Department of Exploration Geophysics

Joint Non-Linear Inversion of Amplitudes and Travel Times in a Vertical Transversely Isotropic Medium using Compressional and Converted Shear Waves

Dariush Nadri

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University of Technology

May 2008

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature: 

Date: 15 May 2008

Dedicated to the memory of my friend *Hojjat Nadri (1960-2003)*.

Abstract

Massive shales and fractures are the main cause of seismic anisotropy in the upper-most part of the crust, caused either by sedimentary or tectonic processes. Neglecting the effect of seismic anisotropy in seismic processing algorithms may incorrectly image the seismic reflectors. This will also influence the quantitative amplitude analysis such as the acoustic or elastic impedance inversion and amplitude versus offsets analysis. Therefore it is important to obtain anisotropy parameters from seismic data. Conventional layer stripping inversion schemes and reflector based reflectivity inversion methods are solely dependent upon a specific reflector, without considering the effect of the other layers. This, on one hand, does not take the effect of transmission in reflectivity inversion into the account, and on the other hand, ignores the information from the waves travelling toward the lower layers. I provide a framework to integrate the information for each specific layer from all the rays which have travelled across this layer. To estimate anisotropy parameters I have implemented unconstrained minimization algorithms such as nonlinear conjugate gradients and variable metric methods, I also provide a nonlinear least square method, based on the Levenberg-Marquardt algorithm. In a stack of horizontal transversely isotropic layers with vertical axis of symmetry, where the layer properties are laterally invariant, we provide two different inversion schemes; traveltime and waveform inversion. Both inversion schemes utilize compressional and joint compressional and converted shear waves. A new exact traveltime equation has been formulated for a dipping transversely isotropic system of layers. These traveltimes are also parametrized by the ray parameters for each ray element. I use the Newton method of minimization to estimate the ray parameter using a random prior model from a uniform distribution. Numerical results show that with the assumption of weak anisotropy, Thomsen's anisotropy parameters can be estimated with a high accuracy. The inversion algorithms have been implemented as a software package in a C++ object oriented environment.

Acknowledgments

I would like to thank my supervisor Dr Bruce Hartley for his support, patience, and availability. Without his support completing this thesis would not be easy. I appreciate my chairperson Professor Boris Gurevich for his invaluable comments on most of the chapters. His careful reading improved the accuracy and clarity of this thesis. I also have to acknowledge his financial supports for conference through Curtin Reservoir Geophysics Consortium (CRGC). I would like to thank Dr Milovan Urosevic, who brought me to seismic anisotropy area. He has been always cheerful and encouraging me during my study. I also express my thanks to Dr Andrej Bóna who carefully reviewed the chapter on minimization.

Many other people have to be acknowledged. I particularly thank Dr Ruping Li for her discussions about seismic inversion and seismic anisotropy at the beginning of this research. Special thank goes to Dr James Gunning for his comments on stochastic seismic inversion and his time while I visited him in Melbourne. I would like to thank other people which I have been in contact with them via email. Dr Terje Dahl for clarification about his paper, Dr Andreas Rüger, Dr Alexey Stovas, Dr Ross Lippert for his comments on the Hessian matrix and preconditioning, and Professor Zhen-Jun Shi for his comments on the convergence rate of the objective function.

I would like to thank all the staff member of Department of Exploration Geophysics, who have helped me during my study especially Ms Deirdre Hollingsworth for her administrative help, Dr Mark Lwin for his help in the physical modelling lab, and Mr Robert Verstandig for his computing help. I also thank Dr Don Sherlock for providing the physical model and his cooperation.

I should thank the PhD students for providing a friendly environment and specially thank to my colleague Mr Abdulla Al Ramadhan for technical discussions. I have to acknowledge all my friends, while I was living in campus. I had a wonderful time with them in an amazingly diverse multicultural environment.

I thank National Iranian Oil Company for sponsoring me during most of my study and also the Department of Exploration Geophysics for provide the financial support for the rest of my tuition fees.

Table of Contents

Introduction……..1
Background…… 2
Methodology……..... 4
Thesis outline…….. 5
References……7

Chapter 1……..9
Review of seismic anisotropy……...9
Introduction……...9
1.1 Hooke's law and Stiffness Tensor……10
1.2 Seismic wave propagation in transversely isotropic (TI) media……15
1.3 Numerical examples……...22
References……...27

Chapter 2……30
Ray tracing in transversely isotropic media……...30
Introduction……30
2.1 Parametric wave propagation in terms of ray parameter……...30
2.2 Traveltime and offset in a horizontally layered media……..33
2.3 Two point ray tracing in a horizontally layered media……..36
2.4 Two point ray tracing in a dipping layered media……41
2.4.1 Snell's law in a dipping layered media……..41
2.4.2 Traveltime and offset in a dipping layered media……42
Appendices……51
Appendix A……51
Appendix B……54
Appendix C……55
Appendix D……56
Appendix E……57
Appendix F……57
Appendix G……59
References……..61

Chapter 3……63
Reflection and transmission of plane waves in transversely isotropic media……63
Introduction……63
3.1 Polarization vectors and vertical slowness in a TI medium……...65
3.2 Boundary conditions at the interface of two solid elastic TI media……..66
3.3 Exact plane wave reflection and transmission coefficients……...68
3.4 Approximate plane wave reflection and transmission coefficients……...71
3.5 Numerical examples of exact and approximate R/T coefficients……..76
3.6 Computation of synthetic seismograms in a layered media……...81
References……..88

Chapter 4……90
A review of parameter estimation……..90
Introduction……90
4.1 The anatomy of inverse problems……..90
4.2 Nonlinear least square optimization methods ……92
4.3 Least square inversion of traveltimes in a multi layered medium……...101
4.4 Unconstrained minimization using iterative gradient methods……...106
4.4.1 Golden section search method……...…….107
4.4.2 Inexact line search methods……..…….109
4.4.3 Exact line search methods……..……111
4.4.4 The Method of steepest descent……..……...114
4.4.5 Conjugate gradient algorithm……..……..117
4.4.6 Variable metric (quasi-Newton) methods……..……127
References……131

Chapter 5……..134
Unconstrained inverse models for traveltime and amplitude……..134
Introduction……134
5.1 Partial derivatives of traveltime in a layered TI media……134
5.2 Partial derivatives of reflection and transmission coefficients in a TI horizontal layered media……..136
5.3 Unconstrained inverse model for traveltime inversion in a multi layered Media……...140
5.4 Unconstrained inverse model for waveform inversion in a multi layered Media……...141
Appendices……147
Appendix A: Partial derivatives of offset and traveltime in a horizontal VTI layered medium……..147
Appendix B: Partial derivatives of offset and traveltime in a dipping TI layered Medium……...149
Appendix C: Partial derivatives of and with respect to model parameters and ray parameter……..152
References……..155

Chapter 6……..156
Numerical results of unconstrained inversion of traveltime and amplitude…….156
Introduction……...156
6.1 Traveltime inversion in a horizontally layered VTI media……..156
6.2 Waveform inversion in horizontally layered VTI media…….181
6.3 Traveltime inversion in a dipping layered TI media……191
Comments……..…198
Appendix ……..….199
References……..….200

Conclusions, Discussions, and Future prospects……201
Conclusion and discussion……...201
Future prospects……...203
References……204

List of Figures

Figure 1.1: Ray and phase angles in a homogeneous VTI medium……...…13
Figure 1.2: A schematic illustration of a VTI medium…….……...……..15
Figure 1.3: A schematic illustration of a HTI medium……...……...15
Figure 1.4: Seismic wave propagation and polarization pattern of P, SV, and SH waves for Dog Creek Shale……...24
Figure 1.5: Seismic wave propagation and polarization pattern of P, SV, and SH waves for Mesaverde Shale……..25
Figure 1.6: Seismic wave propagation and polarization pattern of P, SV, and SH waves for Green River Shale…….……...26
Figure 2.1: Ray paths for $P \to P$ and $P \to S_v$ travelling through horizontal anisotropic media……..35
Figure 2.2: A ray element travelling upward through a horizontal layer……...35
Figure 2.3: A typical ray parameter solution space……...36
Figure 2.4: Ray parameters estimated for $P \to P$ rays from the horizontal layered media.....39
Figure 2.5: Ray parameters estimated for $P \to S_v$ rays from the horizontal layered media……..39
Figure 2.6: Traveltimes calculated for $P \to P$ rays from the horizontal layered media....40
Figure 2.7: Traveltimes calculated for $P \to S_v$ rays from the horizontal layered media...40
Figure 2.8: Incidence and refraction phase angles in a dipping media……..43
Figure 2.9: Ray paths for $P \to P$ and $P \to S_v$ travelling through a dipping layered anisotropic media……..….45
Figure 2.10: A ray element travelling downward through a dipping layer……....45
Figure 2.11: Solution space of objective function in a dipping medium for near offsets..47
Figure 2.12: Solution space of the logarithmic objective function in a dipping medium for larger offsets……..47

Figure 2.13: Jump detection based on the curvature. The red star is a local minimum solution for ray parameter has resulted in a drop in traveltime…….48
Figure 2.14: Analytic and numerical derivatives for the ray path reflecting from the top of layer 10 in Table 2.2…….……48
Figure 2.15: Traveltimes calculated for $P \to P$ rays from the dipping layered medium in Table 2.2……....50
Figure 2.16: Traveltimes calculated for $P \to S_v$ rays from the dipping layered medium in Table 2.2……....50
Figure D.1: A ray element travelling downward through a dipping layer…….56
Figure 3.1: Geometry of incidence $P \,\&\, SV$ waves, correspondent polarization vectors and slowness components at the interface in the symmetry plane of two TI medium…..67
Figure 3.2: Reflection coefficients for $P \to P$ waves from the interface between layer 1 and layer 2……..77
Figure 3.3: Reflection coefficients for $P \to S_v$ waves from the interface between layer 1 and layer 2……..77
Figure 3.4: Reflection coefficients for $P \to P$ waves from the interface between layer 5 and layer 6……..79
Figure 3.5: Reflection coefficients for $P \to S_v$ waves from the interface between layer 5 and layer 6……..79
Figure 3.6: Reflection coefficients for $P \to P$ waves from the interface between layer 9 and layer 10……....80
Figure 3.7: Reflection coefficients for $P \to S_v$ waves from the interface between layer 9 and layer 10……....80
Figure 3.8: A minimum phase wavelet with a 25 Hz dominant frequency and 2ms sampling interval……83
Figure 3.9: Exact and approximate plane compressional and converted shear waves of a single traces at near and far offset..…….……..84
Figure 3.10: A synthetic seismogram from the exact and approximate compressional plane wavefield …….……....85

Figure 3.10: A synthetic seismogram from the exact and approximate shear plane wavefield……....86
Figure 3.11: A synthetic seismogram from the exact and approximate compressional and converted shear plane wavefield ..……..……...……....87
Figure 4.1: A schematic view of an inverse problem……92
Figure 4.2: Reflection traveltimes of compressional waves in a layered medium……..102
Figure 4.3: Golden ratio search……...108
Figure 4.4: Golden section search bracketing……108
Figure 4.5: Line of sufficient decrease and Wolfe conditions for the objective function……111
Figure 4.6: Successive fitting a parabola with the objective function……....113
Figure 4.7: Minimization of a two parameter linear equation using the method of steepest descent……..116
Figure 4.8: Convergence pattern of different nonlinear conjugate gradient method…...122
Figure 4.9: Convergence pattern of preconditioned nonlinear conjugate gradient using the method of Polak-Ribiére……..…126
Figure 4.10: Convergence pattern of variable metric method using the BFGS algorithm.....130
Figure 5.1: Reflection and transmission coefficients of a plane wave across a M layer TI media……....136
Figure 6.1: Convergence pattern of the 20 realizations of P-wave velocity of layer 1 from the inversion of far offset traveltimes for different δ and ε . ..……..164
Figure 6.2: Convergence pattern of the 20 realizations of P-wave velocity of layer 1 from the inversion of near offset traveltimes for different δ and ε …….. ..……..164
Figure 6.3: Convergence pattern of the 20 realizations of thickness of layer 1 from the inversion of far offset traveltimes for different δ and ε . ..……..…..165
Figure 6.4: Convergence pattern of the 20 realizations of thickness of layer 1 from the inversion of near offset traveltimes for different δ and ε . ..…….…..…..165
Figure 6.5: Convergence pattern of the 20 realizations of objective function from the inversion of far offset traveltimes of layer 1 for different δ and ε ……..…….166

Figure 6.6: Convergence pattern of the 20 realizations of objective function from the inversion of near offset traveltimes of layer 1 for different δ and ε ……166
Figure 6.7: Convergence pattern of the thickness of layer 1 from inversion of the traveltime data……..168
Figure 6.8: Convergence pattern of the thickness of the layer 7 from inversion of the traveltime data……..…168
Figure 6.9: Convergence pattern of the P-wave velocity of layer 1 from inversion of the traveltime data……..169
Figure 6.10: Convergence pattern of the P-wave velocity of layer 7 from inversion of the traveltime data……..169
Figure 6.11: Convergence pattern of the S-wave velocity of layer 1 from inversion of the traveltime data……..170
Figure 6.12: Convergence pattern of the S-wave velocity of layer 7 from inversion of the traveltime data……..170
Figure 6.13: The Frobenius Norm shows the convergence rate of the simultaneous traveltime inversion of traveltimes for $P \to P$ and joint $P \to P \,\&\, P \to S_V$ …...……..……...……..……...171
Figure 6.14: The Frobenius Norm shows the convergence rate of the five realizations with different random prior model of simultaneous $P \to P$ traveltime inversion ….....175
Figure 6.15: Convergence pattern of the δ for layer 1 from five realizations…..……..176
Figure 6.16: Convergence pattern of the ε for layer 1 from five realizations……....176
Figure 6.17: Convergence pattern of the δ for layer 7 from five realizations……....177
Figure 6.18: Convergence pattern of the ε for layer 7 from five realizations……....177
Figure 6.19: Convergence pattern of the δ for layer 9 from five realizations……....178
Figure 6.20: Convergence pattern of the ε for layer 9 from five realizations……....178
Figure 6.21: The effect of preconditioning on the convergence rate in the conjugate gradient algorithm……..……..180
Figure 6.22: Convergence rate of the simultaneous $P \to P$ and joint $P \to P \,\&\, P \to S_V$ waveform inversion……..184
Figure 6.23: Convergence pattern of the δ for layer 1 from waveform inversion…….185
Figure 6.24: Convergence pattern of the ε for layer 1 from waveform inversion…..…185

Figure 6.25: Convergence pattern of the δ for layer 4 from waveform inversion…….186
Figure 6.26: Convergence pattern of the ε for layer 4 from waveform inversion……..186
Figure 6.27: Convergence pattern of the δ for layer 9 from waveform inversion…….187
Figure 6.28: Convergence pattern of the ε for layer 8 from waveform inversion……..187
Figure 6:29: Synthetic seismograms from a true model, the amplitude residual from a prior model, and the residual from the estimated model parameter by waveform inversion of compressional wavefield…….188
Figure 6:30: Synthetic seismograms from a true model, the amplitude residual from a prior model, and the residual from the estimated model parameter by waveform inversion of joint compressional and converted shear wavefield.....……189
Figure 6.31: Convergence rate of the simultaneous inversion of $P \to P$ and joint $P \to P$ & $P \to S_V$ traveltimes in dipping TI media……....193
Figure 6:32 Convergence pattern of the δ for layer 1 from traveltime inversion of $P \to P$ and joint $P \to P$ & $P \to S_V$ in dipping TI media……...…194
Figure 6:33 Convergence pattern of the ε for layer 1 from traveltime inversion of $P \to P$ and joint $P \to P$ & $P \to S_V$ in dipping TI media……....194
Figure 6:34 Convergence pattern of the δ for layer 3 from traveltime inversion of $P \to P$ and joint $P \to P$ & $P \to S_V$ in dipping TI media……...……..…195
Figure 6:35 Convergence pattern of the ε for layer 3 from traveltime inversion of $P \to P$ and joint $P \to P$ & $P \to S_V$ in dipping TI media……195
Figure 6:36 Convergence pattern of the δ for layer 7 from traveltime inversion of $P \to P$ and joint $P \to P$ & $P \to S_V$ in dipping TI media……196
Figure 6:37 Convergence pattern of the ε for layer 7 from traveltime inversion of $P \to P$ and joint $P \to P$ & $P \to S_V$ in dipping TI media……196
Figure 6:38 Convergence pattern of the δ for layer 9 from traveltime inversion of $P \to P$ and joint $P \to P$ & $P \to S_V$ in dipping TI media……197
Figure 6:39 Convergence pattern of the ε for layer 9 from traveltime inversion of $P \to P$ and joint $P \to P$ & $P \to S_V$ in a dipping TI media……..…197
Figure 6.40: A schematic class diagram illustrates different classes of the package…..199

List of Tables

Table 1.1: Stiffness tensor for (a) Dog Creek Shale, (b) Mesaverde Shale, and (c) Green River Shale…...……..23

Table 2.1: Layer properties used to calculate $P \to P$ and $P \to S_v$ traveltimes……..… 38

Table 2.2: Layer properties used to calculate $P \to P$ and $P \to S_v$ traveltimes for the dipping layered…….…… 49

Table 5.1: the prior model has been used to compute the analytic and numerically derived partial derivatives…… 144

Table 5.2: Analytically computed partial derivatives of the objective function using the amplitudes of compressional waves…….145

Table 5.3: Numerically computed partial derivatives of the objective function using the amplitudes of compressional waves…….145

Table 5.4: Analytically computed partial derivatives of the objective function using the amplitudes of compressional waves…….146

Table 6.1: Prior and estimated model of the 20 realizations from traveltime inversion of far offset data for layer 1…..……167

Table 6.2: Prior and estimated model of the 20 realizations from traveltime inversion of near offset data for layer 1…...……167

Table 6.3: True layer properties which have been used to create observed $P \to P$ and $P \to S_v$ traveltimes……172

Table 6.4: The prior model which has been used in minimization of $P \to P$ and $P \to S_v$ traveltimes……172

Table 6.5: Estimated model parameters from minimization of the $P \to P$ for the last iteration in Figure 6.13……...173

Table 6.6: Estimated model parameters from minimization of the joint $P \to P$ and $P \to S_v$ traveltimes for the last iteration in Figure 6.13..……..173

Table 6.7: Estimated model parameters from minimization of the $P \to P$ for the last iteration in Figure 6.13….……...174

Table 6.8: Estimated model parameters from minimization of the joint $P \to P$ and $P \to S_v$ traveltimes for the last iteration in Figure 6.13……..174

Table 6.9: Prior model and estimated Thomsen's anisotropy parameters from four different realizations……179

Table 6.10: Estimated model parameters from waveform inversion of $P \to P$ and joint $P \to P \,\&\, P \to S_v$ ……190

Table 6.11: True model and prior models for waveform inversion of $P \to P$ and joint $P \to P \,\&\, P \to S_v$ ……190

Table 6.12: The prior model and estimated parameter models form inversion of $P \to P$ and joint $P \to P \,\&\, P \to S_v$ traveltimes in dipping TI media……198

Introduction

With the advent of new seismic data acquisition technology and processing algorithms, seismic anisotropy, which had been considered as an academic practice two decades ago, has now provided a large impact on the quality of processed data and also on the quantitatively interpreted seismic data. Without considering the effect of anisotropy, exploration seismologist would have ignored the multi component data two decades ago (Alford, 1986; Grechka, 2001). Converted shear wave data recorded from ocean-bottom seismic surveys have gained much from considering anisotropy. Anisotropy is mainly due to massive shales and fractured rocks under sedimentary or tectonic processes. Transverse isotropy with the vertical axis of symmetry (Polar anisotropy) is one of the most common types of seismic anisotropy routinely used in seismic imaging and quantitative seismic interpretation such as amplitude versus offset analysis and impedance inversion. P-wave prestack time and depth migration is further extended to include transverse isotropy with the horizontal axis of symmetry (Azimuthal anisotropy) due to vertical fractures (Thomsen, 2001). These models, despite the simplicity in modelling the anisotropy of geological formations are still the most commonly used types of seismic anisotropy.

After introducing the Thomsen's anisotropy parameters (Thomsen, 1986a), P-wave and S-wave velocity expressions become simpler and give more intuitive insight. They further simplified the reflectivity to more understandable equations (Rüger, 1996). Estimation of Thomsen's anisotropy parameters has been an important area of research in seismic anisotropy and research is still going on to provide more accurate estimates.

In this thesis I provide two inversion strategies to estimate Thomsen's anisotropy parameters from seismic data; traveltime inversion and waveform inversion. Due to the importance of converted shear waves, we implement both compressional and joint compressional and converted shear wave data. I use exact traveltimes and approximate reflection and transmission coefficients for different modes in horizontally layered media. I also provide a new formulation for exact P-wave and SV-wave traveltimes in dipping transversely isotropic media. These equations are also used to estimate Thomsen's anisotropy parameters in dipping layered media.

Background

Estimation of elastic stiffness coefficients and anisotropic structure of the earth has been of interest, particularly in earthquake seismology, from the earliest days of seismic anisotropy studies. Crampin and Bamford (1977) studied the anisotropy structure of the upper mantle from azimuthal P-wave data using transverse isotropy and orthorhombic models. From the earliest works on anisotropic parameter estimation in modern exploration seismology, I can refer to works of Miller and Spencer (1994) and Miller et al. (1994). They expressed the vertical slowness of quasi-compressional waves in a transverse isotropic medium with vertical axis of symmetry (VTI), as a linear function of elastic stiffness coefficients. Assuming a horizontally layered medium and using wide aperture walk-away vertical seismic profile data, they have estimated the elastic stiffness coefficients for a shale layer from the horizontal and vertical slowness of quasi compressional waves, however, their methodology resolved C_{55} poorly.

Li (2002) introduced a layer striping inversion scheme using the Levenberg-Marquardt algorithm to estimate Thomsen's anisotropy parameters in layered VTI media from quasi-compressional wave traveltimes from VSP data. In a stack of n horizontally homogeneous, transverse isotropic layers, she considers the $n-1$ upper layers over the layer n as a single effective layer with the same effect on the traveltime as if the ray travels across the $n-1$ layers. Layer stripping based inversion schemes do not incorporate the ray information from all the ray paths travelling across a system of n layers. This would cause the lack of robustness in parameter estimation.

Hyperbolic and nonhyperbolic moveout based anisotropic parameter estimation is extensively studied and practically implemented in both transversely isotropic and orthorhombic media (Tsvankin and Thomsen, 1995; Tsvankin, 1997; Grechka and Tsvankin, 1998; Grechka et al., 1999; Grechka and Tsvankin, 1999; Thomsen et al., 1999; Grechka and Tsvankin, 2000). These are a class of layer stripping linear inversion methods which mostly minimize traveltimes for both vertical normal moveout velocity and processing parameter (anellipticity) η. The effectiveness of η is in its parameterization which combines two Thomsen's anisotropy parameter δ and ε. They have been implemented on P-wave, S-wave, and joint $P \to P$ & $P \to S_v$ traveltimes. These

algorithms do not use the exact traveltime equations. Instead, they use the Taylor approximations. Most of these algorithms are processing algorithms and are used in anisotropic NMO correction and seismic imaging. They do not decouple processing parameter η into δ and ε. Hence, these algorithms may not be very helpful in the anisotropic reflectivity analysis such as amplitude versus offset (AVO) analysis and impedance inversion.

Traveltime based inversion methods generally estimate the coarse part of the structure. In successive sand shale layers where the thickness of layers is less than the Rayleigh resolution limit (a quarter of dominant wavelength), discrimination of the top and the base of the layers is impossible. Constructive and destructive interference makes it difficult to distinguish the layer boundaries. Prestack and particularly poststack inversion has been widely used in isotropic elastic and acoustic impedance successfully. Isotropic poststack inversion is routinely used in commercial applications. With the knowledge of the source wavelet, the high frequency part of the structure can be successfully estimated. Estimation of Thomsen's anisotropic parameters from partial stack seismic data is unreliable and a wide range of different angles of incidence are needed to recover these parameters. Hence, prestack anisotropic inversion algorithms should be used instead. The prestack anisotropic inversion is computationally intensive, so full waveform and even exact plane wave formalism may not be economically justified. Approximate analytic anisotropic reflectivity equations of Rüger (1996), Thomsen (1993), and Jílek (2002), could be used instead. These linear equations are generally valid for pre-critical angles of incidence.

Jílek (2002) proposed linear and nonlinear algorithms using approximate and exact plane waves respectively for joint $P \to P \,\&\, P \to S_v$ reflection coefficients at the boundary of two transversely isotropic media. Incorporating the converted shear waves increases the nonlinearity but stabilizes estimates of Thomson's anisotropy parameters.

Simões-Filho et al. (1999) inverted the elastic stiffness coefficients at the interface of a VTI/HTI media. They used a genetic algorithm to invert the exact plane wave reflectivity equations of Schoenberg and Protázio (1992). Free partial derivative minimization algorithms such as simplex method and especially global optimization algorithms such as Genetic algorithm and Markov Chain Monte Carlo require extensive

function evaluation. Global optimization methods converge most likely to the global minimum in nonlinear functions. Jílek (2002) and Simões-Filho et al. (1999) inversions are single interface parameter estimation algorithms and neither consider transmission effect nor extra information from the other ray paths which contribute to a specific reflector, in the favour of reducing dimensionality and nonlinearity.

Methodology

To address some of the above issues in this thesis, I provide a framework to incorporate all the information from different rays which passes through a specific layer. This is in contrast to the layer stripping inversion methods which rely only on the ray passing through the layer under estimation.

Most of the nonlinear geophysical inversions rely on the popular least square minimization methods, such as Levenberg-Marquardt, which require the prior model to be close to the solution. Instead, I implement the unconstrained minimization methods, such as variable metric and especially conjugate gradient methods, which are more tolerant to errors in prior information in estimating model parameters simultaneously from different layers. I also provide a simultaneous least square inversion model.

I provide two prestack inversion algorithms for P-wave and joint $P \rightarrow P \,\&\, P \rightarrow S_v$ waves: traveltime and waveform inversion. Both inversions use the exact traveltimes based on ray parameter in horizontally layered transversely isotropic media given by Ursin and Stovas (2006). Layers are considered as laterally homogeneous. I compute the ray parameter by inversion of the offset equation using the Newton method. To be able to compute the derivatives of reflection and transmission coefficients analytically, I use the Rüger's equations (Rüger, 1996) which are computationally fast and valid for pre-critical incidence angles. Traveltime inversions are used to estimate Thomsen's anisotropy parameters and layer thicknesses while waveform inversion is mostly used in estimation of Thomsen's anisotropy parameters.

