavefield inversion for spectrally shaped marine streamer data: SEG Technical Program Expanded Abstracts, 2433–2438.

Rüger, A., 1997, P-wave reflection coefficients for transversely isotropic models with vertical and horizontal axis of symmetry: Geophysics, **62**, 713–722.

Rusmanugroho, H., Modrak, R., and Tromp, J., 2017, Anisotropic full-waveform inversion titl-angle recovery: Geophysics, **82**, R135–R151.

Russell, B., 1998, Introduction to Seismic Inversion Methods: Scoiety Exploration Geophysicists, Tulsa.

Saad, Y., 2003, Iterative methods for sparse linear systems: SIAM.

Sacchi, M. D., Wang, J., and Kuehl, H., 2007, Estimation of the diagonal of the migration blurring kernel through a stochastic approximation: SEG Expanded Abstracts, 2437–2441.

Sainath, T. N., Horesh, L., Kingsbury, B., Aravkin, A. Y., and Ramabhadran, B., 2013, Accelerating Hessian-free optimization for Deep Neural Networks by implicit preconditioning and sampling: IEEE Workshop on Automatic Speech Recognition and Understanding, 303–308.

Santosa, F., and Symes, W., 1988, Computation of the hessian for least-squares solutions of inverse problems of reflection seismology: Inverse Problems, **4**, 211–233.

Schoenberg, M., 1983, Reflection of elastic waves from periodically stratified media with interfacial slip: Geophysical Prospecting, **31**, 265–292.

Shanno, D. F., 1970, Conditioning of quasi-Newton methods for function minimization: Mathematics of Computation, **24**, 647–656.

Shin, C., and Cha, Y. H., 2008, Waveform inversion in the Laplace domain: Geophysical Journal International, **173**, 922–931.

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

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

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

Son, W., Pyun, S., Shin, C., and Kim, H. J., 2014, An algorithm adapting encoded simultaneous-source full-waveform inversion to marine-streamer acquisition data: Geophysics, **79**, R183–R193.

Spakman, W., 1991, Delay-time tomography of the upper mantle below Europe, the Mediterranean and Asia minor: Geophysical Journal International, **107**, 309–332.

Stolt, R. H., and Weglein, A. B., 2012, Seismic Imaging and Inversion: Application of Linear Inverse Theory: Cambriage University Press.

Stork, C., and Kapoor, J., 2004, How many p values do you want to migrate for delayed shot wave equation migration?: SEG Technical Program Expanded Abstracts, 1041–1044.

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

Tang, Y., and Lee, S., 2015, Multi-parameter full wavefield inversion using non-stationary point-spread functions: SEG Technical Program Expanded Abstracts, 1111–1115.

Tao, Y., and Sen, M. K., 2013, Frequency-domain full waveform inversion with plane-wave data: Geophysics, **78**, R13–R23.

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

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

Tarantola, A., 2005, Inverse Problem Theory and Methods for Model Parameter Estimation: Society of Industrial and Appplied Mathematics.

Thomsen, L., 1988, Reflection seismology over azimuthally anisotropic media: Geophysics, **53**, 304–313.

Trad, D. O., Ulrych, T. J., and Sacchi, M. D., 2002, Accurate interpolation with high-resolution time-variant Randon transforms: Geophysics, **67**, 664–656.

Trad, D. O., Ulrych, T. J., and Sacchi, M. D., 2003, Latest views of the sparse Radon transform: Geophysics, **68**, 386–399.

Trampert, J., Fichtner, A., and Ritsema, J., 2013, Resolution tests revisited: the power of random numbers: Geophysical Journal International, **192**, 676–680.

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

Tsvankin, I., 1997a, Anisotropic parameters and p-wave velocity for orthorhombic media: Geophysics, **62**, 1292–1309.

Tsvankin, I., 1997b, Reflection moveout and parameter estimation for horizontal transverse isotropy: Geophysics, **62**, 614–629.

Tsvankin, I., and Grechka, V., 2011, Seismology of azimuthally anisotropic media and seismic fracture characterization: SEG.

Valenciano, A., 2008, Imaging by wave-equation inversion: Ph.D. thesis, Stanford University.