I formulate the new offset and exact traveltime equations in a dipping transversely isotropic layered medium based on ray parameters. I also use these equations to estimate the Thomsen's anisotropy parameters from inversion of $P \rightarrow P$ and joint inversion of

$P \to P$ & $P \to S_v$ traveltimes. Moveout based anisotropic parameter estimations only assumes a single dipping layer at the bottom of the model.

Thesis outline

In the first chapter I review seismic wave propagation in anisotropic media particularly in a transversely isotropic medium. The solution of Christoffel equations in a transversely isotropic medium will be discussed. At the end of this chapter the vertical slowness and polarization directions of different modes from few examples will be demonstrated.

In the second chapter I review the exact traveltime and offset equations for a horizontal VTI medium for $P \to P$ and $P \to S_v$ waves. I use the Newton method to estimate the ray parameter by inversion of the offset equations from a random prior model. I also derive the new offset and exact traveltime equations in a dipping transversely isotropic layered medium for $P \to P$ and $P \to S_v$ waves. Because the law of conservation of horizontal slowness is no longer valid in a dipping layer, I estimate the first ray parameter in the interface of the first two top layers and the other ray parameters will be computed successively based on Snell law at the interface of two media.

In the third chapter I review the computation of approximate reflection and transmission coefficients based on the Rüger's equations. To understand the limitation of the approximate equations I compute the exact plane wave reflection and transmission coefficients and compare them with approximate ones at the interface between two VTI layers. I also compute the offset gathers of synthetic seismograms from a multilayered VTI medium for both approximate and exact reflectivity equations for compressional and joint compressional and shear wave wavefield respectively.

In chapter four I briefly explain the anatomy of an inverse problem and review the nonlinear least square inversion and unconstrained minimization algorithms which have been used in this thesis. The significance of preconditioning conjugate gradient in fast convergence will be discussed. Different conjugate gradient algorithms and their impact on the objective function will be discussed. I compare the variable metric and

preconditioning conjugate gradient algorithms in terms of their efficiency in fast convergence from an example of a traveltime inversion.

In chapter five I show how to compute the analytic partial derivatives in a system of layered media. I demonstrate the equations of partial derivatives of traveltimes and effective reflection coefficient with respect to model parameters which will be followed, respectively by the derivative of their objective functions. At the end of this chapter I compare the computation of the analytic and numerical partial derivatives of the objective function of the effective reflection coefficient with respect to the model parameters for a given layered VTI model.

In chapter six the numerical results from the unconstrained nonlinear inversion of $P \to P$ and joint $P \to P \,\&\, P \to S_v$ traveltimes in horizontal VTI media will be demonstrated. I investigate the effect of random prior information from a uniform distribution on the estimated Thomsen's anisotropy parameters. This will be followed by nonlinear waveform inversion of $P \to P$ and joint $P \to P \,\&\, P \to S_v$ respectively. At the end of this chapter the traveltime inversion in dipping transversely isotropic layered media will be discussed using an example of the preconditioning conjugate gradient method for both $P \to P$ and joint $P \to P \,\&\, P \to S_v$ prestack data.

At the end of this thesis I address the potential future works and the conclusion.

References

Alford, R. M., 1986, Shear data in the presence of azimuthal anisotropy: Presented at the Annual International Meeting, Society of Exploration Geophysicists, 476-489.

Crampin, S., and D. Bamford, 1977, Inversion of P-wave velocity anisotropy: Geophys. J. R. astr. Soc, **49**, 123-132.

Grechka, V., 2001, Seismic anisotropy: Yesterday, Today, Tomorrow: CSEG Recorder, **26**, 9-10.

Grechka, V., S. Theophanis, and I. Tsvankin, 1999, Joint inversion of P- and PS-waves in orthorhombic media: Theory and a physical modeling study: Geophysics, **64**, 146-161.

Grechka, V., and I. Tsvankin, 1998, Feasibility of nonhyperbolic moveout inversion in transversely isotropic media: Geophysics, **63**, 957-969.

Grechka, V., and I. Tsvankin, 1999, 3-D moveout velocity analysis and parameter estimation for orthorhombic media: Geophysics, **64**, 820-837.

Grechka, V., and I. Tsvankin, 2000, Inversion of azimuthally dependent NMO velocity in transversely isotropic media with a tilted axis of symmetry: Geophysics, **65**, 232-246.

Jílek, P., 2002, Modeling and inversion of converted-wave reflection coefficients in anisotropic media: A tool for quantitative AVO analysis: PhD, Colorado School of Mines.

Li, R., 2002, Inversion for the elastic parameters of layered transversely isotropic media: PhD, Curtin University of Technology.

Miller, D. E., S. Leaney, and W. H. Borland, 1994, An in situ estimation of elastic moduli fro a submarine shale: Journal of Geophysical Research, **99**, 21659-21665.

Miller, D. E., and C. Spencer, 1994, An exact inversion for anisotropic moduli from phase slowness data: Journal of Geophysical Research, **99**, 21651-21657.

Rüger, A., 1996, Reflection coefficients and azimuthal AVO analysis in anisotropic media: PhD, Colorado School of Mines.

Schoenberg, M., and J. S. Protázio, 1992, Zoeppritz rationalized and generalized to anisotropy: Journal of Seismic Exploration, **1**, 125-144.

Simões-Filho, I. A., F. A. Neves, J. S. Tinen, J. S. Protázio, and C. J. C., 1999, Multiazimuthal modeling and inversion of qP reflection coefficients in fractured media: Geophysics, **64**, 1143-1152.

Thomsen, L., 1986a, Weak elastic anisotropy: Geophysics, **51**, 1954-1966.

Thomsen, L., 1993, Weak anisotropic reflections, *in* J. P. Castagna and M. E. Backus, eds., Offset-dependent reflectivity-Theory and practice of AVO analysis: Society of Exploration Geophysicists, 1, 103-111.

Thomsen, L., 2001, Seismic anisotropy: Geophysics, **66**, 40-41.

Thomsen, L., I. Tsvankin, and M. C. Mueller, 1999, Coarse-layer stripping of vertically variable azimuthal anisotropy from shear-wave data: Geophysics, **64**, 1126-1138.

Tsvankin, I., 1997, Reflection moveout and parameter estimation for horizontal transverse isotropy: Geophysics, **62**, 614-629.

Tsvankin, I., and L. Thomsen, 1995, Inversion of reflection traveltimes for transverse isotropy: Geophysics, **60**, 1095-1107.

Ursin, B., and A. Stovas, 2006, Traveltime approximations for a layered transversely isotropic medium: Geophysics, **71**, D23-D33.

Chapter 1

Review of seismic anisotropy

Introduction

Seismic anisotropy is defined in different ways by different authors (Crampin, 1989; Winterstien, 1990) but perhaps the Thomsen's definition (2002) is the simplest and more intuitive one: "*the dependence of seismic velocity upon angle*". In geological areas where there are massive shales or where Carbonate or Sandstones formations are fractured, Thomsen's definition is well defined.

Despite the formalism of the dynamics of elastic waves in anisotropic media developed in 19th century by Kelvin (Musgrave, 1954a; Helbig and Thomsen, 2005), elastic anisotropy has received little attention until the works of Musgrave in the mid 1950's (Musgrave, 1954a; , 1954b; , 1959). However, for almost two decades this did not attract the attention of seismologist up until the works of Crampin and his colleagues, particularly in the context of earthquake seismology (Bamford and Crampin, 1977; Crampin, 1977; Keith and Crampin, 1977). Seismic anisotropy in transversely isotropic media has been further studied in exploration seismology by Berryman (1979) and Levin (1979). It has also been studied and implemented by Amoco since the early 1980's but it was officially announced in the famous Amoco "Anisotropy Session" at the 1986 SEG convention (Alford, 1986; Lynn and Thomsen, 1986; Rai and Hanson, 1986; Willis et al., 1986; Thomsen, 1986b). Parametrizing the elastic stiffness coefficients in transverse isotropic media into five parameters (vertical compressional and shear wave velocities, δ, ε, and γ) by Thomsen (1986a) was a significant progress in simplifying seismic velocity formulations. These are called as Thomsen's anisotropy parameters. The new anisotropy parameters reveal more intuitive understanding of the seismic velocity of different modes in anisotropic media. A comprehensive history of seismic anisotropy is given by Helbig and Thomsen (2005).

In this chapter I review the theory of wave propagation in anisotropic media with emphasize on body waves in transverse isotropy. We start with the Hooke's law to relate stress to strain through the stiffness tensor in an elastic transversely isotropic medium.

Later I introduce Christoffel matrix to compute the body wave slowness and polarization direction in a VTI medium. I discuss phase and group velocity and at the end of this chapter; to have a more physical insight to wave propagation through anisotropic media, several examples from the sedimentary rocks will be considered. We compute the phase and group velocities for the examples and also calculate the slowness surface and polarization directions for the compressional and shear waves.

1.1 Hooke's law and Stiffness Tensor

Hooke's law is named after 17th Century British physicist Robert Hooke. In Newtonian mechanics Hooke's law is an approximate linear model for elastic materials' behaviour under a load. It relates the amount of extension or compression (strain) of an element directly to the applied force (stress). In an elastic material the small deformation of the atomic structure will return to the stable state (minimal energy level of atoms). This is in fact the linearity limit of the mathematical Hooke's model. The stress-strain relationship of Hooke's law in a tensor notation is given as follows (Nye, 1993).

$$\sigma_{ij} = c_{ijkl}\varepsilon_{kl} \ , \tag{1.1}$$

where ε_{kl} is the second- order strain tensor and is defined as:

$$\varepsilon_{kl} = \frac{1}{2}\left(\frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k}\right) \ . \tag{1.2}$$

σ_{ij} is the second-order stress tensor. Both stress and strain tensors are symmetric tensors; ($\sigma_{ij} = \sigma_{ji}, \varepsilon_{kl} = \varepsilon_{lk}$; $i, j, k, l = 1, 2, 3$). u is the displacement vector. c_{ijkl} is the fourth-rank stiffness tensor and is a function of elastic properties of material. The stiffness tensor is a symmetric tensor $(c_{ijkl} = c_{jikl} = c_{ijlk} = c_{jilk})$. The symmetry allows reducing the number of independent elements of the stiffness tensor from 81 to 21 in an arbitrarily anisotropic medium. The fourth-order stiffness tensor can be mapped into a two dimensional symmetric matrix $C_{ij} = c_{ijkl}$ of 36 elements (Voigt, 1928).

where

$ij \to m, \ kl \to n$

$11 \to 1, \quad 22 \to 2, \quad 33 \to 3, \quad 32 = 23 \to 4, \quad 31 = 13 \to 5, \quad 12 = 21 \to 6$

In the most general anisotropic solid (triclinic symmetry), C_{ij} reduces to 21 different elements, whereas in isotropic fluid to one element (Bulk modulus k). In the most realistic form of anisotropy in rocks, orthorhombic symmetry, C_{ij} reduces to 9 independent elements.

$$C_{ij} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{12} & C_{22} & C_{23} & 0 & 0 & 0 \\ C_{13} & C_{23} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} \end{pmatrix} . \tag{1.3}$$

C_{ij}'s have the unit of Pascal ($ML^{-1}T^{-2}$) and usually are expressed by GPa. An orthorhombic medium is characterized by three mutually orthogonal symmetry planes. A sandstone or carbonate rock unit with two vertical systems of fractures, mutually orthogonal, is an example of orthorhombic symmetry.

The general form of the wave equation in a homogeneous anisotropic media follows from the second Newton's law. Tensor notation of the wave equation for an infinitesimal element deformed by the stress field σ_{ij} in the Cartesian system is expressed as (Aki and Richards, 1980):

$$\rho \frac{\partial^2 u_i}{\partial t^2} - \frac{\partial \tau_{ij}}{\partial x_j} = f_i \quad i = 1,2,3 \ , \tag{1.4}$$

where ρ is density, and f is external force per unit volume. Substitution of Hooke's law (1.1) and stress-strain relation (1.2) into the equation (1.4) and assuming homogeneous stiffness coefficients, results in the wave equation for a homogenous anisotropic medium,

$$\rho \frac{\partial^2 u_i}{\partial t^2} - c_{ijkl} \frac{\partial^2 u_k}{\partial x_j \partial x_l} = f_i \ . \tag{1.5}$$

By dropping the external force f, equation (1.5) reduces to the homogeneous wave equation for plane waves,

$$\rho \frac{\partial^2 u_i}{\partial t^2} - c_{ijkl} \frac{\partial^2 u_k}{\partial x_j \partial x_l} = 0 \ . \tag{1.6}$$

A trial solution for equation (1.6) is represented by:

$$u_k = AU_k e^{i\omega(n_j x_j / V - t)} = AU_k e^{i\omega(p_j x_j - t)} \ , \tag{1.7}$$

where A is the amplitude factor, U_k is the unit polarization vector, ω is the angular frequency, V is the velocity of wave propagation (the phase velocity), n_j is the unit vector normal to the wavefront, p_j is defined as the slowness vector $(p_j = n_j / V)$.

Substituting equation (1.7) into equation (1.6) results in:

$$(c_{ijkl} n_l n_j - \rho V^2 \delta_{ik}) U_k = 0 \ . \tag{1.8}$$

This is the Christoffel equation for phase velocity V and polarization vector U. In matrix notation it is called Christoffel matrix (Musgrave, 1970):

$$\begin{pmatrix} G_{11} - \rho V^2 & G_{12} & G_{13} \\ G_{21} & G_{22} - \rho V^2 & G_{23} \\ G_{31} & G_{32} & G_{33} - \rho V^2 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = 0 \ , \tag{1.9}$$

where G_{ij} depends on stiffness and wave propagation direction:

$$G_{ik} = c_{ijkl} n_j n_l \ . \tag{1.10}$$

For the non-zero solution U in equation (1.9) the determinant of the symmetric positive definite Christoffel matrix should be zero:

$$\det[G_{ij} - \rho V^2 \delta_{ik}] = 0 \ . \tag{1.11}$$

This is a 3×3 classic eigenvalue problem. In an transversely isotropic medium with vertical axis of symmetry (VTI) for any direction n, the three eigenvalues of equation (1.11) correspond to one compressional (P-wave) and two shear waves "normal or vertical component (SV-wave) and parallel or horizontal component (SH-wave)". Each of the three normalized eigenvectors corresponding to each eigenvalue determines the polarization direction of each mode. Polarization directions are always normal to each other but not necessarily to the direction of propagation. In an isotropic medium the polarization direction of plane waves is parallel to the propagation direction of

compressional wave and normal to the direction of shear waves; however, in anisotropic media this is usually untrue for each direction.

Group velocity is the velocity in the direction of energy propagation or along the source receiver direction, whereas phase velocity is normal to the wavefront (Figure 1.1). The group velocity surface may be visualized as the constructive interference of plane wave wavefronts propagating from a point source (Winterstien, 1990). Depending on the degree of anisotropy, shape of the group velocity surface is usually more complicated than the shape of the phase velocity surface. In a TI medium a cusp may occur for the (SV-wave) group velocity surface resulting from a high curvature of its phase velocity (see Figures 1.4-1.6). Unlike the phase velocity which is directly computed from Christoffel equation, group velocity is a function of phase velocity and polarization vector. In dispersive isotropic media, group velocity is only a function of phase velocity and the variation of the phase velocity with the wave number, however, in anisotropic media, it also varies with the polarization direction. The general form of the group velocity is given as follow (Musgrave, 1970):

$$Vg_j = \frac{1}{\rho v} c_{ijkl} U_i U_k n_l \ , \tag{1.12}$$

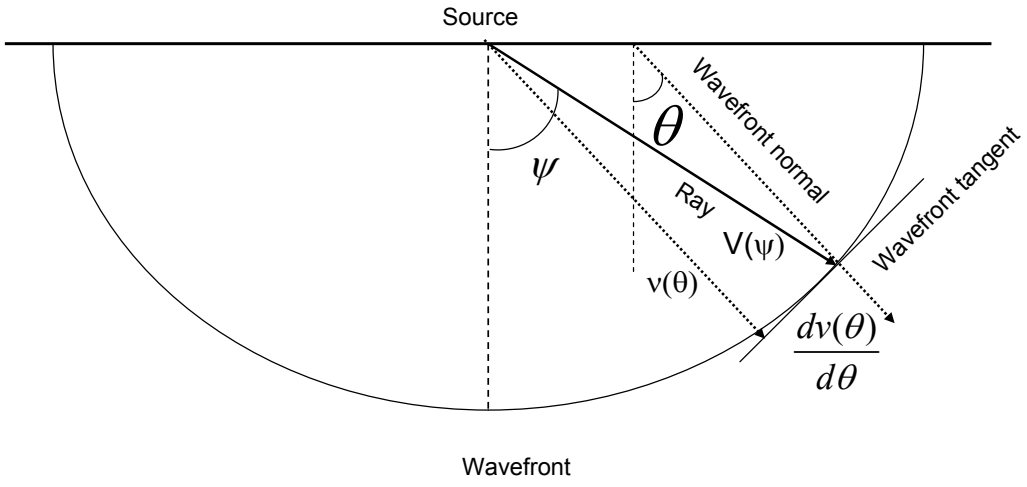
where Vg is the group velocity, v is the phase velocity, and U is the polarization vector. An angle dependent relation between group velocity (ray velocity) and phase velocity for a VTI medium is given by (Berryman, 1979):

$$V_g^2(\psi) = v^2 + \left(\frac{dv}{d\theta}\right)^2 , \tag{1.13}$$

where θ is the phase angle, the angle between normal to the wavefront and symmetry axis. Group (ray) angle ψ is the angle between group velocity vector and symmetry axis. In a VTI medium ray angle is given by (Thomsen, 1986a; Tsvankin, 2001).

$$\tan(\psi) = \left(\tan(\theta) + \frac{1}{v}\frac{dv}{d\theta}\right) \Big/ \left(1 - \frac{\tan(\theta)}{v}\frac{dv}{d\theta}\right) , \tag{1.14}$$

Figure (1.1) shows phase and ray angle and velocities for transversely isotropic medium with vertical axis of symmetry (VTI). Along the vertical and horizontal direction the phase and group angles are the same.



$$V^2(\psi) = \nu^2(\theta) + \left(\frac{dv(\theta)}{d\theta}\right)^2$$

Figure 1.1: Ray and phase angles in a homogeneous VTI (Hexagonal symmetry) medium, θ is the phase angle and ψ is the ray angle.

1.2 Seismic wave propagation in transversely isotropic (TI) media

Transversely isotropy (hexagonal symmetry) is the simplest form of anisotropy to model geological features and is widely used in geophysical applications. It is characterized by an isotropy plane and axis of symmetry normal to it. Any plane which contains axis of symmetry is a symmetry plane. In the isotropy plane the properties of that medium do not change with angle. If the axis of symmetry is rotated with respect to the coordinate axes it creates a general type of TI medium which is called tilted transversely anisotropy (TTI). Dipping transversely isotropic layers such as shales near the flanks of a diaper or in the folded and thrusted regions are examples of TTI media. Vertical transverse isotropy is a special case of a TTI in which the axis of symmetry is vertical (with respect to the earth surface). In some literature VTI is also has been called polar anisotropy (Thomsen, 2002). Shales very often show VTI behaviour. A homogenous isotropic stack of fine layers whose thicknesses are smaller than the seismic wavelength also show VTI behaviour. The effective properties of this medium are calculated by equivalent medium theory (Backus, 1962). This medium can be replaced by an equivalent homogeneous anisotropic medium. Horizontal transverse isotropy (HTI) is also a special case of TTI, in which the axis of symmetry is horizontal. Orthogonal parallel fractures in granitic rocks and sandstones cause a HTI behaviour. Figures (1.2) and (1.3) show schematic illustrations of a VTI and HTI medium.

Stiffness tensor in a VTI medium consists of 5 independent elements (1.15).

$$C_{ij} = \begin{pmatrix} C_{11} & C_{11}-2C_{16} & C_{13} & 0 & 0 & 0 \\ C_{11}-2C_{16} & C_{11} & C_{13} & 0 & 0 & 0 \\ C_{13} & C_{13} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} \end{pmatrix}, \qquad (1.15)$$

where the stability condition is (Musgrave, 1970):

$$\begin{aligned} &C_{11} \geq C_{66} \geq 0 \\ &C_{33} \geq 0 \\ &C_{44} \geq 0 \\ &C_{13}^2 \leq C_{33}(C_{11}-C_{66}) \end{aligned} \quad . \qquad (1.16)$$

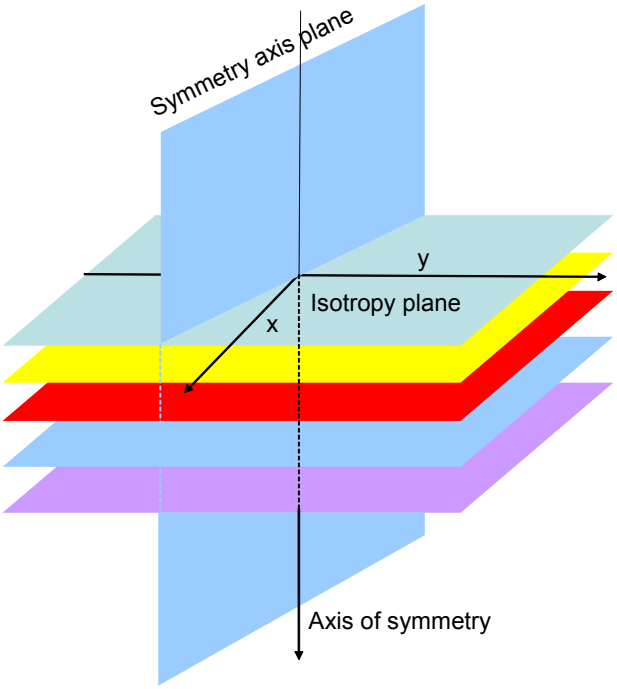


Figure 1.2: A schematic illustration of a VTI medium. Isotropy plane is horizontal and axis of symmetry is normal to it.

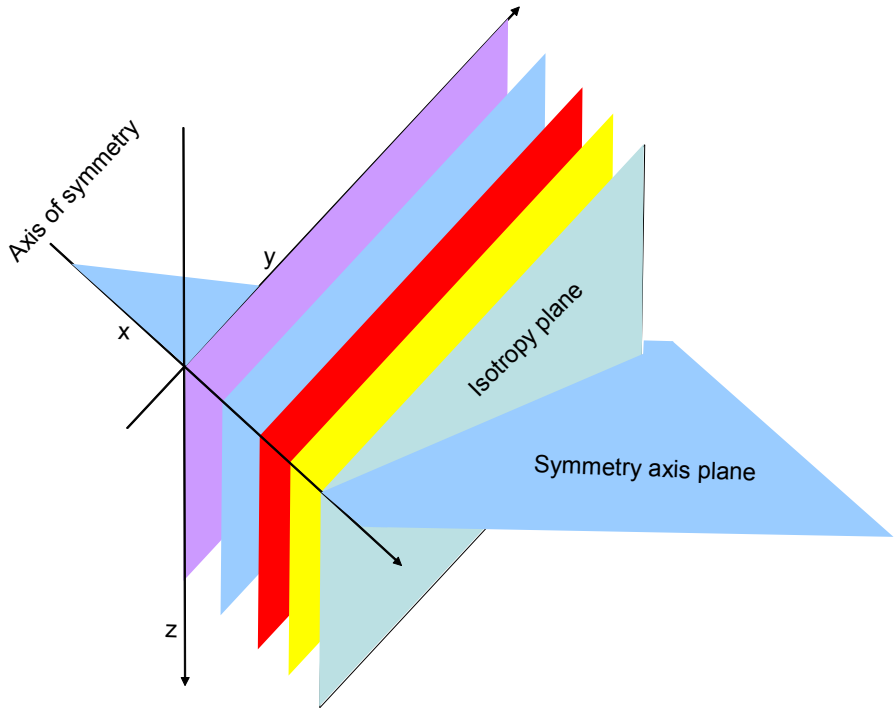


Figure 1.3: A schematic illustration of a HTI medium. Isotropy plane is vertical and axis of symmetry is normal to it.

Another way of representing the stiffness tensor in a VTI medium is given by Lamé constants (Chapman, 2004; Ikelle and Amundsen, 2005):

$$C_{ij} = \begin{pmatrix} \lambda_\perp + 2\mu_\perp & \lambda_\perp & \upsilon & 0 & 0 & 0 \\ \lambda_\perp & \lambda_\perp + 2\mu_\perp & \upsilon & 0 & 0 & 0 \\ \upsilon & \upsilon & \lambda_{\|} + 2\mu_{\|} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu_{\|} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu_{\|} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu_\perp \end{pmatrix}, \tag{1.17}$$

where $\perp$ and $\|$ refer to normal and parallel to the axis of symmetry. Lamé constants can be found by equating the elements of (1.17) into (1.15).

As for VTI medium, the stiffness tensor in a HTI medium consists of 5 independent elements,

$$C_{ij} = \begin{pmatrix} C_{11} & C_{13} & C_{13} & 0 & 0 & 0 \\ C_{13} & C_{33} & C_{33} - 2C_{44} & 0 & 0 & 0 \\ C_{13} & C_{33} - 2C_{44} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{66} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} \end{pmatrix}. \tag{1.18}$$

The stiffness tensor for a TTI medium can be found by rotating the stiffness tensor for a VTI medium normal to the symmetry axis. The stiffness tensor for a TTI medium as a function of dip angle is given by Ikelle and Amundsen (2005).

We mentioned that the eigenvalues and eigenvectors of the Christoffel matrix correspond to phase velocity and polarization direction of three modes (P-waves, SV-waves, and SH-waves) for VTI media. Since all orthogonal planes normal to the isotropy plane are similar, we can restrict the wave propagation to two dimensions (x-z plane). Now the cosine directions are:

$$\begin{aligned} &n_1 = \sin(\theta), \qquad n_2 = 0, \qquad n_3 = \cos(\theta) \\ &\theta = \text{angle between } n_3 \text{ and the wave propagation direction} \end{aligned}. \tag{1.19}$$

By substitution of the stiffness tensor for a VTI medium (1.15) into the equation (1.10), the G_{ij} coefficients are (Tsvankin, 2001):

$$\begin{aligned} G_{11} &= C_{11}n_1^2 + C_{66}n_2^2 + C_{55}n_3^2 \\ G_{22} &= C_{66}n_1^2 + C_{11}n_2^2 + C_{55}n_3^2 \\ G_{33} &= C_{55}(n_1^2 + n_2^2) + C_{33}n_3^2 \\ G_{12} &= (C_{11} - C_{66})n_1n_2 \\ G_{13} &= (C_{13} + C_{55})n_1n_3 \\ G_{23} &= (C_{13} + C_{55})n_2n_3 \end{aligned} \quad . \tag{1.20}$$

Substitution of the equation (1.20) in the Christoffel matrix (1.9) gives:

$$\begin{pmatrix} C_{11}n_1^2 + C_{55}n_3^2 - \rho V^2 & 0 & (C_{13} + C_{55})n_1n_3 \\ 0 & C_{66}n_1^2 + C_{55}n_3^2 - \rho V^2 & 0 \\ (C_{13} + C_{55})n_1n_3 & 0 & C_{55}n_1^2 + C_{33}n_3^2 - \rho V^2 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = 0 \ . \tag{1.21}$$

In a VTI medium, $C_{55} = C_{44}$, so I prefer to use C_{44} instead when it is needed.

Solving the equation (1.21), the phase velocity of a pure shear wave (V_{sh}) polarized in transverse direction can be found by solving the following equation for V :

$$(C_{66}\sin^2(\theta) + C_{44}\cos^2(\theta) - \rho V^2)U_2 = 0 \ , \tag{1.22}$$

$$V_{Sh}^2(\theta) = \frac{1}{\rho}\left[C_{66}\sin^2(\theta) + C_{44}\cos^2(\theta)\right] \ . \tag{1.23}$$

The phase velocity of the horizontal component of the shear wave in the vertical direction of propagation is $V_{sh}(0) = \sqrt{C_{44}/\rho}$, while in the horizontal direction is $V_{sh}(90) = \sqrt{C_{66}/\rho}$.

By letting $U_1 = U_3 \neq 0,\ U_2 = 0$, equation (1.21) turns to the following equation:

$$\begin{pmatrix} C_{11}\sin^2(\theta) + C_{44}\cos^2(\theta) - \rho V^2 & (C_{13} + C_{44})\sin(\theta)\cos(\theta) \\ (C_{13} + C_{44})\sin(\theta)\cos(\theta) & C_{44}\sin^2(\theta) + C_{33}\cos^2(\theta) - \rho V^2 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = 0 \ , \tag{1.24}$$

Solving this equation for the phase velocity (V) for P-waves and SV-waves gives:

$$V_p^2(\theta) = \frac{1}{2\rho}\left[(C_{11}+C_{44})\sin^2(\theta)+(C_{33}+C_{44})\cos^2(\theta)+D\right], \quad (1.25)$$

$$V_{Sv}^2(\theta) = \frac{1}{2\rho}\left[(C_{11}+C_{44})\sin^2(\theta)+(C_{33}+C_{44})\cos^2(\theta)-D\right], \quad (1.26)$$

where,

$$D = \left(\left[(C_{11}-C_{44})\sin^2(\theta)-(C_{33}-C_{44})\cos^2(\theta)\right]^2+4(C_{13}+C_{44})^2\sin^2(\theta)\cos^2(\theta)\right)^{\frac{1}{2}}.$$

The compressional velocity in the vertical direction, $Vp(0)=\sqrt{C_{33}/\rho}$, while in the horizontal direction $Vp(90)=\sqrt{C_{11}/\rho}$. This shows that compressional wave velocity in the vertical and horizontal directions are different unless $C_{11}=C_{33}$. The phase velocity for the SV-wave when it is propagating either in vertical or horizontal direction, is the same, $V_{sv}(0)=V_{sv}(90)=\sqrt{C_{44}/\rho}$. The phase velocity difference of the shear wave modes in the isotropy plane results in shear wave splitting.

Phase and group velocity surfaces can be illustrated by plotting phase and group velocities against their corresponding phase and group (ray) angles in a polar diagram. Slowness surface (inverse of the phase velocity) can be shown in the same way as phase velocity. According to Musgrave (1970) and Helbig (1994), wave propagation direction is normal to the slowness surface. The polarization direction of each mode can be determined by eigenvectors of equation (1.21). In a homogenous isotropic medium the polarization direction of each mode is either parallel (P-waves) or orthogonal (S-waves) to the ray direction, whereas, in homogenous anisotropic media except for the vertical and horizontal ray directions (pure compressional and shear waves), there is no pure compressional and SV-wave modes; Therefore, they are called quasi compressional and quasi SV- waves modes (see Figures 1.4-1.6).

In anisotropic media, the equations governing seismic wave propagation expressed based on the stiffness tensor, are more complicated than for isotropic media. The importance of anisotropy is usually hidden in these equations and can't be intuitively understood. Thomsen (Thomsen, 1986a) introduced simpler notations which yield much insight to wave equation in anisotropic media. In a VTI medium five elastic coefficients

can be replaced by the following three Thomsen's parameters $(\delta, \varepsilon, \gamma)$ and vertical velocity of compressional and shear waves:

$$\alpha_0 = \sqrt{\frac{C_{33}}{\rho}} \ , \tag{1.27}$$

$$\beta_0 = \sqrt{\frac{C_{44}}{\rho}} \ , \tag{1.28}$$

$$\varepsilon = \frac{C_{11} - C_{33}}{2C_{33}} \ , \tag{1.29}$$

$$\delta = \frac{(C_{13} + C_{44})^2 - (C_{33} - C_{44})^2}{2C_{33}(C_{33} - C_{44})} \ , \tag{1.30}$$

$$\gamma = \frac{C_{66} - C_{44}}{2C_{44}} \ , \tag{1.31}$$

where α_0 is vertical P-wave velocity and β_0 is vertical shear wave velocity, $\varepsilon, \delta,$ and γ are dimensionless. Rearranging the equations (1.25), (1.26), and (1.23) in terms of Thomsen's parameters gives:

$$V_p^2(\theta) = \alpha_0^2 \left[1 + \varepsilon \sin^2(\theta) + D' \right] \ , \tag{1.32}$$

$$V_{Sv}^2(\theta) = \beta_0^2 \left[1 + \varepsilon \frac{\alpha_0^2}{\beta_0^2} \sin^2(\theta) - \frac{\alpha_0^2}{\beta_0^2} D' \right] \ , \tag{1.33}$$

$$V_{Sh}^2(\theta) = \beta_0^2 \left[1 + 2\gamma \sin^2(\theta) \right] \ , \tag{1.34}$$

where

$$D' = \frac{1 - \beta_0^2 / \alpha_0^2}{2} \left[\left\{ 1 + \frac{4(2\delta - \varepsilon)}{1 - \beta_0^2 / \alpha_0^2} \sin^2(\theta) \cos^2(\theta) + \frac{4(1 - \beta_0^2 / \alpha_0^2 + \varepsilon)\varepsilon}{(1 - \beta_0^2 / \alpha_0^2)^2} \sin^4(\theta) \right\}^{\frac{1}{2}} - 1 \right] .$$

For $\theta = 90$ equation (1.32) reduces to:

$$V_p^2(90) = V_p^2(0)\left(1 + 2\varepsilon\right) \ . \tag{1.35}$$

An approximate Taylor expansion of (1.35) gives:

$$\varepsilon \approx \frac{V_p(90) - V_p(0)}{V_p(0)} \ . \tag{1.36}$$

Equation (1.36) shows why ε is usually called as P-waves anisotropy. The same procedure with V_{Sh} in equation (1.34) results in:

$$\gamma \approx \frac{V_{Sh}(90) - V_{Sh}(0)}{V_{Sh}(0)} \ . \tag{1.37}$$

γ is usually called the SH-wave anisotropy.

The definition of δ, however, is not as clear as ε and γ from (1.30). Taking the second derivative of V_p with respect to $\theta = 0$ in equation (1.32) results in (Tsvankin, 2001):

$$\left. \frac{d^2 V_p(\theta)}{d\theta^2} \right|_{\theta=0} = 2V_p(0)\delta \ . \tag{1.38}$$

Since $\left. \frac{dV_p}{d\theta} \right|_{\theta=0} = 0$, this shows that away from the vertical direction V_p will increase if δ is positive and decrease if δ is negative. In processing of seismic reflection data δ has more influence than ε and is directly connected to normal moveout velocity. The linearized equations of reflectivity (Thomsen, 1993; Rüger, 1996) show that δ has a first-order effect on reflection coefficients in smaller angles of incidence, while ε has more influence at larger angles of incidence. Following is the approximate moveout velocity (Thomsen, 1986a):

$$V_{pNMO} \approx V_p(0)\left(1+\delta\right) \ . \tag{1.39}$$

In moderately anisotropic sedimentary rocks ε varies from 0.1-0.3 and even in compacted shale formations it may be 0.4-0.5 (Thomsen, 1986a; Alkhalifah, 1996). Typical values for δ are in the order of 0.05-0.2, while in a TI medium interbedding of thin isotropic layers produces small negative values (Berryman et al., 1999).

Transformation from stiffness coefficients (C_{ij}) to $\alpha_0, \beta_0, \delta, \varepsilon, \gamma$ is straightforward but the inverse transformation is not unique for C_{13} unless the sign of $(C_{13} + C_{44})$ is specified, since the stability condition (1.16) requires C_{44} be always positive (elastic medium) except for anomalous large values C_{44} the sign of $(C_{13} + C_{44})$ is always positive.

$$C_{33} = \rho\alpha_0^2 \ , \tag{1.40}$$

$$C_{44} = \rho\beta_0^2 \ , \tag{1.41}$$

$$C_{11} = (2\varepsilon + 1)C_{33} \ , \tag{1.42}$$

$$C_{13} = \left[\left(C_{33} - C_{44} \right) \left(\left(2\delta + 1 \right) C_{33} - C_{44} \right) \right]^{\frac{1}{2}} - C_{44} \ . \tag{1.43}$$

1.3 Numerical examples

To demonstrate the seismic wave propagation pattern and polarization direction in a transversely isotropic medium with vertical axis of symmetry, three examples (data after Thomsen, 1986) from the VTI media are illustrated in Figures 1.4-1.6 in a polar coordinate system. The vertical axis corresponds to the symmetry axis and phase (θ) and group (ψ) angles are measure from it. These examples correspond to shale samples from Dog Creek shale, Mesaverde shale, and Green River shale respectively. Phase (solid-blue) and group (dashed-red) velocities for P-wave, SV-wave, and SH-wave are illustrated in Figure 1.4-a, 1.5-a, and 1.6-a. Strong anisotropy has resulted in triplication effects (cusps) on SV-wave phase velocity (Figures 1.5(a) and 1.6(a)). Shear wave singularities (where the phase velocity or slowness surfaces of SV-wave and SH-wave touch each other) can be seen in all three Figures. In Figures 1.4-1.6 (b), slowness and polarization directions are illustrated for three modes. Polarization direction of SH-wave is in the horizontal direction (isotropy plane) and does not change with phase angle, however, the polarization directions of P-wave and SV-wave change as the phase angle changes. Except in $\theta = 0$ and $\theta = \pm 90$ in which, polarization vectors are parallel (P-wave) or orthogonal (SV-wave) to the ray direction (pure modes), other directions are neither parallel (Quasi P-wave) nor orthogonal (Quasi SV-wave). Table (1.1) illustrates the stiffness tensor for three shale samples mentioned above, measured in the lab with ultrasonic waves.

$$C_{ij} = \begin{pmatrix} 10.2 & 5.58 & 4.96 & & & \\ 5.58 & 10.2 & 4.96 & & & \\ 4.96 & 4.96 & 7. & & & \\ & & & 1.36 & & \\ & & & & 1.36 & \\ & & & & & 2.3 \end{pmatrix}$$

(a)

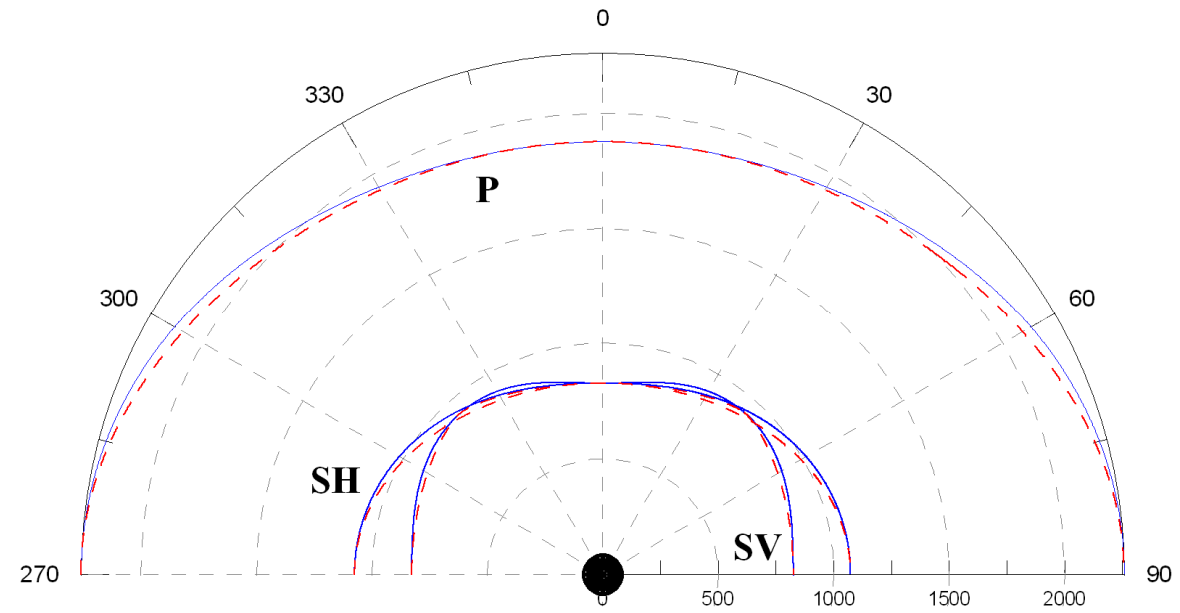
$$C_{ij} = \begin{pmatrix} 66.6 & 19.6 & 39.4 & & & \\ 19.6 & 66.6 & 39.4 & & & \\ 39.4 & 39.4 & 39.9 & & & \\ & & & 10.9 & & \\ & & & & 10.9 & \\ & & & & & 23.5 \end{pmatrix}$$

(b)

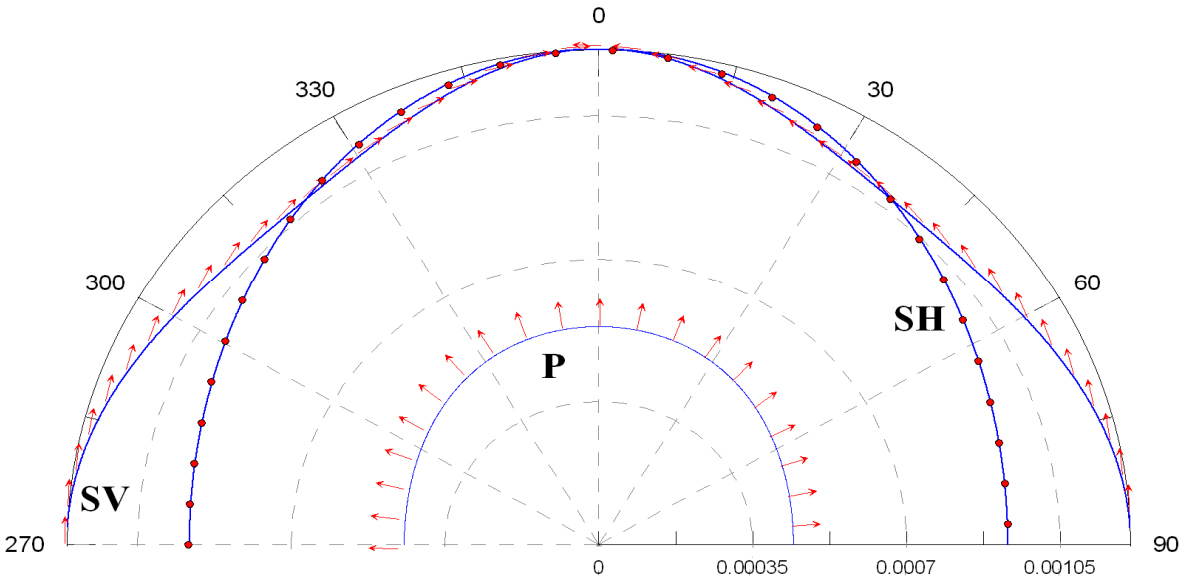
$$C_{ij} = \begin{pmatrix} 31.25 & 13.6 & 3.4 & & & \\ 13.6 & 31.25 & 3.4 & & & \\ 3.4 & 3.4 & 22.48 & & & \\ & & & 6.48 & & \\ & & & & 6.48 & \\ & & & & & 8.82 \end{pmatrix}$$

(c)

Table 1.1: Stiffness tensor for (a) Dog Creek Shale, (b) Mesaverde Shale, and (c) Green River Shale. Numbers are in GPa,

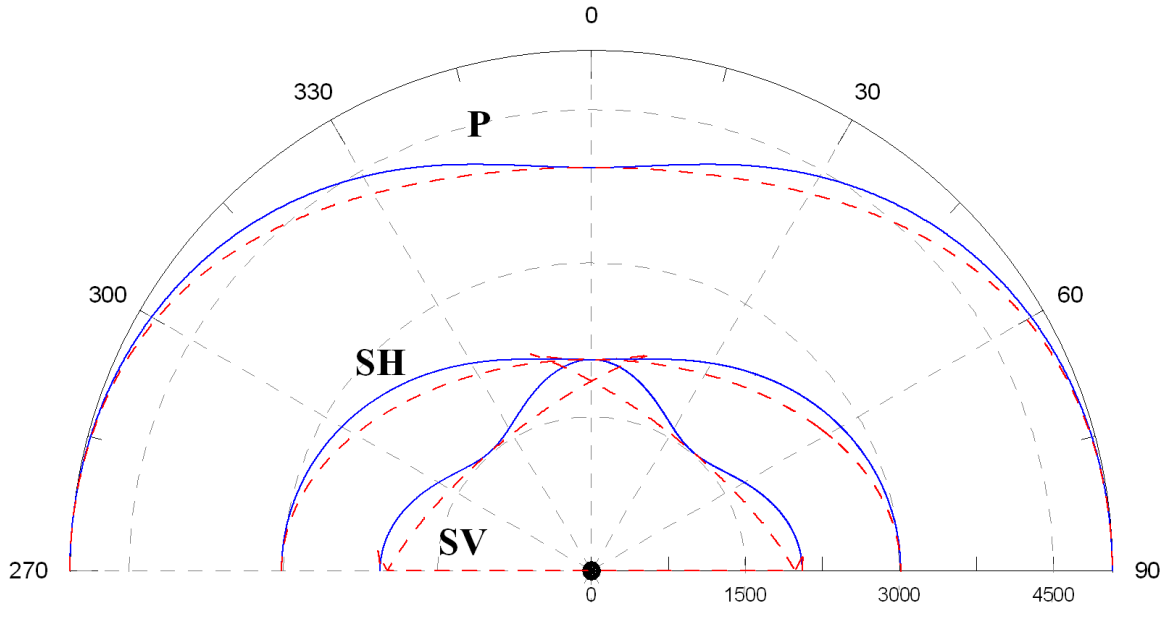


(a)

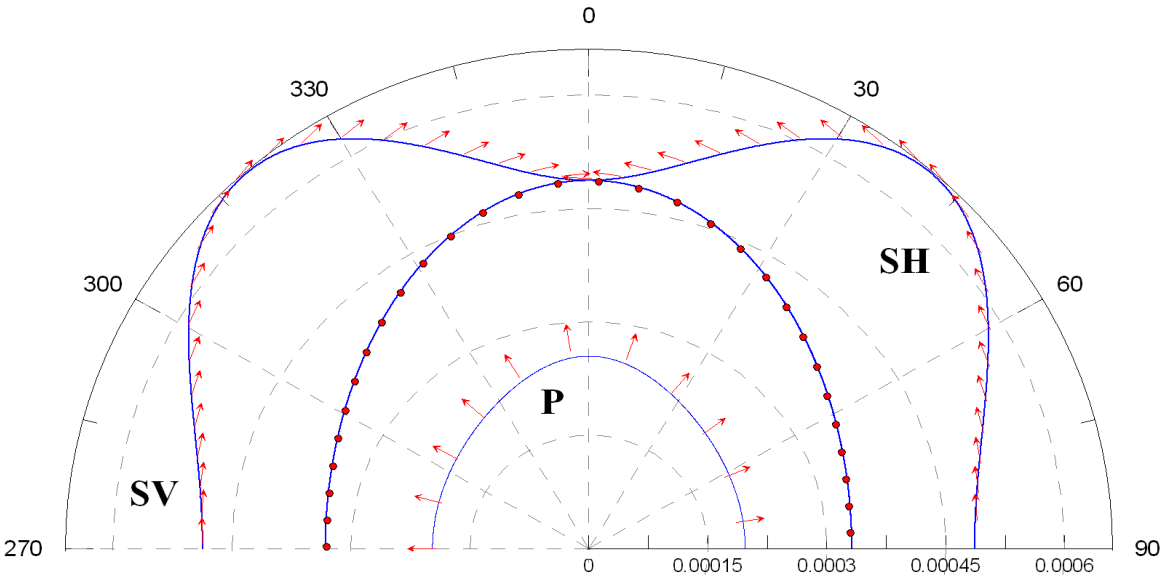


(b)

Figure 1.4: Seismic wave propagation and polarization pattern of P, SV, and SH waves for Dog Creek Shale (VTI anisotropy). (a) Phase velocity (blue curve) and group velocity (red curve) surfaces. (b) Slowness surfaces (blue) and polarization direction (red arrows). (Data after Thomsen, 1986). The elastic and anisotropic parameters are: $\alpha_0 = 1875\, m/s, \beta_0 = 826\, m/s, \rho = 2.00\, g/cm^3, \delta = 0.1, \varepsilon = 0.225, \gamma = 0.345$

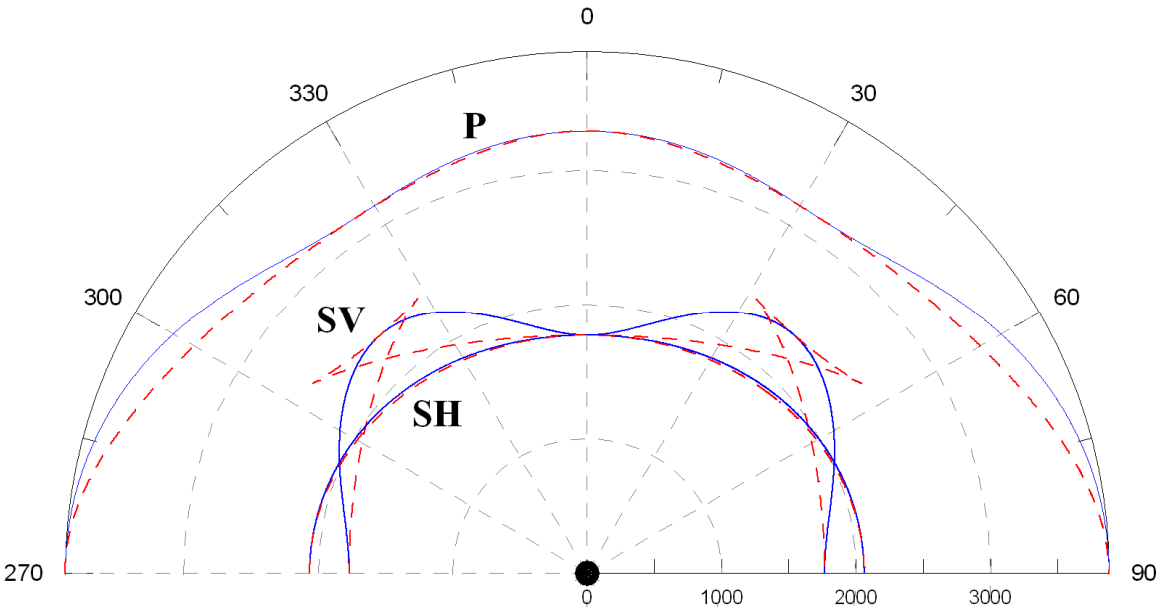


(a)

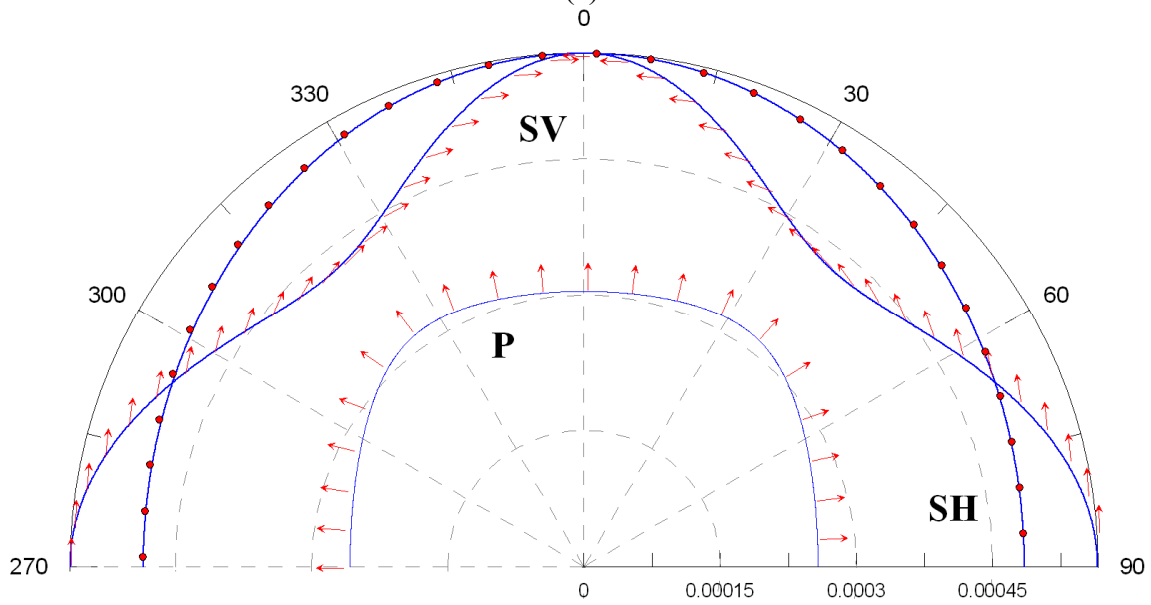


(b)

Figure 1.5: Seismic wave propagation and polarization pattern of P, SV, and SH waves for Mesaverde (5501) Shale (VTI anisotropy). (a) Phase velocity (blue curve) and group velocity (red curve) fields. Triplication effects (Cusps) are occurred for S_v near 0° and $\pm 90^\circ$ from vertical axis. (b) Slowness surfaces (blue) and polarization direction (red arrows). (Data after Thomsen, 1986). The elastic and anisotropic parameters are: $\alpha_0 = 3928\, m/s, \beta_0 = 2055\, m/s, \rho = 2.59\, g/cm^3, \delta = 0.73, \varepsilon = 0.334, \gamma = 0.575$



(a)



(b)

Figure 1.6: Seismic wave propagation and polarization pattern of P, SV, and SH waves for Green River Shale (VTI anisotropy). (a) Phase velocity (blue curve) and group velocity (red curve) fields. Triplication effects (Cusps) are occurred for S_v near 0° and $\pm 90^\circ$ from vertical axis. (b) Slowness surfaces (blue) and polarization direction (red arrows). (Data after Thomsen, 1986). The elastic and anisotropic parameters are: $\alpha_0 = 3292\, m/\sec, \beta_0 = 1768\, m/\sec, \rho = 2.075\, g/cm^3, \delta = -0.220, \varepsilon = 0.195, \gamma = 0.180$

References

Aki, K., and P. G. Richards, 1980, Quantitative Seismology: Theory and methods: W. H. Freeman.