Valenciano, A. A., Biondi, B., and Guitton, A., 2006, Target-oriented wave-equation inversion. geophysics: Geophysics, **71**, A35–A38.

van Leeuwen, T., Aravkin, A. Y., and Herrmann, F. J., 2011, Seismic waveform inversion by stochastic optimization: International Journal of Geophysics, 1–18.

van Leeuwen, T., and Mulder, W. A., 2010, A correlation-based mifit criterion for wave-equation traveltime tomography: Geophysical Journal International, **182**, 1383–1394.

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

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

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

Wang, B., Gao, F., Wheaton, D., and Dirks, V., 2006, Model-based decimation of input data for delayed-shot/plane-wave migration for the purpose of subsalt velocity model building: SEG Technical Program Expanded Abstracts, 2186–2190.

Wang, T., and Cheng, J. B., 2017, Elastic full-waveform inversion based on mode decomposition: the approach and mechanism: Geophysical Journal International, **209**, 606–622.

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

Warner, M., Ratclie, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N., Vinje, V., Stekl, I., Guasch, L., Win, C., Conroy, G., and Bertrand, A., 2013, Anisotropic 3D full-waveform inversion: Geophysics, **78**, R59–R80.

Waters, K. H., 1978, Reflection seismology: a tool for energy resource exploration: John Wiley and Sons Inc., New York.

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

Wu, S., Wang, Y., Zheng, Y., and Chang, X., 2015, Limited-memory bfgs based least-squares pre-stack Kirchhoff depth migration: Geophysical Journal International, **202**, 738–747.

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

Xu, K., and McMechan, G. A., 2014, 2D frequency-domain elastic full-waveform inversion using time-domain modeling and a multistep-length gradient approach: Geophysics, **79**, R41–R53.

Xu, S., Wang, D., Chen, F., Zhang, Y., and Lambare, G., 2012, Full waveform inversion for reflected seismic data: EAGE Technical Program Expanded Abstracts.

Yang, J., Liu, Y., and Dong, L., 2016, Simultaneous estimation of velocity and density in acoustic multiparameter full-waveform inversion using an improved scattering-integral approach: Geophysics, **81**, R399–R415.

Yuan, Y. O., and Simons, F. J., 2014, Multiscale adjoint waveform-difference tomography using wavelets: Geophysics, **79**, WA79–WA95.

Yuan, Y. O., Simons, F. J., and Bozdağ, E., 2015, Multiscale adjoint waveform tomography for surface and body waves: Geophysics, **80**, R281–R302.

Zhang, J., and Toksöz, M. N., 1998, Nonlinear refraction traveltime tomography: Geophysics, **63**, 1726–1737.

Zhang, S., Schuster, G., and Luo, Y., 2011, Wave-equation reflection traveltime inversion: SEG Expanded Abstracts, 2705–2710.

Zhang, Y., Sun, J., Notfors, C., Gray, S. H., Chernis, L., and Young, J., 2005, Delayed-shot 3D depth migration: Geophysics, **70**, E21–E28.

Zheng, Y., Fang, X., Fehler, M. C., and Burns, D. R., 2013, Seismic characterization of fractured reservoirs by focusing gaussian beams: Geophysics, **78**, A23–A28.

Zhu, H., Bozdag, E., and Tromp, J., 2015, Seismic structure of the european upper mantle based on adjoint tomography: Geophysical Journal Interational, **201**, 18–52.

Zhu, H., and Fomel, S., 2016, Building good starting models for full-waveform inversion using adaptive matching filtering misfit: Geophysics, **81**, U61–U72.

Zhu, H., Li, S., Fomel, S., Stadler, G., and Ghattas, O., 2016, A Bayesian approach to estimate uncertainty for full-waveform inversion using a prior information from depth migration: Geophysics, **81**, R307–R323.

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

Zhu, X., Sixta, D. P., and Angstman, B. G., 1992, Tomostatics:tuning-ray tomography + static corrections: The Leading Edge, **11**, 15–23.

Zhu, X., Valasek, P., Roy, B., Shaw, S., Howell, J., Whitney, S., Whitmore, N. D., and Anno, P., 2008, Recent applications of turning-ray tomography: Geophysics, **73**, VE243–VE254.