Alford, R. M., 1986, Shear data in the presence of azimuthal anisotropy: Presented at the Annual International Meeting, Society of Exploration Geophysicists, 476-489.

Alkhalifah, T., 1996, Seismic processing in transversely isotropic media: PhD, Colorado School of Mines.

Backus, G. E., 1962, Long-wave elastic anisotropy produced by horizontal layering: Journal of Geophysical Research, **67**, 4427-4440.

Bamford, D., and S. Crampin, 1977, Seismic anisotropy - the state of the art: Geophys. J. R. astr. Soc, **49**, 1-8.

Berryman, J. G., 1979, Long-wave elastic anisotropy in transversely isotropic media: Geophysics, **44**, 896-917.

Berryman, J. G., V. Grechka, and P. A. Berge, 1999, Analysis of Thomsen parameters for finely layered VTI media: Geophysical Prospecting, **47**, 959-978.

Chapman, C. H., 2004, Fundamentals of seismic wave propagation: Cambridge University Press.

Crampin, S., 1977, A review of the effects of anisotropic layering on the propagation of seismic waves: Geophys. J. R. astr. Soc, **49**, 9-27.

Crampin, S., 1989, Suggestions for a consistent terminology for seismic anisotropy: Geophysical Prospecting, **37**, 753-770.

Helbig, k., 1994, Foundations of anisotropy for exploration seismics: Pergamon.

Helbig, k., and L. Thomsen, 2005, 75-plus years of anisotropy in exploration and reservoir seismics: A historical review of concepts and methods: Geophysics, **70**, 9ND-23ND.

Ikelle, T. L., and L. Amundsen, 2005, Introduction to petroleum seismology: Society of Exploration Geophysics.

Keith, C. M., and S. Crampin, 1977, Seismic body waves in anisotropic media: reflection and refraction at a plane interface: Geophys. J. R. astr. Soc, **49**, 181-208.

Levin, F. K., 1979, Seismic velocities in transversely isotropic media: Geophysics, **44**, 918-936.

Lynn, H., and L. Thomsen, 1986, Shear-wave exploration along the principle axis: Presented at the Annual International Meeting, Society of Exploration Geophysicists, 473-476.

Musgrave, M. J. P., 1954a, On the propagation of elastic waves in aeolotropic media. I. General principles: Proc. R. Soc. A., **226**, 339-355.

Musgrave, M. J. P., 1954b, On the propagation of elastic waves in aeolotropic media. II. Media of hexagonal symmetry: Proc. R. Soc. A., **226**, 356-366.

Musgrave, M. J. P., 1959, The propagation of elastic waves in crystals and other anisotropic media: Rep. Prog. Phys, **22**, 74-96.

Musgrave, M. J. P., 1970, Crystal acoustics: introduction to study of elastic waves and vibrations in crystals: Holden-day.

Nye, J. F., 1993, Physical properties of crystals: their representation by tensors and matrices: Clarendon Press.

Rai, C., and k. E. Hanson, 1986, Shear-wave birefringence: A laboratory study: Presented at the Annual International Meeting, Society of Exploration Geophysicists, 471-473.

Rüger, A., 1996, Reflection coefficients and azimuthal AVO analysis in anisotropic media: PhD, Colorado School of Mines.

Thomsen, L., 1986a, Weak elastic anisotropy: Geophysics, **51**, 1954-1966.

Thomsen, L., 1986b, Reflection seismology in azimuthally anisotropic media: Presented at the Annual International Meeting, Society of Exploration Geophysicists, 468-470.

Thomsen, L., 1993, Weak anisotropic reflections, *in* J. P. Castagna and M. E. Backus, eds., Offset-dependent reflectivity-Theory and practice of AVO analysis: Society of Exploration Geophysicists, 1, 103-111.

Thomsen, L., 2002, Understanding seismic anisotropy in exploration and exploitation: Society of Exploration Geophysics.

Tsvankin, I., 2001, Seismic signature and analysis of reflection data in anisotropic media: Pergamon.

Voigt, W., 1928, Lehrbuch der Kristallphysik: Teubner Verlag.

Willis, H., G. Rethford, and E. Bielanski, 1986, Azimuthal anisotropy: Occurrence and effect on shear wave data quality: Presented at the Annual International Meeting, Society of Exploration Geophysicists, 479-481.

Winterstien, D. F., 1990, Velocity anisotropy terminology for geophysicists: Geophysics, **55**, 1070-1088.

Chapter 2

Ray tracing in transversely isotropic media

Introduction

In the previous chapter we showed the Christoffel matrix based on the stiffness tensor elements and the directional cosines as a function of phase angle and I expressed phase velocities as the eigenvalues of this matrix. In this chapter I will express seismic wave propagation in terms of ray parameter which will subsequently be used in ray tracing in a homogeneous transversely isotropic layered medium. This approach has been widely used in elastic and viscoelastic isotropic and general anisotropic media (Daley and Hron, 1977; Chapman and Woodhouse, 1981; Ursin, 1983; Graebner, 1992; Chapman, 1994; Ĉervenŷ, 2001; Stovas and Ursin, 2003; Ikelle and Amundsen, 2005). Using the offset equation given by Ursin and Stovas (2006), I will optimize an objective function composed of the calculated and measured offsets to estimate a ray parameter for every ray path for both $P \to P$ and converted $P \to S_v$ waves. This will be done using the Newton method of minimization (Nocedal and Wright, 1999). At the end of this chapter I will formulate the new offset and traveltime equations for a dipping transversely isotropic medium and estimate ray parameter for each ray path for both $P \to P$ and converted $P \to S_v$ waves.

2.1 Parametric wave propagation in terms of ray parameter

In a transversely isotropic medium with a vertical axis of symmetry we ignore the transverse direction and restrict ourselves to the vertical plane (so we only consider one component of horizontal slowness $p_1 = p, \quad p_2 = 0$). Fourier transform of the equation of motion and Hooke's law for a VTI medium can be written as a system of $2n$ linear ordinary differential equation expressed by a $2n \times 2n$ coefficient matrix which is partitioned into the 4 $n \times n$ submatrices (Chapman, 1994; Stovas and Ursin, 2003):

$$\frac{db}{dz} = \omega \begin{pmatrix} 0 & A \\ B & 0 \end{pmatrix} b \ , \tag{2.1}$$

where $b = \left[\omega U_z, -S_r, S_z, \omega U_r\right]$, U_z and U_r are transformed vertical and horizontal displacements components, and S_r, S_z are transformed vertical and horizontal source components in the horizontal plane. A and B are real symmetric matrices in an elastic medium and are functions of slowness and stiffness coefficients,

$$A = \begin{pmatrix} C_{33}^{-1} & -C_{13}C_{33}^{-1}p \\ -C_{13}C_{33}^{-1}p & \rho - p^2(C_{11} - C_{13}^2 C_{33}^{-1}) \end{pmatrix}, \qquad B = \begin{pmatrix} -\rho & p \\ p & -C_{44}^{-1} \end{pmatrix}, \tag{2.2}$$

where ρ is the density.

A product matrix $H = AB$ will be used to extract vertical slowness using dispersion equations,

$$H = \begin{pmatrix} -\rho C_{33}^{-1} - C_{13}C_{33}^{-1}p^2 & pC_{33}^{-1} + C_{13}C_{33}^{-1}C_{44}^{-1}p \\ \rho C_{13}C_{33}^{-1}p + \rho p - p^2\left(C_{11} - C_{13}^2 C_{33}^{-1}\right)^2 & -C_{13}C_{33}^{-1}p^2 - \left(\rho - p\left(C_{11} - C_{13}^2 C_{33}^{-1}\right)^2\right)C_{44}^{-1} \end{pmatrix}. \tag{2.3}$$

$$q_\alpha^2 = \alpha^{-2} - p^2, \qquad q_\beta^2 = \beta^{-2} - p^2 \ . \tag{2.4}$$

The eigenvalues q of (2.3) after substitution with the Thomsen's anisotropy parameters are in fact vertical slowness for P-waves and SV-waves given by Stovas and Ursin (2003) and Ursin and Stovas (2006),

$$q^2 = -\frac{1}{2}\left[-q_{\alpha 0}^2 - q_{\beta 0}^2 + 2p^2(\sigma + \delta)\right] \tag{2.5}$$

$$\pm\sqrt{\left(q_{\beta 0}^2 - q_{\alpha 0}^2\right)^2 - 4\frac{p^2}{\alpha_0^2}\left(\gamma_0^2 - 1\right)(\sigma - \delta) + 4p^4\left[2\frac{\left(\gamma_0^2 - 1\right)}{\gamma_0^2}\sigma + (\sigma + \delta)^2\right]} \ ,$$

where $q_{\alpha 0}^2$ and $q_{\beta 0}^2$ are vertical slowness for P and S_v waves travelling vertically:

$$\gamma_0 = \alpha_0 / \beta_0 \ , \tag{2.6}$$

$$\sigma = \gamma_0^2 \left(\varepsilon - \delta \right) \ , \tag{2.7}$$

$$q_{\beta 0}^2 = \beta_0^{-2} + p^2 \ , \tag{2.8}$$

$$q_{\alpha 0}^2 = \alpha_0^{-2} + p^2 \ , \tag{2.9}$$

where α_0 and β_0 are the vertical compressional and shear wave velocities. By substitution of $q_{\alpha 0}$ and $q_{\beta 0}$ respectively in (2.5), the vertical slowness for P-waves and SV-waves become:

$$q_\alpha^2 = q_{\alpha 0}^2 - p^2 \left(\sigma + \delta \right) - \psi \ , \tag{2.10}$$

$$q_\beta^2 = q_{\beta 0}^2 - p^2 \left(\sigma + \delta \right) + \psi \ , \tag{2.11}$$

where

$$\psi = \frac{\left(\gamma_0^2 - 1 \right)}{2\alpha_0^2} \left[\sqrt{1 + bp^2\alpha_0^2 + cp^4\alpha_0^4} - 1 \right] , \tag{2.12}$$

$$b = -\frac{4}{\gamma_0^2 - 1}\left(\sigma - \delta \right), \quad c = \frac{4}{\left(\gamma_0^2 - 1 \right)^2} \left[\frac{2\left(\gamma_0^2 - 1 \right)}{\gamma_0^2} \sigma + \left(\sigma + \delta \right)^2 \right] .$$

The square root in (2.12) can be written in a Taylor series with a few terms as:

$$\begin{aligned} \sqrt{1 + bp^2\alpha_0^2 + cp^4\alpha_0^4} &= 1 + \frac{1}{2} bp^2\alpha_0^2 + \frac{1}{2}\left(c - \frac{1}{4} b^2 \right) p^4\alpha_0^4 \\ &- \frac{1}{4} b \left(c - \frac{1}{4} b^2 \right) p^6\alpha_0^6 - \frac{1}{8}\left(c - \frac{1}{4} b^2 \right)\left(c - \frac{5}{4} b^2 \right) p^8\alpha_0^8 + \ldots \end{aligned} \ . \tag{2.13}$$

Combining (2.10) and (2.8) results in the $P-Wave$ phase velocity:

$$\frac{1}{v_p^{\ 2}} = \frac{1}{\alpha_0^2} - p^2 H_\alpha \ , \tag{2.14}$$

where

$$H_\alpha = \sum_{j=0}^{\infty} a_j \left(p\alpha_0 \right)^{2j} \ . \tag{2.15}$$

The first few terms of a_j are given as:

$$
\begin{aligned}
a_0 &= 1+2\delta \\
a_1 &= 2(\varepsilon-\delta)\left(1+\frac{2\gamma_0^2\delta}{\gamma_0^2-1}\right) \\
a_2 &= -\frac{4}{\gamma_0^2-1}(\varepsilon-\delta)\left(1+\frac{2\gamma_0^2\delta}{\gamma_0^2-1}\right)(\delta-\sigma) \\
a_3 &= -\frac{4}{\gamma_0^2\left(\gamma_0^2-1\right)}(\varepsilon-\delta)\left(1+\frac{2\gamma_0^2\delta}{\gamma_0^2-1}\right)\left[\sigma\left(1+\frac{2\gamma_0^2\delta}{\gamma_0^2-1}\right)-\frac{2\gamma_0^2\left(\delta-\sigma\right)^2}{\gamma_0^2-1}\right].
\end{aligned}
\tag{2.16}
$$

By adding (2.10) and (2.11), then using (2.8) and (2.9), and combining with (2.14) we can express the SV-wave phase velocity as:

$$
\frac{1}{v_s^{\,2}} = \frac{1}{\beta_0^2} - p^2 H_\beta \ , \tag{2.17}
$$

where

$$
H_\beta = 2\sigma - \sum_{j=1}^{\infty} a_j \left(p\alpha_0\right)^{2j} \ . \tag{2.18}
$$

Equations (2.14) and (2.17) are the key equations which will be used for the ray tracing. We investigated the accuracy of equations (2.14) and (2.17) for wide varieties of Thomsen's anisotropy parameters and ray parameters. Equation (2.14) is accurate almost under any degree of anisotropy and every ray parameter, however, equation (2.17) seems to breaks down for larger ray parameters or phase angles of incidence. Since, in this thesis we are dealing with $P \to S_v$ rather than $S_v \to S_v$, and also SV-wave incidence angles from conversion are much smaller than P-wave phase angles of incidence, there is no concern about them during the ray tracing.

2.2 Traveltime and offset in a horizontally layered media

In this section we will represent the traveltime and offset equations for a homogeneous horizontally layered medium as a function of ray parameter for both compressional $(P \to P)$ and converted shear waves $(P \to S_v)$. We will still consider the medium as transversely isotropic with a vertical axis of symmetry. Figure 2.1 shows a typical configuration for ray tracing in the symmetry plane for both $P \to P$ and $P \to S_v$ rays. The classic ray theory will be used and, to do so, only two conditions need to be

met. The ray trajectories have to be in the symmetry plane in each layer and rays have to obey Snell's law (Hubral and Krey, 1980).

Let the source and receiver be separated by offset X on the earth's surface and let the traveltime of the ray path be T. Each ray elements will have the traveltime t and offset x in a layer with the thickness z (Figure 2.2). The traveltime T and offset X can be represented by the summation of all down going and up going ray elements.

$$T = \sum_{k=1}^{2n} \frac{z_k}{V_k cos(\alpha_k)} , \tag{2.19}$$

$$X = \sum_{k=1}^{2n} z_k tan(\alpha_k) . \tag{2.20}$$

where V_k is the ray velocity (either compressional or converted shear wave), α_k is the ray angle for each ray element, and n is the number of layers. As we saw in the previous section, wave propagation equations contain phase velocity rather than ray velocity, so the ray tracing equation has to be expressed by phase velocity and phase angle.

The following equation (Ĉervenŷ, 2001) along with (1.13) and their derivatives, after some trigonometric manipulations (Ursin and Hokstad, 2003), (Ursin and Stovas, 2006), map the ray angle and ray velocity to phase angle and phase velocity (see appendix A),

$$cos(\alpha - \theta) = \frac{v}{V} . \tag{2.21}$$

where α is the ray angle, θ is the phase angle, and v is phase velocity. By substitution of ray angle and ray velocity with phase angle and phase velocity, equations (2.19) and (2.20) can be expressed as (see Appendix B):

$$T = \sum_{k=1}^{2n} \frac{z_k}{v_k cos(\theta_k)} \left(1 + \frac{p}{v_k} v_k' \right) , \tag{2.22}$$

$$X = \sum_{k=1}^{2n} \frac{z_k v_k}{cos(\theta_k)} \left(p + \frac{v_k'}{v_k^3} \right) . \tag{2.23}$$

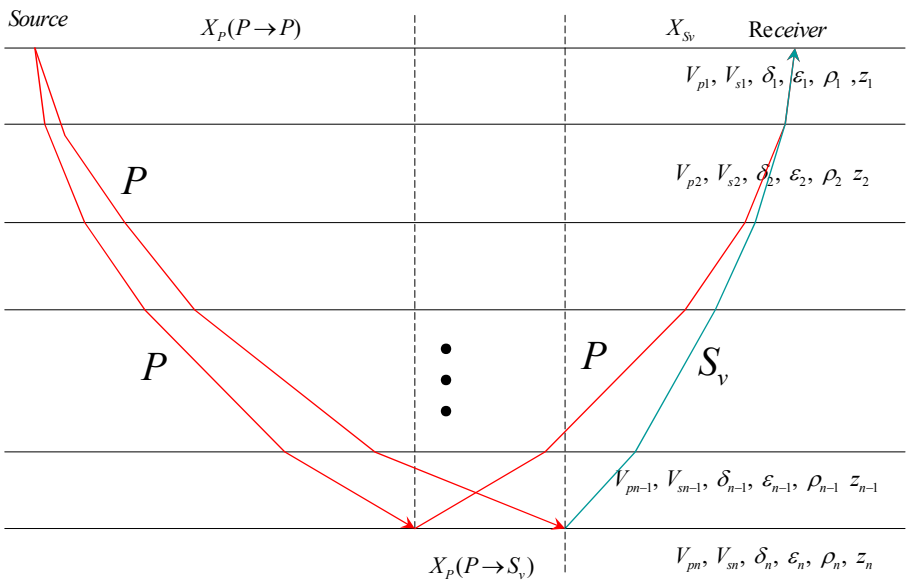


Figure 2.1: Ray paths for $P \to P$ and $P \to S_v$ travelling through horizontal anisotropic media (Z is the thickness).

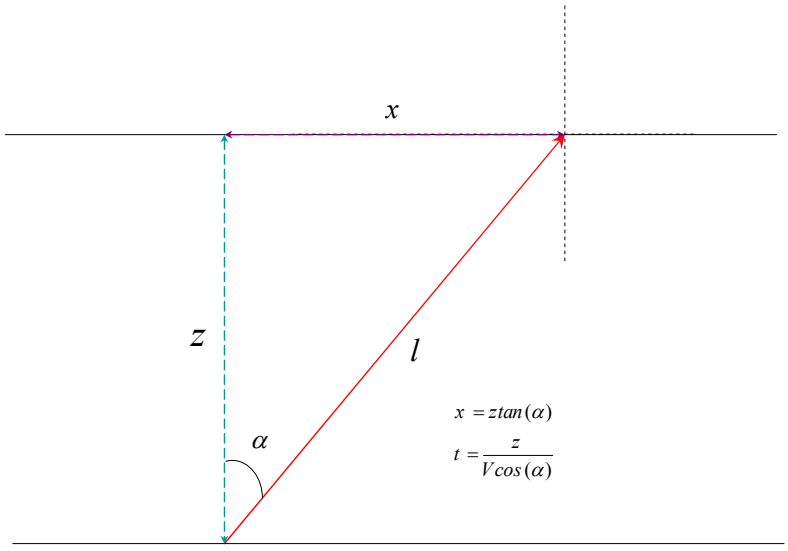


Figure 2.2: A ray element travelling upward through a horizontal layer

Equations (2.22) and (2.23) need to be expressed in terms of vertical velocities, so using the equations (2.14) and (2.17), and their derivative we rewrite equations (2.22) and (2.23) as (see appendix B):

$$T = \sum_{k=1}^{2n} \frac{z_k}{v_{0,k}} \frac{1 + \frac{1}{2} p^3 v_{0,k}^2 H_k'}{\sqrt{1 - p^2 v_{0,k}^2 \left(1 + H_k\right)}} \ , \tag{2.24}$$

$$X = p \sum_{k=1}^{2n} z_k v_{0,k} \frac{1 + H_k + \frac{1}{2} p H_k'}{\sqrt{1 - p^2 v_{0,k}^2 \left(1 + H_k\right)}} \ . \tag{2.25}$$

where v_0 is the vertical velocity (either compressional or converted shear waves) and $H' = dH / dp$. In the case of converted shear waves H and H' should be replaced by H_s and H_s' for the second leg of the ray path.

2.3 Two point ray tracing in a horizontally layered media

In the previous section we derived parametric traveltime and offset equations for the rays which travel in the symmetry plane of a VTI horizontally layered medium. Unfortunately there is no explicit method to find the ray parameter for a multilayered medium. In the ray tracing algorithm we seek a ray parameter in which a ray after shooting can reach the receiver within an error limit. We use Newton method (Nocedal and Wright, 1999) which is fast and works under almost any condition in terms of degree of anisotropy and velocity contrast. In the models we are dealing with in this thesis we assume there would be no triplications in SV-wavefronts. To optimize the ray parameter we choose a quadratic objective function (2.26) which is the square function of measured offset x_0 and calculated offset x by (2.25).

$$f = \left(x_0 - x\right)^2 \ . \tag{2.26}$$

The objective function is discontinuous (Figure 2.3) and for the cases, in which the ray parameter is close to the critical point, minimization is unstable and will break down. To avoid the instability, minimization will alternate to a logarithmic scaled objective function (2.27) and it will effectively stabilize the objective function (Figure 2.3).

$$f = \left(\ln \frac{x_0}{x} \right)^2 . \tag{2.27}$$

We draw a random ray parameter from a uniform distribution in a certain interval (0.00001 - 0.0001, the random ray parameter outside the interval will be rejected) and check for consistency with the following equation to avoid the imaginary part. So,

$$1 - p^2 v_{0,k}^2 \left(1 + H_k\right) , \tag{2.28}$$

only the ray parameters which results in positive values of (2.28) will be accepted. In the case of tracing the $P \to S_v$ rays, equation (2.28) also has to be satisfied for the converted shear wave part. This means that the H and v_0 terms has to be replaced with H_s and β_0. After selecting a ray parameter we calculate the offset according to equation (2.25) and the derivatives of offset with respect to ray parameter (see appendix 2. C). the prior ray parameter can be updated according to:

$$p_{n+1} = p_n - \mu \frac{\partial f}{\partial p} / \frac{\partial^2 f}{\partial p^2} , \tag{2.29}$$

where

$$\begin{aligned} \frac{\partial f}{\partial p} &= -2 \frac{\partial x}{\partial p} \left(x_0 - x \right) , \\ \frac{\partial^2 f}{\partial p^2} &= -2 \frac{\partial^2 x}{\partial p^2} \left(x_0 - x \right) + 2 \left(\frac{\partial x}{\partial p} \right)^2 , \end{aligned} \tag{2.30}$$

μ is the step length from the line minimization, and n is the iteration number. The Newton method usually converges to the minimum in a few iterations to 1.e-06 meter of accuracy (difference between the measured offset and calculated offset from the estimated ray parameter). Sometimes where the angle of incidence is very large (it occurs for the shallow reflectors in long offsets or where the velocity contrast is high), minimization may need to be restarted via a recursive algorithm.

To test the method we have used a numerical model of Ursin and Stovas (2006) after some modifications (Table 2.1) to calculate the traveltimes. This model was composed of 10 layers with different thicknesses. A large variety of VTI anisotropy was assigned to layers to investigate the effect of anisotropy on wave propagation. We will use this layered model as a benchmark to estimate its properties by different inversion

algorithms in later chapters. Figures 2.4 and 2.5 illustrate the ray parameters for $P \to P$ and $P \to S_v$ rays which have been estimated for the layered model (Table 2.1). The first receiver is kept at 200 meter away from the source and the rest of the receivers are spaced at every 50 meters. The maximum offset is 4950 metre. Figures 2.6 and 2.7 illustrate the traveltimes for $P \to P$ and $P \to S_v$ respectively which are calculated using the estimated ray parameters.

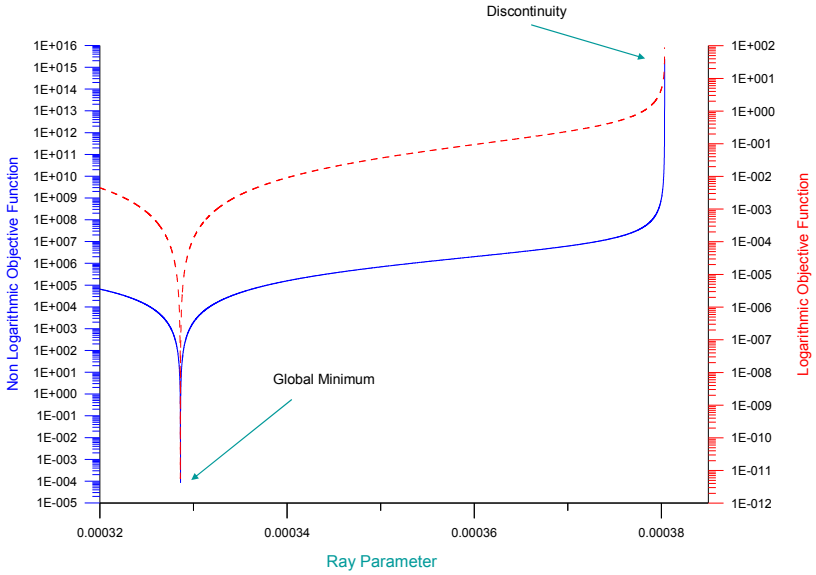


Figure 2.3: A typical ray parameter solution space.

Layer	$V_p(m/s)$	$V_s(m/s)$	δ	ε	$\rho(g/cm^3)$	$\Delta z(metre)$
1	1940	780	0.03	0.1	2.03	1000
2	2140	860	-0.02	0.14	2.06	160
3	2220	890	-0.05	0.1	2.08	90
4	2000	1000	0.1	0.14	2.04	40
5	1990	990	0.05	0.1	2.04	100
6	1900	950	0.04	0.12	2.02	190
7	2200	1150	0.06	0.18	2.08	270
8	2050	1130	0.1	0.2	2.05	170
9	2650	1500	0.07	0.1	2.15	310
10	2750	1530	0.1	0.14	2.17	300

Table 2.1: Layer properties used to calculate $P \to P$ and $P \to S_v$ traveltimes.

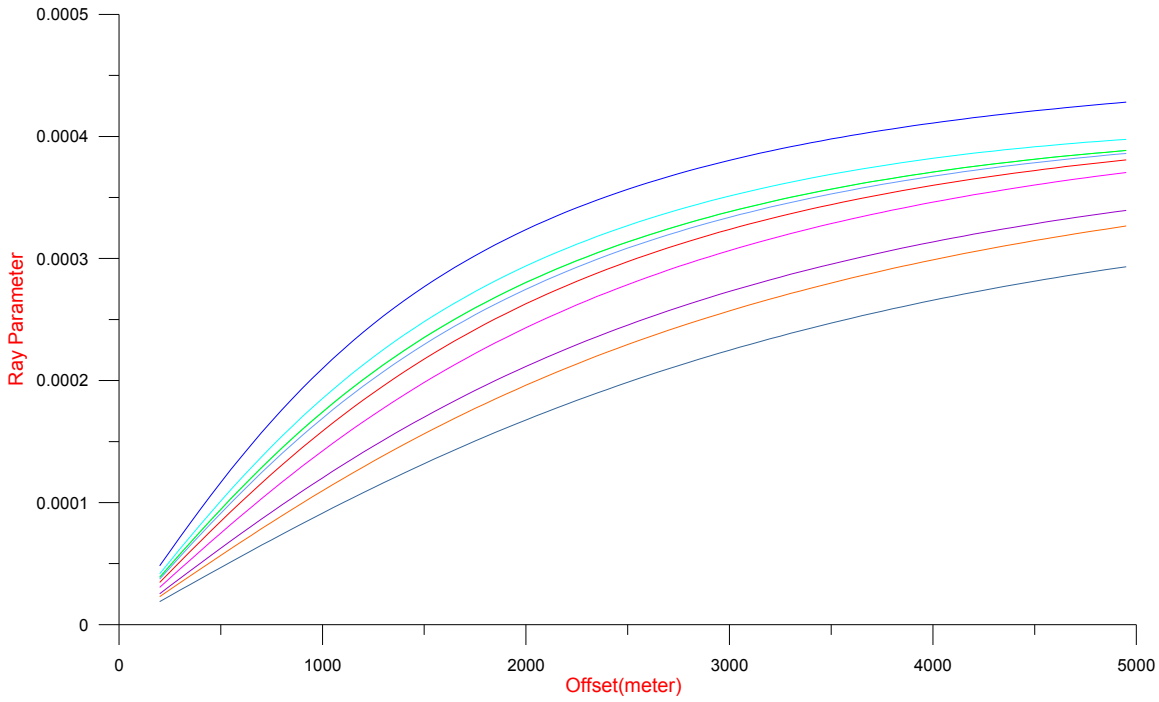


Figure 2.4: Ray parameters estimated for $P \to P$ rays from the horizontal layered media (Table 2.1). The top curve corresponds to layer 1, while the lowest one corresponds to layer 9.

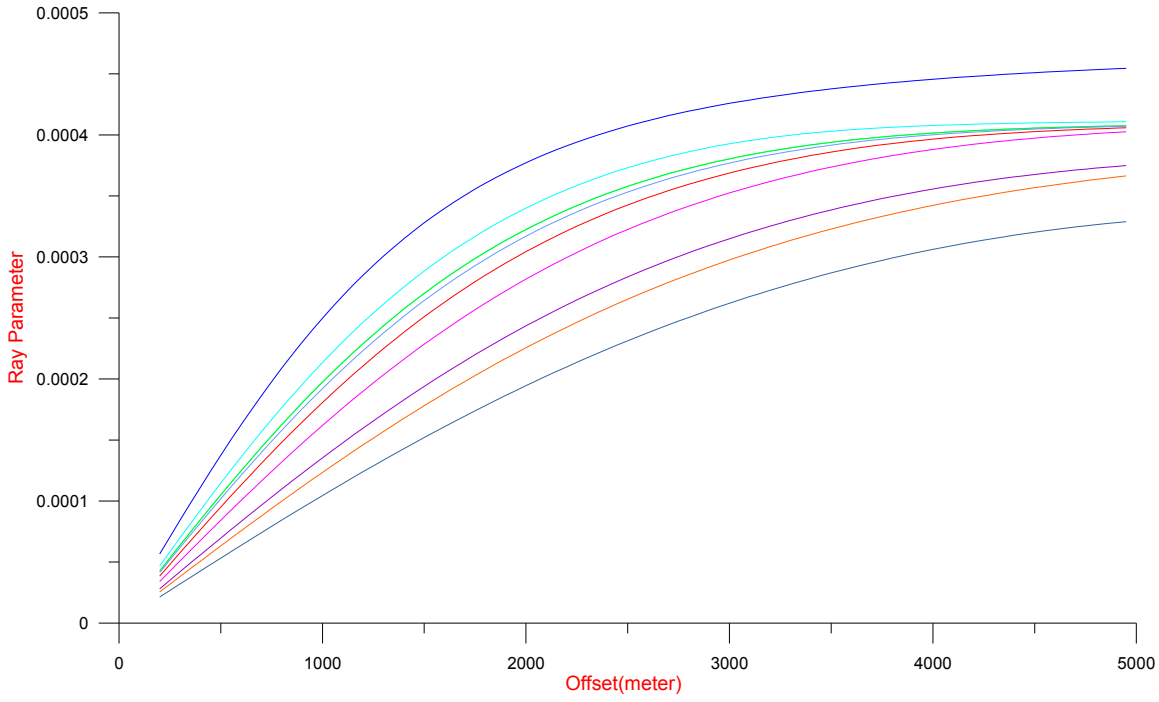


Figure 2.5: Ray parameters estimated for $P \to S_v$ rays from the horizontal layered media (Table 2.1). The top curve corresponds to layer 1, while the lowest one corresponds to layer 9.

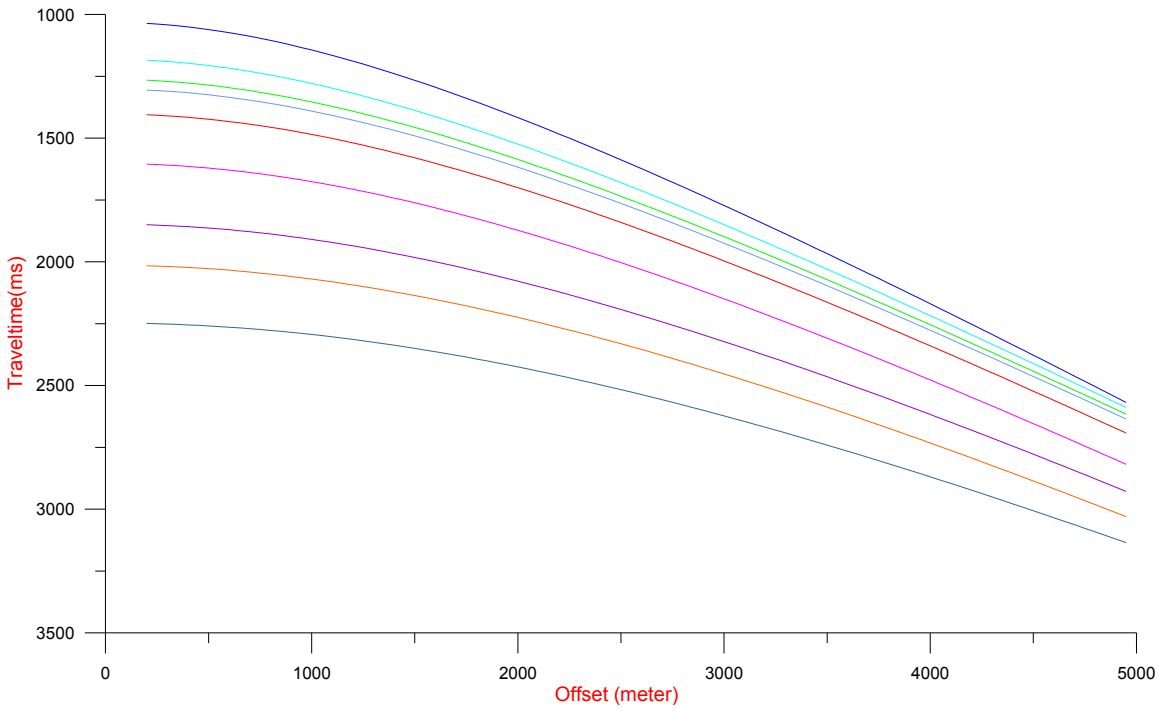


Figure 2.6: Traveltimes calculated for $P \to P$ rays from the horizontal layered media (Table 2.1). The top curve corresponds to layer 1, while the lowest one corresponds to layer 9.

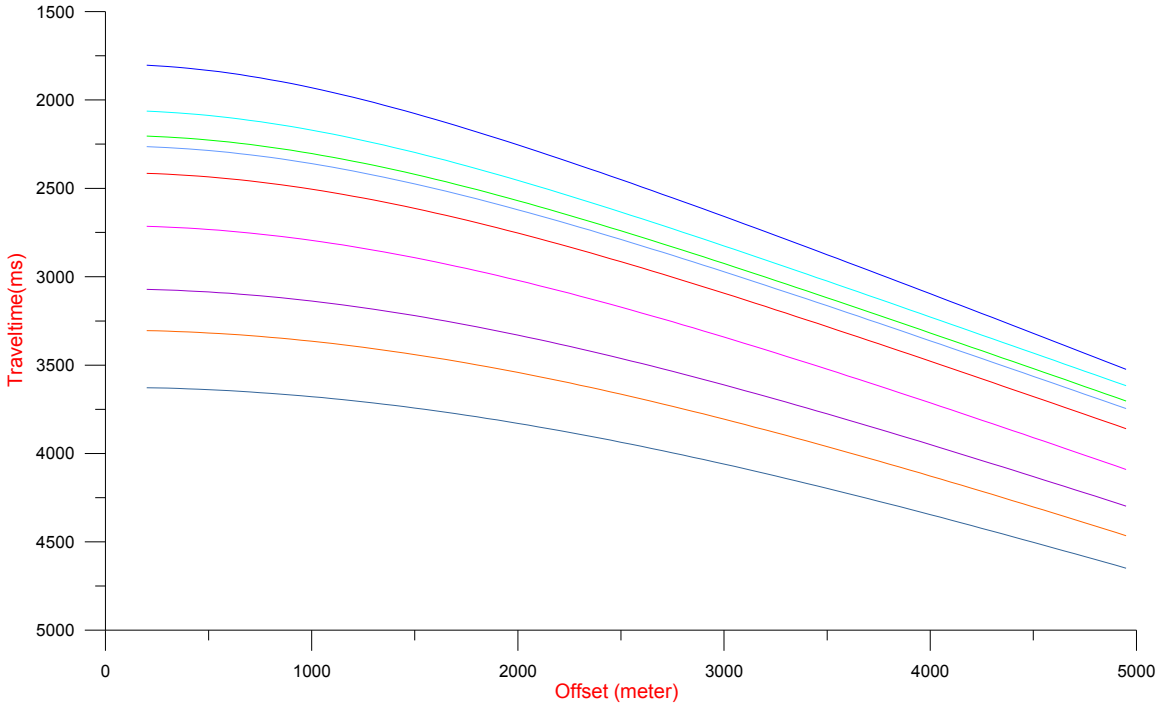


Figure 2.7: Traveltimes calculated for $P \to S_v$ rays from the horizontal layered media (Table 2.1). The top curve corresponds to layer 1, while the lowest one corresponds to layer 9.

2.4 Two point ray tracing in a dipping layered medium

2.4.1 Snell's law in a dipping layered media

In this section we express equations for traveltime and offset in a dipping layered medium. We assume that each layer is transversely isotropic, the axis of symmetry normal to the interface and layer properties do not change laterally. In a dipping layered medium the law of conservation of horizontal slowness is no longer valid and each ray element has its own ray parameter. Along the interface of two dipping layers, the horizontal slowness is constant and we can calculate a ray parameter by projection of vertical and horizontal slowness on the interface (Tsvankin, 2001). Then, using Snell's law along the interface we can compute refraction phase angle (Byun, 1982; Levin, 1990). Snell's law along the interface for different rays can be written as:

Down going waves:

$$\begin{aligned} p^{r}_{ifc} &= \frac{sin(\theta_r - \gamma_{ifc} + \psi_u)}{v_u(\theta_r)} \\ p^{t}_{ifc} &= \frac{sin(\theta_t - \gamma_{ifc} + \psi_l)}{v_l(\theta_t)} \\ p^{r}_{ifc} &= p^{t}_{ifc} \end{aligned} \quad , \tag{2.31}$$

where θ_r is the incidence phase angle, θ_t is the refraction phase angle, γ is the dip angle of the interface, and ψ is dip angle of the symmetry plane. u and l stand for upper and lower layers, v is phase velocity, p is ray parameter, and ifc stand for interface .

Reflected waves:

$$\begin{aligned} p_r &= \frac{sin(\theta_r - \gamma_r + \psi_r)}{v_r(\theta_r)} \\ p_{r'} &= \frac{sin(\theta_{r'} + \gamma_r - \psi_r)}{v_r(\theta_{r'})} \\ p_r &= p_{r'} \end{aligned} \quad , \tag{2.32}$$

where r and r' stand for incidence and reflected waves.

Up going waves:

$$p_{ifc}^{r} = \frac{sin(\theta_r + \gamma_{ifc} - \psi_u)}{v_l(\theta_r)}$$
$$p_{ifc}^{t} = \frac{sin(\theta_t + \gamma_{ifc} - \psi_l)}{v_u(\theta_t)} \quad . \tag{2.33}$$
$$p_{ifc}^{r} = p_{ifc}^{t}$$

Equations (2.31) – (2.33) are the general form of Snell's law and have to be corrected for the dip direction of the layer and the symmetry plane. This could be done by multiplying by +1 for positive dips (assuming the dipping directions are counter clockwise) and -1 for negative dips (assuming the dipping directions are clockwise).

Suppose a ray is propagating downward. For a given ray parameter along an interface (p_{ifc}^{r}), we can calculate the phase velocity in the upper layer (v_u), then the phase angle of incidence can be calculated using equation (2.29). Now, phase velocity in the lower layer (v_l) for the given ray parameter along the interface can be calculated and, therefore, the phase angle of refraction or equally, the phase angle of incidence for the lower layer is found. In each layer the horizontal slowness (p_1) and vertical slowness (p_3) can be calculated simply by Snell's law (equation(2.31)).

$$p_1 = \frac{sin(\theta)}{v(\theta)}$$
$$p_3 = \frac{cos(\theta)}{v(\theta)} \quad , \tag{2.34}$$

which, in fact are related to the ray parameter along an interface via:

$$p_{ifc} = p_1 cos(\gamma_{ifc}) + p_3 sin(\gamma_{ifc}) \quad . \tag{2.35}$$

2.4.2 Traveltime and offset in a dipping layered media

We assume a dipping transversely isotropic medium composed of a stack of layers in a 2-D plane (Figure 2.8). We also assume that ray trajectory has to be in the axis of the symmetry plane, so there is no out of plane reflection. At the interface of two media, in

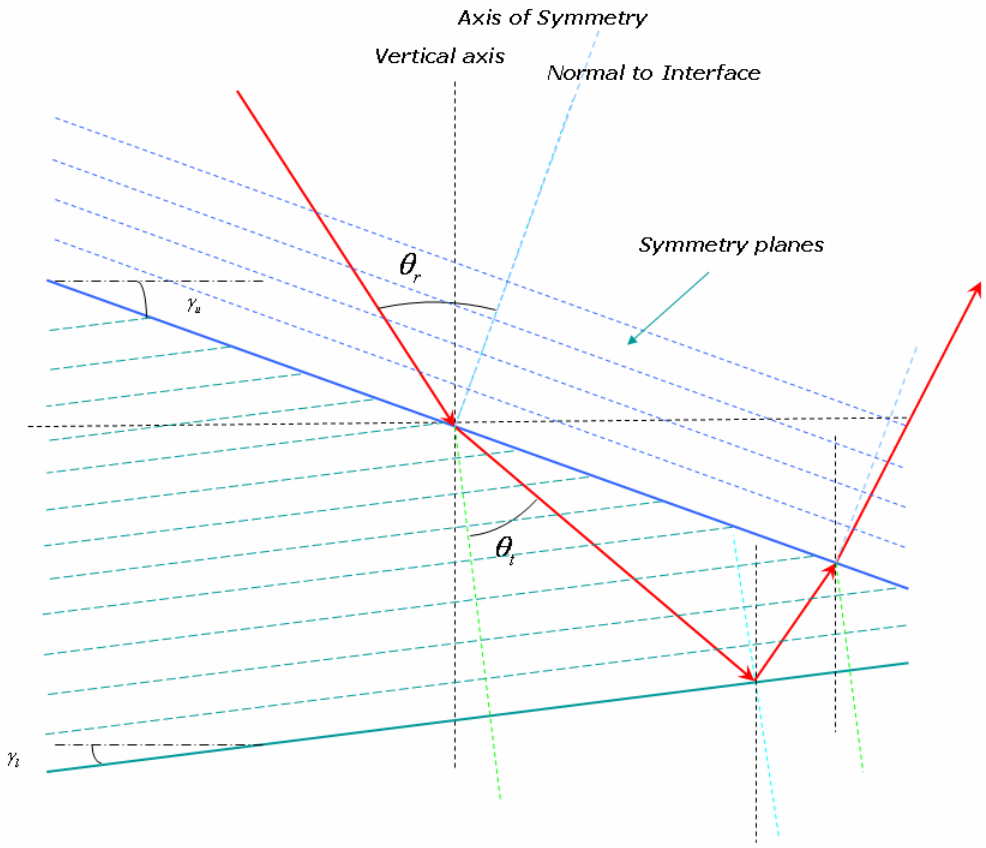


Figure 2.8: Incidence and refraction phase angles in a dipping media

addition to P-wave reflections, shear wave conversions will also be created (Figure 2.9). We consider a down-going ray element (Figure 2.10) between two dipping interfaces travelling left to right. The source- receiver offset (X) is the sum of the projection of all ray elements (down-going and up-going) on the horizontal plane (equation (2.35)). Using some geometrical and trigonometry relationships, the projection of ray element (X_k) can be expressed based on ray angle (α), dip angles (γ_u , γ_l), thickness (Z, measured from the middle of the model) as (see appendix 2.D and 2.E) :

$$X_k = \frac{Z_k - a_k tan(\gamma_k^u) + a_k tan(\gamma_k^l)}{cot(\alpha_k) + tan(\gamma_k^l)} , \tag{2.36}$$

$$X = \sum_{k=0}^{n-1} X_k , \tag{2.37}$$

where

$$cot(\alpha_k) = \frac{\sqrt{1 - p_k^2 v_0^{\,2}\left(1 + H_k\right)}}{p_k v_{0,k}\left(1 + H_k + \frac{1}{2} p_k H'_k\right)}\ , \tag{2.38}$$

$$a_k = \frac{Offset}{2} - X_{c,k-1}\ ,\ a_0 = 0,\ \ X_{c,-1} = 0\ ,\ k = 0,1,2...n-1,\ n = 2N - 2\,.$$

X_c is sum of the projection of all ray elements from left to right up to the current ray element k. N is number of the layers.

Traveltime T_k for each ray element can be expressed by (see appendix 2.F):

$$T_k = \frac{1 + \frac{1}{2} p_k^3 v_{0,k}^2 H'_k}{p_k v_{0,k}^2\left(1 + H_k + \frac{1}{2} p_k H'_k\right)} X_k\ , \tag{2.39}$$

$$T = \sum_{k=0}^{n-1} T_k\ . \tag{2.40}$$

In the case of a converted shear waves H and H' should be replaced by H_s and H'_s.

Equation (2.36) is the general form and the sign of the *tan* terms has to be changed according to the ray propagation direction and dip direction of interfaces. For the clockwise dip directions they have to be multiplied by (+1) and for counter clockwise dip directions they have to be multiplied by (-1). In addition, the *tan* terms need to be multiplied by (+1) in the case of down-going rays and (-1) for the up-going rays.

In an anisotropic dipping layered medium calculation of the ray parameter is more complicated and computationally intensive. We draw a random ray parameter in the interface of the first and second layer and all the ray parameters for each ray element will be computed using Snell's law. A very small variation in ray parameter $\left(1 \times 10^{-6}\right)$ will results in a few millisecond error in traveltime. Therefore, ray parameters have to be estimated very precisely, so that the difference in computed offset and original offset should be less then $\left(1 \times 10^{-8}\right)$ metre. This very high precision will put an extra burden on minimization and require more iterations. Apart from the discontinuity problem associated with the objective function, for near offsets, small ray parameters result in a

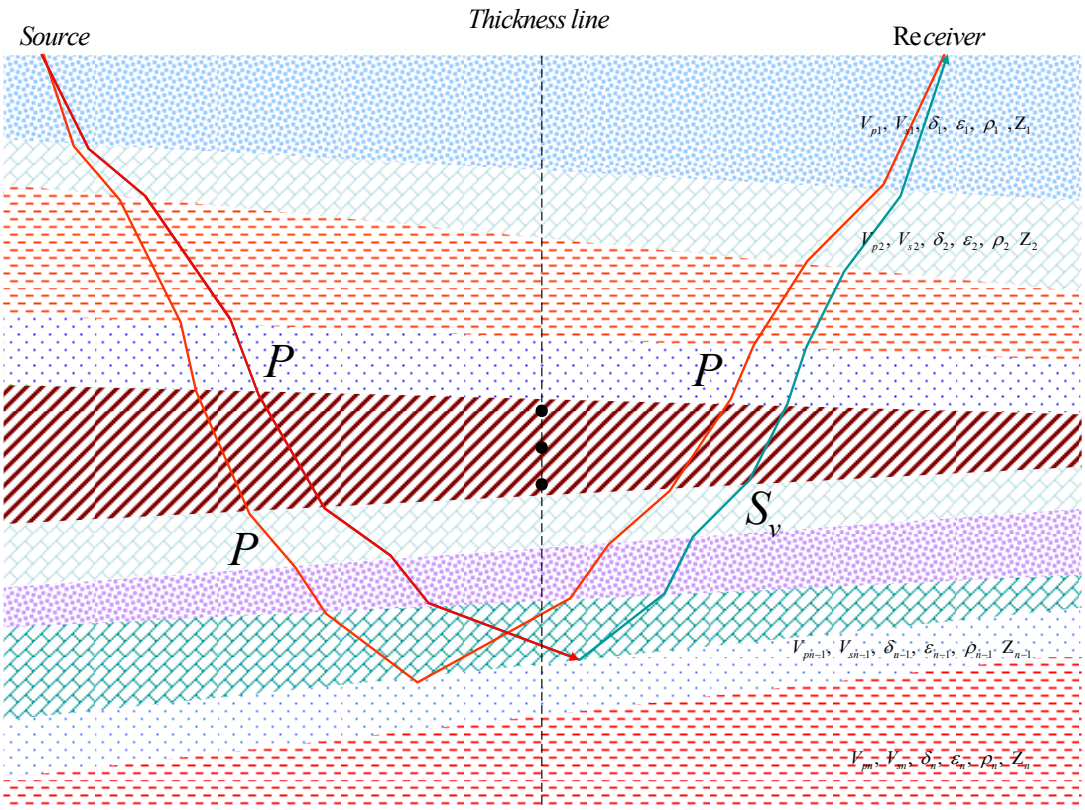


Figure 2.9: Ray paths for $P \rightarrow P$ and $P \rightarrow S_v$ travelling through a dipping layered anisotropic media (Z is the thickness measured from the middle of model).

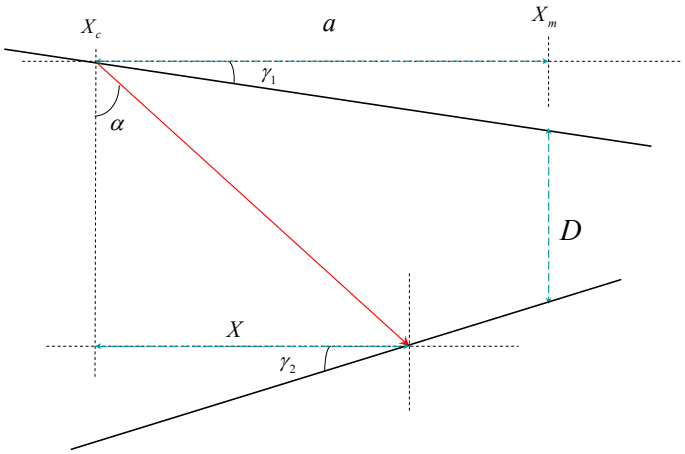


Figure 2.10: A ray element travelling downward through a dipping layer

negative offset, which can't be handled by a logarithmic objective function. Fortunately, this range of ray parameters is quite far from the discontinuity region and minimization will alternate to the non-logarithmic objective function (Figure 2.11).

The objective function f is a monotonically decreasing function for the range of ray parameters from zero to the solution point. Away from the solution it increases sharply and suddenly drops, making a secondary minimum (Figure 2.12). If the ray parameter during the optimization falls in the right hand side of the global minimum, minimization algorithms converge to a local minimum and it results in a jump in traveltime and ray parameter curves in an offset gather. In this case we use a jump detection algorithm based on local curvature (Figure 2.13) to identify the jump and restart the minimization.

To minimize the objective function we need the derivative of (2.36) with respect to the ray parameter. We know that all ray parameters are related to each other through Snell's law in the interface, so it turns out that derivative with respect to every ray parameter should be almost the same. To test this approximation we computed the numerical derivatives of (2.37) with respect to the ray parameter at the first interface for a ray path. We realized that the numerical derivatives, to a good approximation, are equal to the analytic derivative of (2.37) when the analytic differentiation is taken by summing (2.37) over all the ray elements for each ray path (Figure 2.14). In the next chapters whenever we need to calculate the derivative of any parameter with respect to ray parameter, we use this approach.

The traveltimes for a dipping layered medium (Table 2.2) for both compressional and converted shear waves for an offset gather consisting of 96 receivers have been computed. The first receiver is kept at 200 meter from the source and the rest of the receivers are spaced every 50 meters. The maximum offset is 4950 meter. Figures 2.15 and 2.16 illustrate the traveltimes for $P \to P$ and $P \to S_v$ respectively. The non-hyperbolic behaviour of the traveltimes is quite obvious.

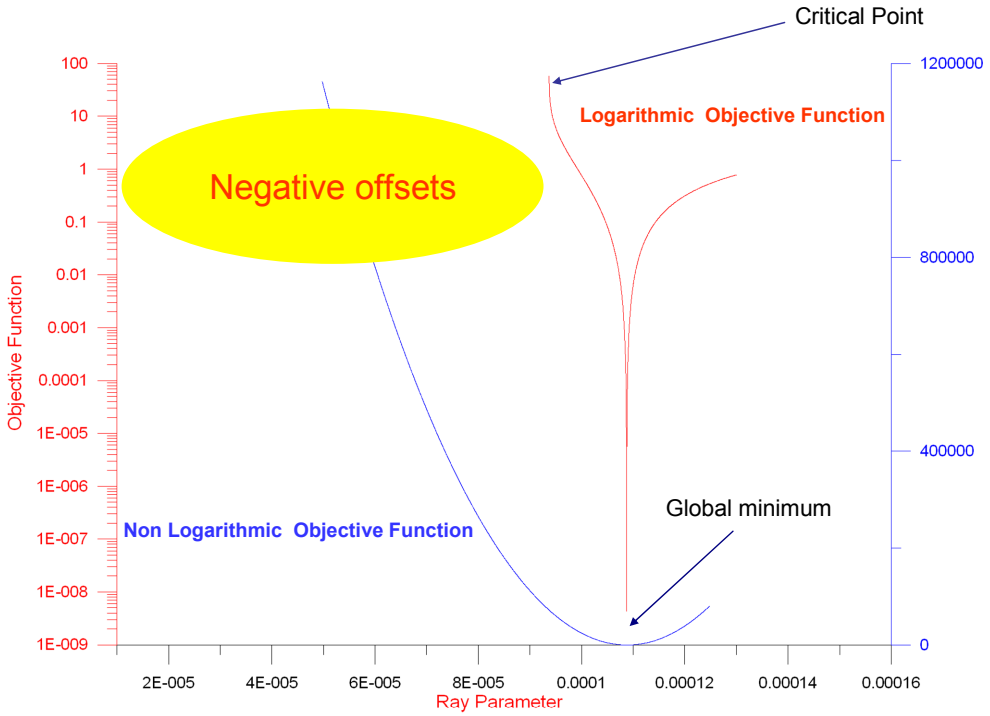


Figure 2.11: Solution space of objective function in a dipping medium for near offsets.

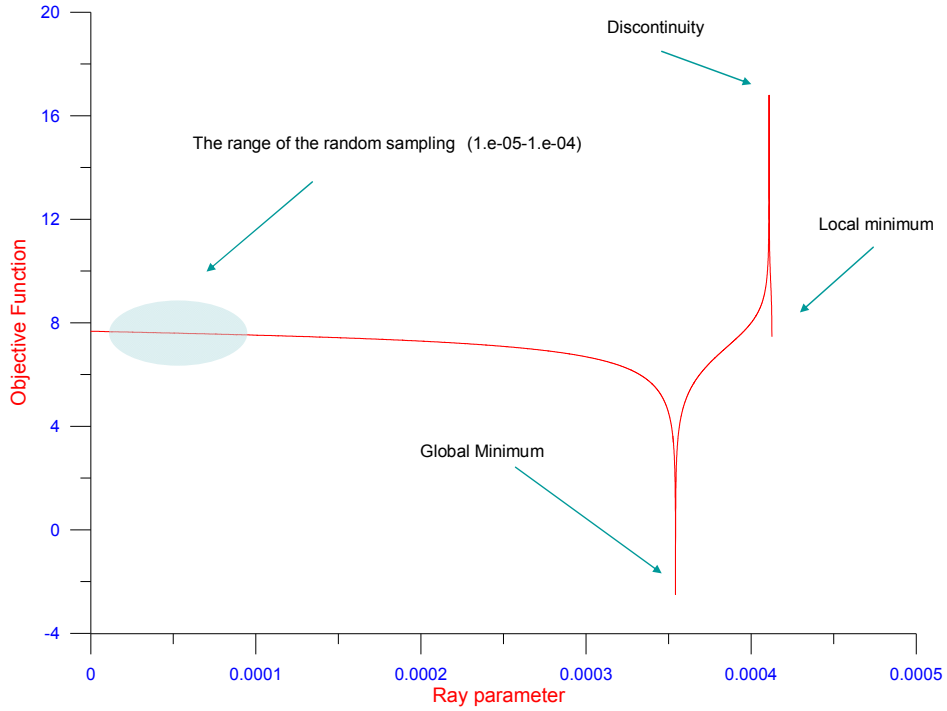


Figure 2.12: Solution space of the logarithmic objective function in a dipping medium for larger offsets.

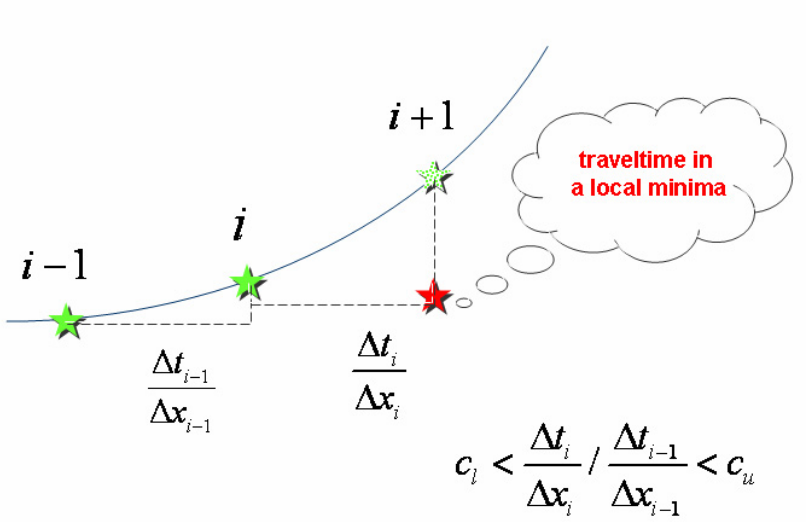


Figure 2.13: Jump detection based on the curvature. The red star is a local minimum solution for ray parameter has resulted in a drop in traveltime

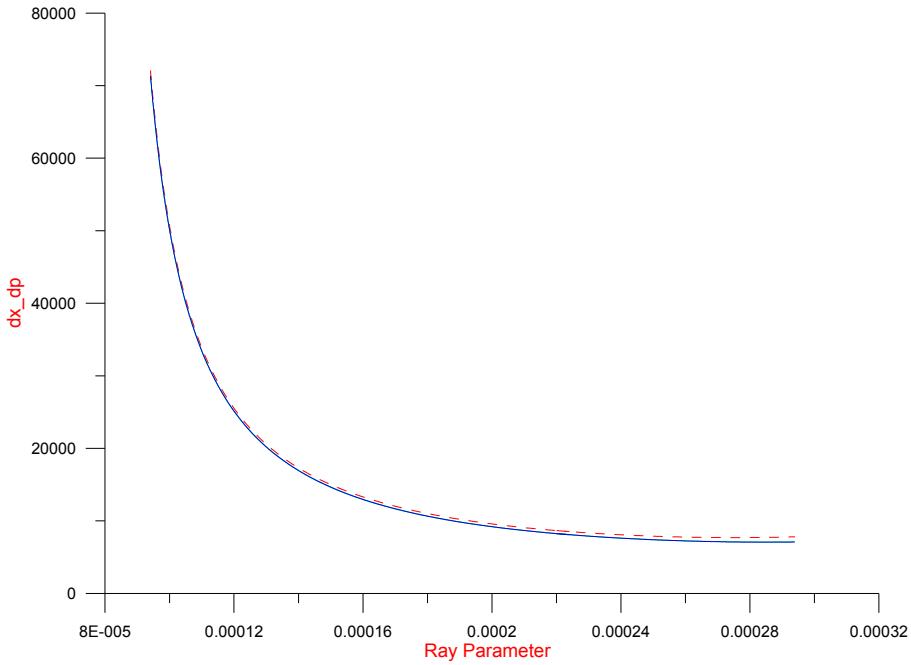


Figure 2.14: Analytic (red-dashed) and numerical (blue-solid) derivatives for the ray path reflecting from the top of layer 10 (Table 2.2). Model width is kept at 3000 metres. Numerical derivatives are calculated from the ray parameter at the interface between the first and second layer. The derivatives with respect to all other ray parameters are the same as the Blue curve (not shown here).

Layer	V_p (m/s)	V_s (m/s)	δ	ε	Layer/ Symmetry plane dip	Z $(meter)$
1	1940	780	0.03	0.1	0	477
2	2140	860	-0.02	0.14	-3	295
3	2220	890	-0.05	0.1	-6	414
4	2000	1000	0.1	0.14	-3	254
5	1990	990	0.05	0.1	-2	398
6	1900	950	0.04	0.12	3	214
7	2200	1150	0.06	0.18	4	231
8	2050	1130	0.1	0.2	3	255
9	2650	1500	0.07	0.1	6	206
10	2750	1530	0.1	0.14	7	374

Table 2.2: Layer properties used to calculate $P \to P$ and $P \to S_v$ traveltimes for the dipping layered.

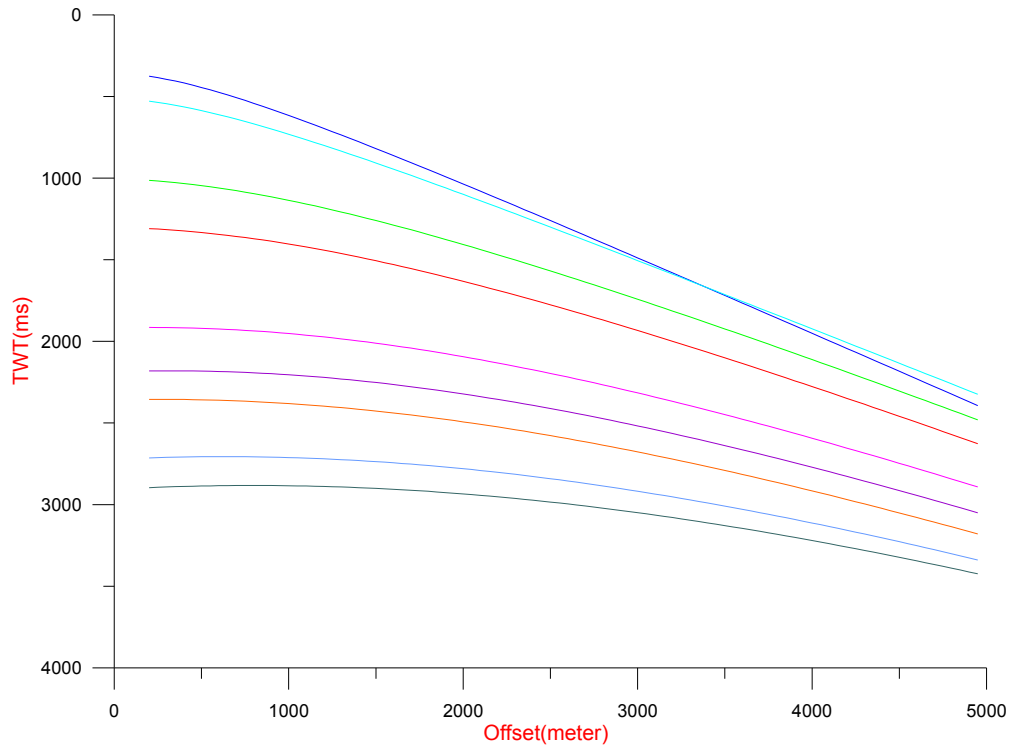


Figure 2.15: Traveltimes calculated for $P \rightarrow P$ rays from the dipping layered medium (Table 2.2). The top curve corresponds to layer 1, while the lowest one corresponds to layer 9

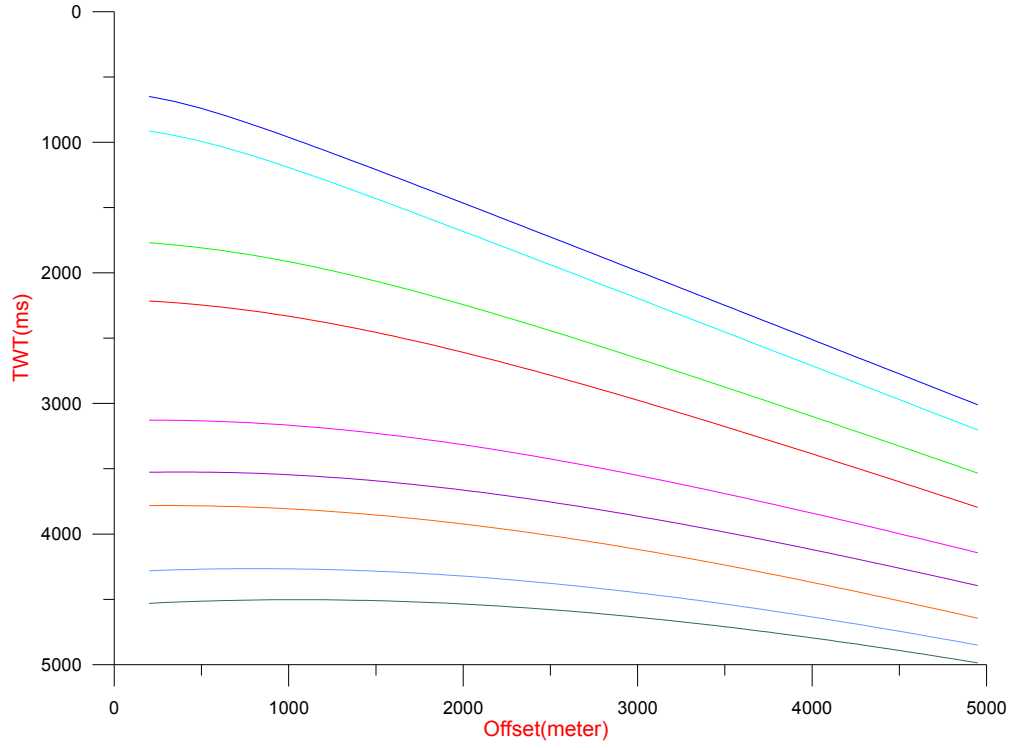


Figure 2.16: Traveltimes calculated for $P \rightarrow S_v$ rays from the dipping layered medium (Table 2.2). The top curve corresponds to layer 1, while the lowest one corresponds to layer 9.

Appendices

Appendix A

Taking the derivative of both sides of equation 2.21 or (A.1) respect to phase angle (θ) we have:

$$cos(\alpha-\theta)=\frac{v}{V} \quad , \tag{A.1}$$

$$\frac{\partial}{\partial\theta}\big(cos(\alpha-\theta)\big)=\frac{\partial}{\partial\theta}\left(\frac{v}{V}\right) , \tag{A.2}$$

$$-sin(\alpha-\theta)\frac{d\alpha}{d\theta}+sin(\alpha-\theta)=\frac{1}{V}\frac{dv}{d\theta}-\frac{v}{V^2}\frac{dV}{d\theta} \quad . \tag{A.3}$$

From equation (1.13) we have:

$$V^2=v^2+\left(\frac{dv}{d\theta}\right)^2 , \tag{A.4}$$

Dividing (A.4) to V^2 and v^2 respectively and using (A.1) results in:

$$sin(\alpha-\theta)=\frac{1}{V}\frac{dv}{d\theta} \quad , \tag{A.5}$$

$$tan(\alpha-\theta)=\frac{1}{v}\frac{dv}{d\theta} \quad , \tag{A.6}$$

using the chain rule we have:

$$\frac{dV}{d\alpha}=\frac{dV}{d\theta}\frac{d\theta}{d\alpha} , \tag{A.7}$$

Then

$$\frac{d\alpha}{d\theta}=\frac{dV}{d\theta}/\frac{dV}{d\alpha} \quad , \tag{A.8}$$

replacing (A.5) , (A.8) and (A.1) in (A.3) results in:

$$tan(\alpha-\theta)=\frac{1}{V}\frac{dV}{d\alpha}=\frac{1}{v}\frac{dv}{d\theta} \quad . \tag{A.9}$$

Let's write:

$$\begin{aligned} &cos(\alpha)=cos(\theta)cos(\alpha-\theta)-sin(\theta)sin(\alpha-\theta) \\ &=cos(\theta)cos(\alpha-\theta)[1-tan(\theta)tan(\alpha-\theta)] \end{aligned} \quad , \tag{A.10}$$

Replacing (A.1) and (2.A.6) in (A.10) results in:

$$cos(\alpha) = cos(\theta)\frac{v}{V}[1 - tan(\theta)\frac{1}{v}\frac{dv}{d\theta}] \quad . \tag{A.11}$$

Let's write:

$$\begin{aligned} sin(\alpha) &= sin(\theta)cos(\alpha-\theta) + cos(\theta)sin(\alpha-\theta) \\ &= sin(\theta)cos(\alpha-\theta)[1 + \frac{1}{tan(\theta)}tan(\alpha-\theta)] \end{aligned} \quad . \tag{A.12}$$

Replacing (A.1) and (A.6) in (A.12) results in:

$$sin(\alpha) = sin(\theta)\frac{v}{V}[1 + \frac{1}{tan(\theta)}\frac{1}{v}\frac{dv}{d\theta}] \quad . \tag{A.13}$$

Now, by dividing (A.12) over (A.10), we obtain (Ursin and Hokstad, 2003):

$$tan(\alpha) = \frac{1}{cos(\theta)}\frac{sin(\theta) + \frac{cos(\theta)}{v}\frac{dv}{d\theta}}{1 - \frac{tan(\theta)}{v}\frac{dv}{d\theta}} \quad . \tag{A.14}$$

Taking the derivative of ray parameter $p = \frac{sin(\theta)}{v}$ with respect to phase angle we obtain:

$$cos(\theta)\frac{d\theta}{dp} = v + p\frac{dv}{dp} \quad , \tag{A.15}$$

$$\frac{dp}{d\theta} = \frac{cos(\theta)}{v + p\frac{dv}{dp}} \quad . \tag{A.16}$$

Using the chain rule we have:

$$\frac{dv}{d\theta} = \frac{dv}{dp}\frac{dp}{d\theta} \quad , \tag{A.17}$$

substitution (A.16) in (A.17) we obtain:

$$\frac{dv}{d\theta} = \frac{cos(\theta)\frac{dv}{dp}}{v + p\frac{dv}{dp}} \quad . \tag{A.18}$$

Now, by substitution of (A.18) in (A.11) we map ray angle and ray velocity to phase angle and phase velocity:

$$Vcos(\alpha) = vcos(\theta)[1 - tan(\theta)\frac{1}{v}\frac{cos(\theta)\frac{dv}{dp}}{v + p\frac{dv}{dp}}]$$

$$= vcos(\theta)[1 - \frac{sin(\theta)}{v}\frac{\frac{dv}{dp}}{v + p\frac{dv}{dp}}]$$

$$= vcos(\theta)[1 - \frac{p\frac{dv}{dp}}{v + p\frac{dv}{dp}}] \quad . \qquad \text{(A.19)}$$

$$= vcos(\theta)[\frac{v}{v + p\frac{dv}{dp}}]$$

Choosing the notation:

$$v' = \frac{dv}{dp} \ ,$$

$$Vcos(\alpha) = vcos(\theta)\frac{1}{1 + \frac{p}{v}v'} \quad . \qquad \text{(A.20)}$$

In the same way, by substitution of (A.18) in (A.14) we obtain (Ursin and Stovas, 2006):

$$tan(\alpha) = \frac{v}{cos(\theta)}[p + \frac{\frac{dv}{dp}}{v^3}] \ , \qquad \text{(A.21)}$$

$$tan(\alpha) = \frac{v}{cos(\theta)}[p + \frac{v'}{v^3}] \quad . \qquad \text{(A.22)}$$

Equations (A.20) and (A.22) will have a main role in traveltime equations from ray domain to phase domain.

Appendix B

By taking the derivative of a general form of equations 2.14 or 2.17, (B.1) respect to ray parameter we obtain:

$$\frac{1}{v^2} = \frac{1}{v_0^2} - p^2 H \quad , \tag{B.1}$$

$$\frac{v'}{v^3} = pH + \frac{1}{2} p^2 H' \ , \tag{B.2}$$

where $H' = \frac{dH}{dp}$ and $v' = \frac{dv}{dp}$.

If we rearrange the equation (B.1) in terms of phase velocity we obtain:

$$v^2 = \frac{v_0^2}{1 - p^2 v_0^2 H} \quad . \tag{B.3}$$

If we express $cos(\theta)$ in terms of ray parameter we obtain:

$$cos(\theta) = \sqrt{\frac{1 - p^2 v_0^2 (1+H)}{1 - p^2 v_0^2 H}} \quad . \tag{B.4}$$

Now, by substitution of (B.3) in (B.4) we obtain:

$$cos(\theta) = \frac{v\sqrt{1 - p^2 v_0^2 (1+H)}}{v_0} \quad . \tag{B.5}$$

By substitution of (B.5) and (B.2) in equation 2.23 (for each ray element), (B.6) we obtain:

$$x = \frac{zv}{cos(\theta)} \left(p + \frac{v'}{v^3} \right) \quad , \tag{B.6}$$

$$x = \frac{z v_0 p \left(1 + H + \frac{1}{2} pH' \right)}{\sqrt{1 - p^2 v_0^2 (1+H)}} \quad . \tag{B.7}$$

In the same way by substitution of (B.5) and (B.2) in equation 2.22 (for each ray element), (B.8) we obtain:

$$t = \frac{z}{v cos(\theta)} \left(1 + p \frac{v'}{v} \right) \ , \tag{B.8}$$

$$t = \frac{z}{v_0} \frac{1 + \frac{1}{2} p^3 v_0^2 H'}{\sqrt{1 - p^2 v_0^2 (1 + H)}} \quad . \tag{B.9}$$

Appendix C

$$\frac{dx}{dp} = \tag{C.1}$$

$$\sum_{k=0}^{2n} \frac{\left(1 + H_k(p) + \frac{1}{2} p \left(\frac{\mathrm{d}}{\mathrm{d}p} H_k(p)\right)\right) z_k v_{0,k}}{\sqrt{1 - \left(1 + H_k(p)\right) p^2 v_{0,k}^2}}$$

$$+ p \left(\sum_{k=0}^{2n} \left(\frac{\left(\frac{3}{2} \frac{\mathrm{d}}{\mathrm{d}p} H_k(p) + \frac{1}{2} p \left(\frac{\mathrm{d}^2}{\mathrm{d}p^2} H_k(p)\right)\right) z_k v_{0,k}}{\sqrt{1 - \left(1 + H_k(p)\right) p^2 v_{0,k}^2}} \right.\right.$$

$$\left.\left. - \frac{1}{2} \frac{\left(1 + H_k(p) + \frac{1}{2} p \left(\frac{\mathrm{d}}{\mathrm{d}p} H_k(p)\right)\right) z_k v_{0,k} \left(- \left(\frac{\mathrm{d}}{\mathrm{d}p} H_k(p)\right) p^2 v_{0,k}^2 - 2 \left(1 + H_k(p)\right) p v_{0,k}^2 \right)}{\left(1 - \left(1 + H_k(p)\right) p^2 v_{0,k}^2\right)^{3/2}} \right) \right)$$

$$\frac{d^2x}{dp^2} = \tag{C.2}$$

$$2 \left(\sum_{k=0}^{2n} \left(\frac{\left(\frac{3}{2} \frac{\mathrm{d}}{\mathrm{d}p} H_k(p) + \frac{1}{2} p \left(\frac{\mathrm{d}^2}{\mathrm{d}p^2} H_k(p)\right)\right) z_k v_{0,k}}{\sqrt{1 - \left(1 + H_k(p)\right) p^2 v_{0,k}^2}} \right.\right.$$

$$\left.\left. - \frac{1}{2} \frac{X_k \left(- \left(\frac{\mathrm{d}}{\mathrm{d}p} H_k(p)\right) p v_{0,k}^2 - 2 \left(1 + H_k(p)\right) v_{0,k}^2 \right)}{1 - \left(1 + H_k(p)\right) p^2 v_{0,k}^2} \right) \right) + \sum_{k=0}^{2n}$$

$$\left(\frac{\left(2 p \left(\frac{\mathrm{d}^2}{\mathrm{d}p^2} H_k(p)\right) + \frac{1}{2} p^2 \left(\frac{\mathrm{d}^3}{\mathrm{d}p^3} H_k(p)\right)\right) z_k v_{0,k}}{\sqrt{1 - \left(1 + H_k(p)\right) p^2 v_{0,k}^2}} \right.$$

$$- \frac{\left(p \left(\frac{3}{2} \frac{\mathrm{d}}{\mathrm{d}p} H_k(p) + \frac{1}{2} p \left(\frac{\mathrm{d}^2}{\mathrm{d}p^2} H_k(p)\right)\right)\right) z_k v_{0,k} \left(- \left(\frac{\mathrm{d}}{\mathrm{d}p} H_k(p)\right) p^2 v_{0,k}^2 - 2 \left(1 + H_k(p)\right) p v_{0,k}^2 \right)}{\left(1 - \left(1 + H_k(p)\right) p^2 v_{0,k}^2\right)^{3/2}}$$

$$+ \frac{3}{4} \frac{p^2 X_k \left(- \left(\frac{\mathrm{d}}{\mathrm{d}p} H_k(p)\right) p v_{0,k}^2 - 2 \left(1 + H_k(p)\right) v_{0,k}^2 \right)^2}{\left(1 - \left(1 + H_k(p)\right) p^2 v_{0,k}^2\right)^2}$$

$$\left. - \frac{1}{2} \frac{X_k \left(- \left(\frac{\mathrm{d}^2}{\mathrm{d}p^2} H_k(p)\right) p^2 v_{0,k}^2 - 4 \left(\frac{\mathrm{d}}{\mathrm{d}p} H_k(p)\right) p v_{0,k}^2 - 2 \left(1 + H_k(p)\right) v_{0,k}^2 \right)}{1 - \left(1 + H_k(p)\right) p^2 v_{0,k}^2} \right)$$

.

Appendix D

From Figure D.1:

$$z_2 = a.tan(\gamma_u) \ , \tag{D.1}$$

$$Tan(\beta) = \frac{z_1 + Z + z_2}{a} = \frac{z_1 + Z + a.tan(\gamma_u)}{a} \ , \tag{D.2}$$
$$z_1 = a.tan(\beta) - a.tan(\gamma_u) - Z \ ,$$

$$cos(\beta) = \frac{a}{l + l_1} \ , \tag{D.3}$$
$$l = \frac{a}{cos(\beta)} - l_1 \ ,$$

$$l_1 = \frac{cos(\gamma_l)}{sin(\beta + \gamma_l)} z_1 = \frac{cos(\gamma_l)}{sin(\beta)cos(\gamma_l) + cos(\beta)sin(\gamma_l)} \tag{D.4}$$
$$= \frac{1}{cos(\beta)} \frac{z_1}{tan(\beta) + tan(\gamma_l)} \ .$$

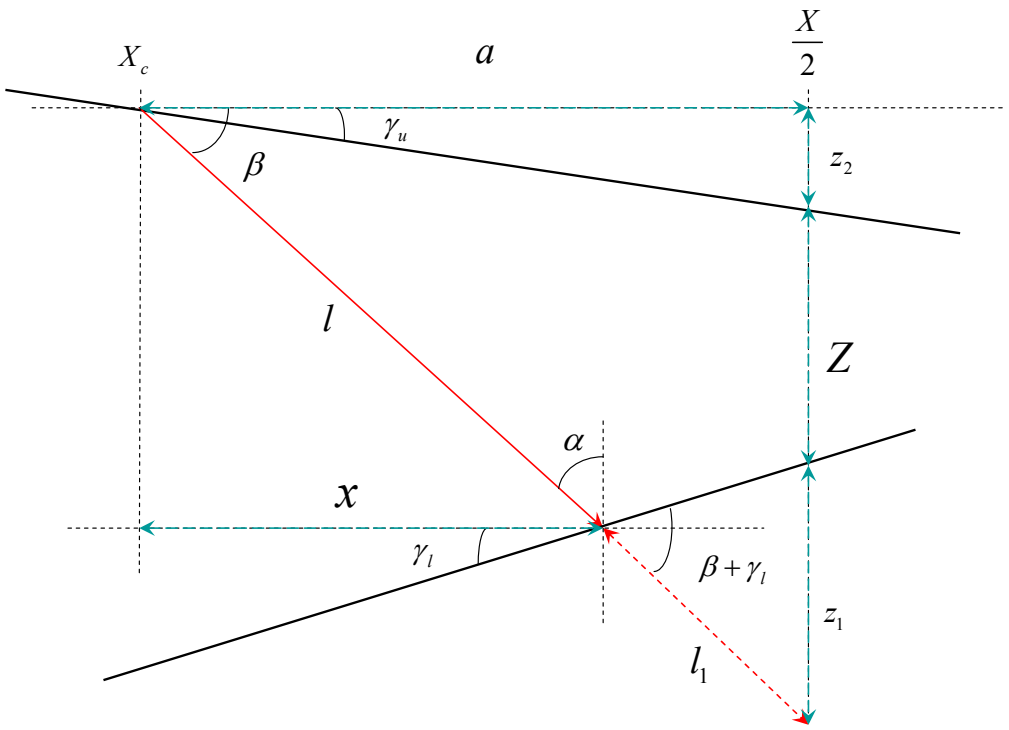


Figure D.1: A ray element travelling downward through a dipping layer

$$l_1 = \frac{1}{cos(\beta)} \frac{a.tan(\beta) - a.tan(\gamma_u) - Z}{tan(\beta) + tan(\gamma_l)} \quad ,$$

$$l = \frac{a}{cos(\beta)} - \frac{1}{cos(\beta)} \frac{a.tan(\beta) - a.tan(\gamma_u) - Z}{tan(\beta) + tan(\gamma_l)} \quad , \tag{D.5}$$

$$x = l.cos(\beta) = a - \frac{a.tan(\beta) - a.tan(\gamma_u) - Z}{tan(\beta) + tan(\gamma_l)}$$

$$= \frac{a.tan(\gamma_l) + a.tan(\gamma_u) + Z}{tan(\beta) + tan(\gamma_l)} \tag{D.6}$$

$$cot(\alpha) = tan(\beta) \ ,$$

so, the projection of ray element l on the horizontal axis is:

$$x = \frac{a.tan(\gamma_l) + a.tan(\gamma_u) + Z}{cot(\alpha) + tan(\gamma_l)} \quad , \tag{D.7}$$

where α is the ray angle.

Appendix E

By substitution of (B.2) in (A.22):

$$tan(\alpha) = \frac{vp}{cos(\theta)}[1 + H + \frac{1}{2} pH'] \quad . \tag{E.1}$$

Now, by substitution of (B.3) and (B.5) in (E.1), we can express (E.1) for $cot(\alpha)$ as:

$$cot(\alpha) = \frac{\sqrt{1 - p^2 v_0^2 (1 + H)}}{pv_0 \left(1 + H + \frac{1}{2} pH'\right)} \quad . \tag{E.2}$$

Appendix F

By rewriting the equation (A.13) we obtain:

$$V\sin(\alpha) = v\sin(\theta)\left(1+\frac{1}{\tan(\theta)}\frac{1}{v}\frac{dv}{d\theta}\right) \quad . \tag{F.1}$$

Now, by substitution of (A.18) in (F.1) and few manipulations we will obtain:

$$\begin{aligned}
V\sin(\alpha) &= v\left(\sin(\theta)+\frac{\cos^2(\theta)}{v}\frac{v'}{v+pv'}\right)\\
&= v^2\left(\frac{\sin(\theta)}{v}+\frac{v'-v'\sin^2(\theta)}{v^2(v+pv')}\right)\\
&= v^2\left(p+\frac{v'}{v^2(v+pv')}-\frac{v'p^2}{v+pv'}\right) \quad .\\
&= \left(pv^2+\frac{v'}{v+pv'}-\frac{p^2v^2v'}{v+pv'}\right)\\
V\sin(\alpha) &= pv^2+\frac{v'\left(1-p^2v^2\right)}{v+pv'}
\end{aligned} \tag{F.2}$$

Now, by substitution of (2.B.2) in (2.F.2) we obtain:

$$\begin{aligned}
V\sin(\alpha) &= pv^2+\frac{v^3\left(pH+\frac{1}{2}p^2H'\right)\left(1-p^2v^2\right)}{v+pv^3\left(pH+\frac{1}{2}p^2H'\right)}\\
&= pv^2+\frac{p^2v^2\left(H+\frac{1}{2}pH'\right)\left(1-p^2v^2\right)}{1+p^2v^2\left(H+\frac{1}{2}pH'\right)} \quad .\\
&= pv^2\left(1+\frac{\left(H+\frac{1}{2}pH'\right)\left(1-p^2v^2\right)}{1+p^2v^2\left(H+\frac{1}{2}pH'\right)}\right)\\
V\sin(\alpha) &= pv^2\left(\frac{1+H+\frac{1}{2}pH'}{1+p^2v^2\left(H+\frac{1}{2}pH'\right)}\right)
\end{aligned} \tag{F.3}$$

By substitution of (B.3) in (F.3) we obtain:

$$V\sin(\alpha) = \frac{pv_0^2}{1-p^2v_0^2H}\left(\frac{1+H+\frac{1}{2}pH'}{1+p^2\frac{v_0^2}{1-p^2v_0^2H}\left(H+\frac{1}{2}pH'\right)}\right) \quad . \tag{F.4}$$

$$V\sin(\alpha) = \frac{pv_0^2\left(1+H+\frac{1}{2}pH'\right)}{1+\frac{1}{2}p^3v_0^2H'}$$

Now, for each ray element the traveltime t can be written as:

$$t = \frac{x}{V\sin(\alpha)} \quad , \tag{F.5}$$

after substitution (F.4) in (F.5) we obtain:

$$t = \frac{1+\frac{1}{2}p^3v_0^2H'}{pv_0^2\left(1+H+\frac{1}{2}pH'\right)}x \quad . \tag{F.6}$$

where x is given by (D.7).

Appendix G

$$\frac{d}{dp_k}X = \sum_{k=0}^{2N-1} \tag{G.1}$$

$$-\frac{\left(Z_k + a_k\,Tan\left(\gamma_{u,k}\right) + a_k\,Tan\left(\gamma_{l,k}\right)\right)\left(\frac{\mathrm{d}}{\mathrm{d}p_k}\,Cota(\alpha)\right)}{\left(Cot(\alpha) + Tan\left(\gamma_l\right)\right)^2} \quad .$$

$$\frac{d}{dp_k}cot\left(\alpha\right)= \tag{G.2}$$

$$\frac{1}{2}\frac{-2\,p_k\,v_{0,k}^2\left(1+H_k\left(p_k\right)\right)-p_k^2\,v_{0,k}^2\left(\frac{\mathrm{d}}{\mathrm{d}p_k}H_k\left(p_k\right)\right)}{\sqrt{1-p_k^2\,v_{0,k}^2\left(1+H_k\left(p_k\right)\right)}\;p_k\,v_{0,k}\left(1+H_k\left(p_k\right)+0.5\,p_k\left(\frac{\mathrm{d}}{\mathrm{d}p_k}H_k\left(p_k\right)\right)\right)}$$

$$-\frac{\sqrt{1-p_k^2\,v_{0,k}^2\left(1+H_k\left(p_k\right)\right)}}{p_k^2\,v_{0,k}\left(1+H_k\left(p_k\right)+0.5\,p_k\left(\frac{\mathrm{d}}{\mathrm{d}p_k}H_k\left(p_k\right)\right)\right)}\quad .$$

$$-\frac{\sqrt{1-p_k^2\,v_{0,k}^2\left(1+H_k\left(p_k\right)\right)}\left(1.5\left(\frac{\mathrm{d}}{\mathrm{d}p_k}H_k\left(p_k\right)\right)+0.5\,p_k\left(\frac{\mathrm{d}^2}{\mathrm{d}{p_k}^2}H_k\left(p_k\right)\right)\right)}{p_k\,v_{0,k}\left(1+H_k\left(p_k\right)+0.5\,p_k\left(\frac{\mathrm{d}}{\mathrm{d}p_k}H_k\left(p_k\right)\right)\right)^2}$$

$$\frac{d^2}{dp_k^2}X=\sum_{k=0}^{2N-1}\frac{-X}{cot\left(\alpha\right)+tan\left(\gamma_l\right)}\frac{d^2cot\left(\alpha\right)}{dp_k^2}-2\frac{X}{\left(cot\left(\alpha\right)+tan\left(\gamma_l\right)\right)^2}\frac{dcot\left(\alpha\right)}{dp_k}\quad . \tag{G.3}$$

$$\frac{d^2}{dp_k^2}cot\left(\alpha\right)= \tag{G.4}$$

$$\frac{1}{2}\frac{1}{1-{p_k}^2v_0^2\left(1+H_k\left(p_k\right)\right)}\left(-2\,v_{0,k}^2\left(1+H_k\left(p_k\right)\right)-2\,p_k\,v_{0,k}^2\left(\frac{\mathrm{d}}{\mathrm{d}p_k}H_k\left(p_k\right)\right)-2\,p_k\,v_{0,k}^2\left(\frac{\mathrm{d}}{\mathrm{d}p_k}H_k\left(p_k\right)\right)Cot(\alpha)-{p_k}^2v_{0,k}^2\left(\frac{\mathrm{d}^2}{\mathrm{d}{p_k}^2}H_k\left(p_k\right)\right)Cot(\alpha)\right)-\frac{1}{2}\frac{1}{\left(1-{p_k}^2v_{0,k}^2\left(1+H_k\left(p_k\right)\right)\right)^2}\left(\left(-2\,p_k\,v_{0,k}^2\left(1+H_k\left(p_k\right)\right)-{p_k}^2v_{0,k}^2\left(\frac{\mathrm{d}}{\mathrm{d}p_k}H_k\left(p_k\right)\right)Cot(\alpha)\right)\left(-2\,p_k\,v_{0,k}^2\left(1+H_k\left(p_k\right)\right)-{p_k}^2v_{0,k}^2\left(\frac{\mathrm{d}}{\mathrm{d}p_k}H_k\left(p_k\right)\right)\right)\right)+\frac{Cot(\alpha)}{{p_k}^2}-\frac{Cot(\alpha)\left(2.0\left(\frac{\mathrm{d}^2}{\mathrm{d}{p_k}^2}H_k\left(p_k\right)\right)+0.5\,p_k\left(\frac{\mathrm{d}^3}{\mathrm{d}{p_k}^3}H_k\left(p_k\right)\right)\right)}{1+H_k\left(p_k\right)+0.5\,p_k\left(\frac{\mathrm{d}}{\mathrm{d}p_k}H_k\left(p_k\right)\right)}+\frac{Cot(\alpha)\left(1.5\left(\frac{\mathrm{d}}{\mathrm{d}p_k}H_k\left(p_k\right)\right)+0.5\,p_k\left(\frac{\mathrm{d}^2}{\mathrm{d}{p_k}^2}H_k\left(p_k\right)\right)\right)^2}{\left(1+H\left(p_k\right)+0.5\,p_k\left(\frac{\mathrm{d}}{\mathrm{d}p_k}H_k\left(p_k\right)\right)\right)^2}$$

.

References

Byun, B. S., 1982, Seismic parameters for media with elliptical velocity dependencies: Geophysics, **47**, 1621-1626.

Ĉervenŷ, V., 2001, Seismic ray theory: Camridge Unversity Press.

Chapman, C. H., 1994, Reflection/Transmission coefficient reciprocities in anisotropic media: Geophysical Journal International, **116**, 498-501.

Chapman, C. H., and J. H. Woodhouse, 1981, Symmetry of the wave equation and excitation of body waves: Geophysical J. R. astr. Soc., **65**, 777-782.

Daley, P. F., and F. Hron, 1977, Reflection and transmission coefficients for transversely isotropic media: Bull. Seismic Soc. Am., **67**, 661-675.

Graebner, M., 1992, Plane-wave reflection and transmission coefficients for a transversley isotropic soild: Geophysics, **57**, 1512-1519.

Hubral, P., and T. Krey, 1980, Interval velocities from seismic reflection time measurements: Society of Exploration Geophysicists.

Ikelle, T. L., and L. Amundsen, 2005, Introduction to petroleum seismology: Society of Exploration Geophysics.

Levin, F. K., 1990, Reflection from a dipping plane - Transversely isotropic solid: Geophysics, **55**, 851-855.

Nocedal, J., and S. J. Wright, 1999, Numerical optimization: Springer-Verlag.

Stovas, A., and B. Ursin, 2003, Reflection and Transmission responses of layered transversely isotropic viscoelastic media: Geophysical Prospecting, **51**, 447-477.

Tsvankin, I., 2001, Seismic signature and analysis of reflection data in anisotropic media: Pergamon.

Ursin, B., 1983, Review of elastic and electromagnetic wave propagation in horizontally layered media: Geophysics, **48**, 1063-1081.

Ursin, B., and K. Hokstad, 2003, Geometrical spreading in a layered transversely isotropic medium with vertical symmetry axis: Geophysics, **68**, 2082-2091.

Ursin, B., and A. Stovas, 2006, Traveltime approximations for a layered transversely isotropic medium: Geophysics, **71**, D23-D33.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged

Chapter 3

Reflection and transmission of plane waves in transversely isotropic media

Introduction

Reflection and transmission (R/T) of seismic waves at an interface of two media is a fundamental topic in theoretical seismology. The solution for an interface between two isotropic media was first given by Zoeppritz (1919). Zoeppritz equations are generally arranged in a matrix form and a closed form solution for 16 R/T coefficients is provided for down and up going compressional and shear waves (Aki and Richards, 1980; Waters, 1981). A parametric solution to Zoeppritz equations is given in (Ĉervenŷ, 2001) which is more suitable for ray tracing based algorithms. Zoeppritz equations are exact plane wave solution and are very complicated to infer the effect of each elastic parameter on the reflection coefficient. Based on a small jump assumption in elastic properties in the boundary of two elastic isotropic media, Aki and Richards (1980) linearized the Zoeppritz equations for $P \to P$ waves. The linearized equation for $P \to P$ waves was rewritten further in terms of the Poisson ratio by (Shuey, 1985). In a critical comment on the use of Poisson ratio in seismology, Thomsen (1990) rearranged Shuey's equation in terms of original compressional and shear waves velocities and shear modulus. Thomsen style of linearization is widely used among exploration seismologists. His method of linearization gives more insight into the effect of seismic elastic parameters on R/T coefficients.

Reflection and transmission of plane waves in anisotropic media has been an active area of research since the 1950 and there is an extensive literature on this topic. Musgrave (1954a; , 1954b) studied the propagation of an elastic plane wave in a general anisotropic and transversely isotropic crystals respectively. Musgrave (1959) formulated the R/T coefficients of an elastic plane wave numerically at the boundary of two elastic media with general anisotropy. While Musgrave was concentrating on plane waves, Buchwald (1959) further studied radiation of elastic waves from a point source in an infinite general anisotropic medium. In 1977 one special issue of the Geophysical Journal

of the Royal Astronomical Society (Volume 49, Issue 1) was devoted to the works that have been done by Crampin and his colleagues on the current state of seismic anisotropy, particularly in earthquake seismology (Bamford and Crampin, 1977). Keith and Crampin (1977) formulated equations for calculation of R/T coefficients numerically for an elastic plane wave at an interface of two general anisotropic media.

Daley and Hron (1977) give analytical expressions for R/T coefficients from a plane wave in a transversely isotropic medium. This paper along with Banik's (1987) first order perturbation theory approach was later used by Thomsen (1993) to linearize R/T coefficients. Thomsen (1993) introduced the idea of an average medium "*two identical isotropic layers separated by a phantom interface as the background, perturbed by a small jump in elastic parameters*" which could approximate the R/T coefficients from the interface of two weakly transverse isotropic media with a small jump in elastic parameters. Banik's (1987) equation are only valid for small angles of incidence while Thomsen's (1993) linearized equations may not be valid for moderate angles of incidence too. Thomson's method of linearization later motivated Rüger (1996) to modify Thomsen's reflectivity equations to a better approximation based on Graebner (1992) equations. While Rüger (1996) only gives analytical expressions for converted shear waves at the symmetry planes of TI or Orthorhombic media, Jílek (2002) gives analytical expressions for azimuthally varying reflectivity at the interface of two weakly TI or Orthorhombic anisotropy media. Varvyčuk and Pšenčik (1998), and Pšenčik and Varvyčuk (1998) give analytic expressions for azimuthally varying R/T coefficients for compressional waves at an interface of two weakly but general anisotropic media based on Thomsen's (1993) approach while, Rüger (1998) extracted analytic equations for azimuthally varying reflection coefficient at an interface of two HTI media and also at the symmetry planes of Orthorhombic media. A complete review of reflection-transmission problem in transversely isotropic medium with vertical axis of symmetry, where the reflection and transmission coefficients are expressed in terms of slowness is given by Stovas and Ursin (2003).

In this chapter I review the Graebner (1992) equations. Rüger's (1996) approximations in transversely isotropic media for compressional and converted shear waves will be discussed. Numerical models will be used to compare the exact and

approximate plane wave reflection coefficients and at the end of this chapter I address the reflectivity calculation for a stack of horizontally layered VTI media by a convolution approach along the ray path. I also compute the corresponding synthetic seismograms of compressional and converted shear wave.

3.1 Polarization vectors and vertical slowness in a TI medium

The exact plane wave analytic expression for reflection and transmission coefficient at an interface of two TI medium, given by Graebner (1992), are particularly suitable for ray parameter based ray tracing problems. We review Graebner's expressions after some modifications in notation. We assume that plane waves are propagating in the symmetry plane of TI medium. The vertical slowness for quasi compressional (QP) and shear waves (QSV) and their polarization vectors can be derived from the solution of Christoffel equation given by (White, 1983):

$$\det\begin{vmatrix} C_{11}p^2 + C_{55}q^2 - \rho & \left(C_{13} + C_{55}\right)pq \\ \left(C_{13} + C_{55}\right)pq & C_{33}q^2 + C_{55}p^2 - \rho \end{vmatrix} = 0 \ , \tag{3.1}$$

$$q_\alpha = \frac{1}{\sqrt{2}}\left(K_1 - \sqrt{K_1^2 - 4K_2K_3}\right)^{\frac{1}{2}} \ , \tag{3.2}$$

$$q_\beta = \frac{1}{\sqrt{2}}\left(K_1 + \sqrt{K_1^2 - 4K_2K_3}\right)^{\frac{1}{2}} \ , \tag{3.3}$$

where

$$K_1 = \frac{\rho}{C_{33}} + \frac{\rho}{C_{55}} - \left\{\frac{C_{11}}{C_{55}} + \frac{C_{55}}{C_{33}} - \frac{\left(C_{13} + C_{55}\right)^2}{C_{33}C_{55}}\right\}p^2 \ ,$$

$$K_2 = \frac{C_{11}}{C_{33}}p^2 - \frac{\rho}{C_{33}} \ ,$$

$$K_3 = p^2 - \frac{\rho}{C_{55}} \ .$$

p and q are horizontal and vertical slowness respectively.

The polarization vectors for QP-wave are given as the eigenvectors corresponding to P-wave vertical slowness:

$$l_\alpha = \left(\frac{C'_{33}q_\alpha^2 + C'_{55}p^2 - 1}{\left(C'_{11}p^2 + C'_{55}q_\alpha^2 - 1\right) + \left(C'_{33}q_\alpha^2 + C'_{55}p^2 - 1\right)} \right)^{\frac{1}{2}}$$
$$m_\alpha = \left(\frac{C'_{11}p^2 + C'_{55}q_\alpha^2 - 1}{\left(C'_{11}p^2 + C'_{55}q_\alpha^2 - 1\right) + \left(C'_{33}q_\alpha^2 + C'_{55}p^2 - 1\right)} \right)^{\frac{1}{2}}, \tag{3.4}$$

where the prime indicates quantities normalized by density.

The polarization vectors for QSV are given as the eigenvectors corresponding to SV-wave vertical slowness:

$$l_\beta = \left(\frac{C'_{11}p^2 + C'_{55}q_\beta^2 - 1}{\left(C'_{11}p^2 + C'_{55}q_\beta^2 - 1\right) + \left(C'_{33}q_\beta^2 + C'_{55}p^2 - 1\right)} \right)^{\frac{1}{2}}$$
$$m_\beta = \left(\frac{C'_{33}q_\beta^2 + C'_{55}p^2 - 1}{\left(C'_{11}p^2 + C'_{55}q_\beta^2 - 1\right) + \left(C'_{33}q_\beta^2 + C'_{55}p^2 - 1\right)} \right)^{\frac{1}{2}}. \tag{3.5}$$

3.2 Boundary conditions at the interface of two solid elastic TI media

The boundary conditions at an interface of two media have an essential effect in the reflection and transmission coefficients. At an interface of two solid elastic media both displacement and stress has to be continuous in three directions when a wavefront either P or SV is incident at the interface. These are vertical and horizontal displacements and normal and shear stresses. Continuity of horizontal displacement implies that the interface has to be welded and no slippage is allowed along the interface. Continuity normal to the interface displacement means that materials can't be separated from each other or penetrate one another (Slawinsky, 2003). The continuity of displacement vectors is usually called the kinematic boundary conditions while dynamic boundary conditions refer to continuity of the stresses along the interface. In the symmetry plane of a TI or monoclinic medium (XZ plane) the conditions on the displacements (u) and stresses (σ) can be written as:

$$[u_x] = [u_z] = 0 \ , \tag{3.6}$$

$$[\sigma_{zx}] = [\sigma_{zz}] = 0 \ , \tag{3.7}$$

where brackets denote the difference in displacement and stresses across the interface.

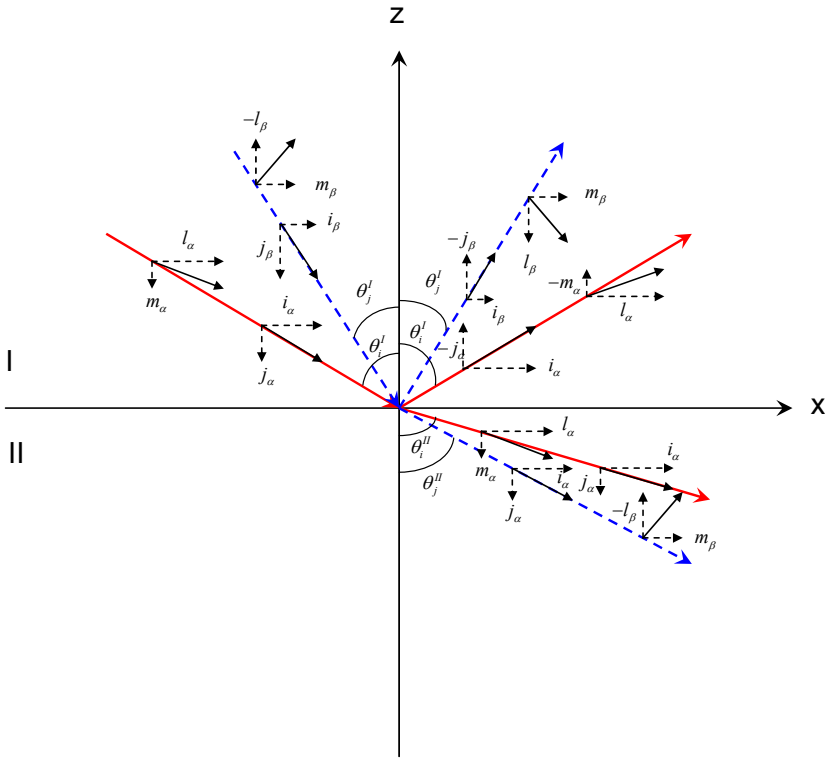


Figure 3.1: Geometry of incidence $P\,\&\,SV$ waves, correspondent polarization vectors (indexed by l, m), and slowness components (indexed by i, j) at the interface in the symmetry plane of two TI medium. Solid lines indicate P-waves and dashed lines SV-waves. (Daley and Hron (1977) after some modifications).

3.3 Exact plane wave reflection and transmission coefficients

If the source is distant and the thickness of the layers is large compared to the wavelength, a wave with a spherical wavefront can be approximated by a plane wave (Daley and Hron, 1977). The particle displacement vector of an elastic plane wave in the symmetry plane is denoted by:

$$u = (l, 0, m) U e^{iw(px+qz-t)} , \tag{3.8}$$

where l and m are polarization vectors for particle motion (Figure 3.1). U is the plane wave amplitude, t is time, and ω angular frequency. The direction of the Z axis is positive downward. The sign of the particle motion has been adopted based on the convention used by Aki and Richards (1980). The positive polarization direction is such that the horizontal component of particle motion has to be in the direction of horizontal slowness. Now, the particle displacement of four incidence waves is denoted by Graebner (1992):

$$u_p^d = (l_\alpha, 0, m_\alpha) U_0 e^{iw(px+q_\alpha z-t)} , \tag{3.9}$$

$$u_p^u = (l_\alpha, 0, -m_\alpha) U_0 e^{iw(px-q_\alpha z-t)} \tag{3.10}$$

$$u_{Sv}^d = (m_\beta, 0, -l_\beta) U_0 e^{iw(px+q_\beta z-t)} , \tag{3.11}$$

$$u_{Sv}^u = (m_\beta, 0, l_\beta) U_0 e^{iw(px-q_\beta z-t)} . \tag{3.12}$$

The particle displacements of the scattered wavefields from a quasi-compressional incident wavefield (3.9) with the unit amplitude at in interface of two elastic medium can be expressed as:

$$u_p^u = (l_{\alpha 1}, 0, -m_{\alpha 1}) r_{pp} e^{iw(px-q_\alpha z-t)} \tag{3.13}$$

$$u_{Sv}^u = (m_{\beta 1}, 0, l_{\beta 1}) r_{ps} e^{iw(px-q_\beta z-t)} , \tag{3.14}$$

$$u_p^d = (l_{\alpha 2}, 0, m_{\alpha 2}) t_{pp} e^{iw(px+q_\alpha z-t)} , \tag{3.15}$$

$$u_{Sv}^d = (m_{\beta 2}, 0, -l_{\beta 2}) t_{ps} e^{iw(px+q_\beta z-t)} . \tag{3.16}$$

where u_p^u, u_{Sv}^u, u_p^d, u_{Sv}^d are displacement of scattered wave mode of reflected QP-wave, reflected QSV-wave, transmitted QP-wave, and transmitted QSV-wave respectively.

The kinematic boundary conditions (3.6) for a quasi compressional incident wave at the interface of two elastic medium and in an arbitrary point $(x=0,\ y=0,\ z=0)$ for a single angular frequency ω and any time t can be constructed by substitution of (3.9) and (3.13)-(3.16) in (3.6).

$$\begin{aligned} -l_{a1} &= r_{pp} l_{a1} - t_{pp} l_{a2} + r_{ps} m_{\beta 1} - t_{ps} m_{\beta 2}\,, \qquad \text{for } x \text{ direction} \\ -m_{\alpha 1} &= -r_{pp} m_{a1} - t_{pp} m_{a2} + r_{ps} l_{\beta 1} + t_{ps} l_{\beta 2}\,, \qquad \text{for } z \text{ direction}\,. \end{aligned} \tag{3.17}$$

In a general anisotropic medium stress-strain relationship can be written using the Hooke's law in a tensor notation as:

$$\sigma_{ij} = \frac{1}{2} C_{ijkl} \left(u_{k,l} + u_{i,k} \right)\,, \tag{3.18}$$

where in $u_{k,l}$ the comma denotes the partial derivative with respect to l component and the Einstein summation over repeated indices is implied. In the symmetry plane of a transversely isotropic medium, the dynamic boundary conditions (3.7) using (3.18) can be written as:

$$\begin{aligned} \sigma_{xz} &= C_{55} \left(\frac{\partial u_x}{\partial z} + \frac{\partial u_z}{\partial x} \right)\,, \\ \sigma_{zz} &= C_{13} \frac{\partial u_x}{\partial x} + C_{33} \frac{\partial u_z}{\partial z}\,. \end{aligned} \tag{3.19}$$

Now, by substitution of (3.17) in (3.19) we have two more equations for a quasi compressional wave incident above the interface at an arbitrary point $(x=0,\ y=0,\ z=0)$ for a single angular frequency (ω). Considering that the time (t) is the same for all incident, reflected and transmitted waves, we have (Rüger, 1996):

Shear Stress

$$\begin{aligned} -C^I_{55}\left(q_{\alpha 1} l_{\alpha 1} + p m_{\alpha 1}\right) &= r_{pp} C^I_{55}\left(-q_{\alpha 1} l_{\alpha 1} - p m_{\alpha 1}\right) + r_{ps} C^I_{55}\left(-q_{\beta 1} m_{\beta 1} + p l_{\beta 1}\right) \\ &\quad + t_{pp} C^{II}_{55}\left(-q_{\alpha 2} l_{\alpha 2} - p m_{\alpha 2}\right) + t_{ps} C^{II}_{55}\left(-q_{\beta 2} m_{\beta 2} + p l_{\beta 2}\right)\,, \end{aligned}$$

Normal Stress

$$\begin{aligned} -C^I_{13} p l_{\alpha 1} - C^I_{33} q_{\alpha 1} m_{\alpha 1} &= r_{pp}\left(C^I_{13} p l_{\alpha 1} + C^I_{33} q_{\alpha 1} m_{\alpha 1}\right) + r_{ps}\left(C^I_{13} p m_{\beta 1} - C^I_{33} q_{\beta 1} l_{\beta 1}\right) \\ &\quad + t_{pp}\left(-C^{II}_{13} p l_{\alpha 2} - C^{II}_{33} q_{\alpha 2} m_{\alpha 2}\right) + t_{ps}\left(-C^{II}_{13} p m_{\beta 2} + C^{II}_{33} q_{\beta 2} l_{\beta 2}\right)\,, \end{aligned} \tag{3.20}$$

where I and II denote the upper or lower media respectively.

The boundary conditions for a shear wave incident from above the interface and also for upcoming waves impinging on the interface from below the interface can be expressed in a similar way.

The four sets of linear equations with four unknown reflection and transmission coefficients can be arranged in a matrix form as (Graebner, 1992):

$$sr = b \ , \tag{3.21}$$

where r is a vector of reflection and transmission coefficients

$$r^T = \left(r_{pp} \ \ r_{ps} \ \ t_{pp} \ \ t_{ps} \right) \ , \tag{3.22}$$

s is a 4×4 matrix:

$$s = \begin{pmatrix} l_{\alpha 1} & m_{\beta 1} & -l_{\alpha 2} & -m_{\beta 2} \\ m_{\alpha 1} & -l_{\beta 1} & m_{\alpha 2} & -l_{\beta 2} \\ a_1 & b_1 & a_2 & b_2 \\ c_1 & d_1 & -c_2 & -d_2 \end{pmatrix} , \tag{3.23}$$

and b is a vector of the form

$$b = \begin{pmatrix} -l_{\alpha 1} \\ m_{\alpha 1} \\ C_{55}^{I} \left(q_{\alpha 1} l_{\alpha 1} + p m_{\alpha 1} \right) \\ -\left(p l_{\alpha 1} C_{13} + q_{\alpha 1} m_{\alpha 1} C_{13} \right) \end{pmatrix} , \tag{3.24}$$

where

$$\begin{aligned} a_1 &= C_{55}^{I} \left(q_{\alpha 1} l_{\alpha 1} + p m_{\alpha 1} \right) \\ b_1 &= C_{55}^{I} \left(q_{\beta 1} m_{\beta 1} - p l_{\beta 1} \right) \\ a_2 &= C_{55}^{II} \left(q_{\alpha 2} l_{\alpha 2} + p m_{\alpha 2} \right) \\ b_2 &= C_{55}^{I} \left(q_{\beta 2} m_{\beta 2} - p l_{\beta 2} \right) \\ c_1 &= p l_{\alpha 1} C_{13}^{I} + q_{\alpha 1} m_{\alpha 1} C_{33}^{I} \\ d_1 &= p m_{\beta 1} C_{13}^{I} - q_{\beta 1} l_{\beta 1} C_{33}^{I} \\ c_2 &= p l_{\alpha 2} C_{13}^{II} + q_{\alpha 2} m_{\alpha 2} C_{33}^{II} \\ d_2 &= p m_{\beta 2} C_{13}^{II} - q_{\beta 2} l_{\beta 2} C_{33}^{II} \end{aligned} \ . \tag{3.25}$$

Equation (3.21) can be extended to include all the 16 reflection and transmission coefficients resulting from 4 incident waves from above and below the interface as:

$$SR = B \ , \tag{3.26}$$

where

$$S = \begin{pmatrix} l_{\alpha 1} & m_{\beta 1} & -l_{\alpha 2} & -m_{\beta 2} \\ c_1 & d_1 & -c_2 & -d_2 \\ m_{\alpha 1} & -l_{\beta 1} & m_{\alpha 2} & -l_{\beta 2} \\ a_1 & b_1 & a_2 & b_2 \end{pmatrix} \ , \tag{3.27}$$

and

$$R = \begin{pmatrix} r_{pp}^D & r_{sp}^D & t_{pp}^U & t_{ps}^U \\ r_{ps}^D & r_{ss}^D & t_{ps}^U & t_{ss}^U \\ t_{pp}^D & t_{sp}^D & r_{pp}^U & r_{sp}^U \\ t_{ps}^D & t_{ss}^D & r_{ps}^U & r_{ss}^U \end{pmatrix} . \tag{3.28}$$

Matrix R is called the scattering matrix too, and

$$B = \begin{pmatrix} -l_{\alpha 1} & -m_{\beta 1} & l_{\alpha 2} & m_{\beta 2} \\ -c_1 & -d_1 & c_2 & d_2 \\ m_{\alpha 1} & -l_{\beta 1} & m_{\alpha 2} & -l_{\beta 2} \\ a_1 & b_1 & a_2 & b_2 \end{pmatrix} . \tag{3.29}$$

Equation (3.26) can be solved numerically. Due to block symmetry in S and B matrices, they can be further simplified to lower dimensional matrices, see Graebner (1992).

3.4 Approximate plane wave reflection and transmission coefficients

The exact plane wave solution to R/T coefficients at the boundary of two elastic TI media is too complex to intuitively infer the physics behind it. It also depends upon all elastic parameters above and below the interface which makes it even more difficult to understand. An R/T phenomenon depends, primarily, on the contrast of elastic parameters rather than each parameter individually. Practically, the contrast in elastic properties in most cases in the exploration seismology is relatively small (10-20-30%) and in the context of weak anisotropy, linearization of exact expressions seems to be reasonable (Thomsen, 1993).

$$\left|\frac{\Delta\alpha_0}{\bar{\alpha}_0}\right| \ll 1, \quad |\varepsilon_i| \ll 1,$$
$$\left|\frac{\Delta\beta_0}{\bar{\beta}_0}\right| \ll 1, \quad |\delta_i| \ll 1, \qquad i = 1,2 \quad , \tag{3.30}$$
$$\left|\frac{\Delta\rho}{\bar{\rho}}\right| \ll 1 \qquad |\gamma_i| \ll 1,$$

where Δ means difference and bar means average. for instance: $\Delta\rho = \rho_2 - \rho_1$ and $\bar{\rho} = (\rho_2 + \rho_1)/2$.

With the above assumptions, Thomsen (1993) proposed two isotropic identical media with a phantom interface in between, where there would be no reflection and conversion. This is also called a background isotropic medium. With the small perturbation in both media, equation (3.26) can be written as:

$$(S_0 + \Delta S)(R_0 + \Delta R) = B_0 + \Delta B \ , \tag{3.31}$$

where 0 indices imply the unperturbed model.

Linearizing (3.31) in terms of Δ and using equation (3.26), equation (3.31) can be written as:

$$S_0 \Delta R + \Delta S R_0 = \Delta B \ , \tag{3.32}$$

where the solution ΔR is:

$$\Delta R = S_0^{-1}(\Delta B - \Delta S R_0) \ . \tag{3.33}$$

In a VTI medium where P and SV waves don't depend on γ, ΔB and ΔS can be computed by linearizing (3.27) and (3.28) using the following partial derivates with respect to all parameters in (3.30) as:

$$\Delta = d_j \frac{\partial}{\partial d_j} \qquad j = 1,2...7 \ . \tag{3.34}$$

Matrices S_0 and R_0 have a simpler structure because only one incident waves is transmitting in a medium for each component, so its inverse is simpler, see Rüger (1996).

To express the reflection and transmission coefficients as a function of incidence phase angle (θ) instead of polarization angles (γ) of (3.4) and (3.5), Rüger (1996)

further extends the Tsvankin (1996) equation for polarization angles in a weak TI medium as:

$$\begin{aligned} m &= cos\left(\gamma\right) = \left(1 - fsin^2\left(\theta\right)\left[\delta + 2\left(\varepsilon - \delta\right)sin^2\left(\theta\right)\right]\right)cos\left(\theta\right) \\ l &= sin\left(\gamma\right) = \left(1 + fcos^2\left(\theta\right)\left[\delta + 2\left(\varepsilon - \delta\right)sin^2\left(\theta\right)\right]\right)sin\left(\theta\right) \end{aligned} , \tag{3.35}$$

where

$$f = \alpha_0^2 / \left(\alpha_0^2 - \beta_0^2\right) , \tag{3.36}$$

and $m \to m_\alpha, m_\beta$ and $l \to l_\alpha, l_\beta$ are the direction cosines (3.4) and (3.5) for the weak TI medium. This implies that phase angle θ in (3.35) corresponds to P-wave and SV-wave incidence angles respectively. These cosine directions have been expressed by Rüger (Rüger, 1998) as:

$$\begin{aligned} l_{\alpha,\beta} &= l_{i,j} sin\left(\theta_{i,j}\right) \\ m_{\alpha,\beta} &= m_{i,j} cos\left(\theta_{i,j}\right) \end{aligned} , \tag{3.37}$$

where θ stands for phase angle and i, j correspond to α, β respectively. Now, the matrices S_0 and R_0, need to be rearranged in terms of velocity and phase angle instead of horizontal and vertical slowness. P-wave and SV-wave velocities should be also linearized in terms of (3.30) using (3.34) (Thomsen, 1986a) . Finally, Using Cramer's rule one can solve equation (3.34) for reflection and transmission coefficients.

Rüger (1996) derived the following equations for the reflection and transmission coefficients for P-waves and SV-waves in a VTI medium. These equations are quite good approximation to the exact plane waves equations for pre-critical angles of incidence (usually up to 40 degrees of phase angle of incidence) in the context of weak anisotropy. To clarify the effect of anisotropy, R/T coefficients are decomposed into isotropic and anisotropic terms (e.g. $r_{pp} = r_{pp}^{iso} + r_{pp}^{aniso}$). The following equations are only R/T coefficients for down-going waves. For up-going waves simply upper and lower media have to be swapped. θ is the phase angle of incidence or reflection (converted shear waves), i stands for P-wave and j for SV-waves. Following, are some of the Rüger's equations for reflection and transmission coefficients at an interface of two VTI media for different modes.

$$r_{pp}^{iso}(\theta_i)=\frac{1}{2}\frac{\Delta Z}{\bar{Z}}+\frac{1}{2}\left[\frac{\Delta\alpha_0}{\bar{\alpha}_0}-\left(\frac{2\beta_0}{\alpha_0}\right)^2\frac{\Delta\mu}{\bar{\mu}}\right]sin^2(\theta_i)+\frac{1}{2}\frac{\Delta\alpha_0}{\bar{\alpha}_0}sin^2(\theta_i)tan^2(\theta_i)\ ,$$
$$r_{pp}^{aniso}(\theta_i)=\frac{1}{2}\Delta\delta sin^2(\theta_i)+\frac{1}{2}\Delta\varepsilon sin^2(\theta_i)tan^2(\theta_i)\ . \tag{3.38}$$

where $Z=\rho\alpha_0$, μ is the shear modulus, $\Delta\delta=(\delta_2-\delta_1)$, and $\Delta\varepsilon=(\varepsilon_2-\varepsilon_1)$.

$$t_{pp}^{iso}(\theta_i)=1-\frac{1}{2}\frac{\Delta Z}{\bar{Z}}+\frac{1}{2}\frac{\Delta\alpha_0}{\bar{\alpha}_0}sin^2(\theta_i)+\frac{1}{2}\frac{\Delta\alpha_0}{\bar{\alpha}_0}sin^2(\theta_i)tan^2(\theta_i)\ ,$$
$$r_{pp}^{aniso}(\theta_i)=\frac{1}{2}\Delta\delta sin^2(\theta_i)+\frac{1}{2}\Delta\varepsilon sin^2(\theta_i)tan^2(\theta_i)-(\Delta\delta-\Delta\varepsilon)sin^4(\theta_i)\ . \tag{3.39}$$

$$r_{ps}^{iso}=-\frac{1}{2}\frac{\Delta\rho}{\bar{\rho}}\frac{sin(\theta_i)}{cos(\theta_j)}-\frac{\bar{\beta}_0}{\bar{\alpha}_0}\left(\frac{\Delta\rho}{\bar{\rho}}+2\frac{\Delta\beta_0}{\bar{\beta}_0}\right)sin(\theta_i)cos(\theta_i)$$
$$+\left(\frac{\bar{\beta}_0}{\bar{\alpha}_0}\right)^2\left(2\frac{\Delta\beta_0}{\bar{\beta}_0}+\frac{\Delta\rho}{\bar{\rho}}\right)\frac{sin^3(\theta_i)}{cos(\theta_j)}\ ,$$
$$r_{ps}^{aniso}=\left[\left(\frac{\bar{\alpha}_0^2}{2(\bar{\alpha}_0^2-\bar{\beta}_0^2)cos(\theta_j)}-\frac{\bar{\alpha}_0\bar{\beta}_0cos(\theta_i)}{2(\bar{\alpha}_0^2-\bar{\beta}_0^2)}\right)\Delta\delta\right]sin(\theta_i) \tag{3.40}$$
$$+\left[\frac{\bar{\alpha}_0\bar{\beta}_0cos(\theta_i)}{(\bar{\alpha}_0^2-\bar{\beta}_0^2)}(\Delta\delta-\Delta\varepsilon)\right]sin^3(\theta_i)-\left[\frac{\bar{\alpha}_0^2}{(\bar{\alpha}_0^2-\bar{\beta}_0^2)cos(\theta_j)}(\Delta\delta-\Delta\varepsilon)\right]sin^3(\theta_i)$$
$$-\left[\frac{\bar{\beta}_0^2}{2(\bar{\alpha}_0^2-\bar{\beta}_0^2)cos(\theta_j)}\Delta\delta\right]sin^3(\theta_i)+\left[\frac{\bar{\beta}_0^2}{(\bar{\alpha}_0^2-\bar{\beta}_0^2)cos(\theta_j)}(\Delta\delta-\Delta\varepsilon)\right]sin^5(\theta_i)\ .$$

$$t_{ps}^{iso}=\frac{1}{2}\frac{\Delta\rho}{\bar{\rho}}\frac{sin(\theta_i)}{cos(\theta_j)}-\frac{\bar{\beta}_0}{\bar{\alpha}_0}\left(\frac{\Delta\rho}{\bar{\rho}}+2\frac{\Delta\beta_0}{\bar{\beta}_0}\right)sin(\theta_i)cos(\theta_i)$$
$$-\left(\frac{\bar{\beta}_0}{\bar{\alpha}_0}\right)^2\left(2\frac{\Delta\beta_0}{\bar{\beta}_0}+\frac{\Delta\rho}{\bar{\rho}}\right)\frac{sin^3(\theta_i)}{cos(\theta_j)}\ ,$$
$$r_{ps}^{aniso}=\left[\left(\frac{\bar{\alpha}_0^2}{2\left(\bar{\alpha}_0^2-\bar{\beta}_0^2\right)cos(\theta_j)}+\frac{\bar{\alpha}_0\bar{\beta}_0cos(\theta_i)}{2\left(\bar{\alpha}_0^2-\bar{\beta}_0^2\right)}\right)\Delta\delta\right]sin(\theta_i) \tag{3.41}$$
$$+\left[\frac{\bar{\alpha}_0\bar{\beta}_0cos(\theta_i)}{\left(\bar{\alpha}_0^2-\bar{\beta}_0^2\right)}(\Delta\delta-\Delta\varepsilon)\right]sin^3(\theta_i)+\left[\frac{\bar{\alpha}_0^2}{\left(\bar{\alpha}_0^2-\bar{\beta}_0^2\right)cos(\theta_j)}(\Delta\delta-\Delta\varepsilon)\right]sin^3(\theta_i)$$
$$+\left[\frac{\bar{\beta}_0^2}{2\left(\bar{\alpha}_0^2-\bar{\beta}_0^2\right)cos(\theta_j)}\Delta\delta\right]sin^3(\theta_i)-\left[\frac{\bar{\beta}_0^2}{\left(\bar{\alpha}_0^2-\bar{\beta}_0^2\right)cos(\theta_j)}(\Delta\delta-\Delta\varepsilon)\right]sin^5(\theta_i)\ .$$

$$t_{ss}^{iso}(\theta_i)=1-\frac{1}{2}\left(\frac{\Delta\rho}{\bar{\rho}}+\frac{\Delta\beta_0}{\bar{\beta}_0}\right)+\frac{1}{2}\frac{\Delta\beta_0}{\bar{\beta}_0}sin^2(\theta_j)+\frac{1}{2}\frac{\Delta\beta_0}{\bar{\beta}_0}sin^2(\theta_j)tan^2(\theta_j)\ ,$$
$$t_{ss}^{aniso}(\theta_i)=-\frac{1}{2}\left(\frac{\bar{\alpha}_0}{\bar{\beta}_0}\right)^2(\Delta\delta-\Delta\varepsilon)sin^2(\theta_j)+\left(\frac{\bar{\alpha}_0}{\bar{\beta}_0}\right)^2(\Delta\delta-\Delta\varepsilon)sin^4(\theta_j)\ . \tag{3.42}$$

$$t_{sp}^{iso}=-\frac{1}{2}\frac{\Delta\rho}{\bar{\rho}}\frac{sin(\theta_j)}{cos(\theta_i)}+\frac{\bar{\beta}_0}{\bar{\alpha}_0}\left(\frac{\Delta\rho}{\bar{\rho}}+2\frac{\Delta\beta_0}{\bar{\beta}_0}\right)sin(\theta_j)cos(\theta_j)$$
$$+\left(2\frac{\Delta\beta_0}{\bar{\beta}_0}+\frac{\Delta\rho}{\bar{\rho}}\right)\frac{sin^3(\theta_j)}{cos(\theta_i)}\ ,$$
$$t_{sp}^{aniso}=\left[\left(\frac{\bar{\alpha}_0\bar{\beta}_0cos(\theta_j)}{2\left(\bar{\alpha}_0^2-\bar{\beta}_0^2\right)}+\frac{\bar{\alpha}_0^2cos^2(\theta_j)}{2\left(\bar{\alpha}_0^2-\bar{\beta}_0^2\right)cos(\theta_i)}\right)\Delta\delta\right]sin(\theta_j) \tag{3.43}$$
$$-\left[\frac{\bar{\alpha}_0^3cos(\theta_j)}{\bar{\beta}_0\left(\bar{\alpha}_0^2-\bar{\beta}_0^2\right)}(\Delta\delta-\Delta\varepsilon)\right]sin^3(\theta_j)-\left[\frac{\bar{\alpha}_0^4cos^2(\theta_j)}{\bar{\beta}_0^2\left(\bar{\alpha}_0^2-\bar{\beta}_0^2\right)cos(\theta_i)}(\Delta\delta-\Delta\varepsilon)\right]sin^3(\theta_j)\ .$$

3.5 Numerical examples of exact and approximate R/T coefficients

Computation of exact plane wave R/T coefficients is intensive and it is not easy to derive analytic expression for their partial derivatives. On the other hand, computation of the Rüger's equations is less intensive compared to the exact one. To compare the computational cost of both approximate and the exact plane wave, we have computed the reflection and transmission coefficients for all 16 modes of exact plane wave and equivalent Rüger's equations. It turns out that CPU time of the exact one is about 50 times more than that the approximate one. It is also easy to derive analytic partial derivatives of the Rüger's equations. However Rüger's equations are subject to errors especially in higher angles of incidence and care has to be taken on their limitation.

To compare the exact plane wave reflection and transmission coefficients in a VTI medium (Graebner, 1992) with the corresponding linear approximations (Rüger, 1996) of $P \to P$ and $P \to S_v$ waves, we have used the model in table 2.1 and computed the reflection coefficient from the base of a reflector and effective reflection coefficient (including transmission effect for all layers travelled by the waves) for pre-critical angles of incident of both $P \to P$ and $P \to S_v$ waves. The nearest angle corresponds to the offset at 200 metre and the far angle corresponds to 2550 meter.

Figures 3.2 and 3.3 illustrate the reflection coefficient for $P \to P$ and $P \to S_v$ waves from the interface of the first and second layer. The solid line shows the exact plane wave reflection coefficient, while the dash line shows the approximate plane wave reflection coefficient. For phase angles of incidence up to 25 degrees the approximation is very close to the exact solution, however for larger angles the approximation tends to diverge from the exact one. The differences in elastic properties are shown below the figure. For converted shear waves the conversion point is shifted from the centre of the source-receiver toward the receiver, hence the P-wave angle of incidence is larger than in the $P \to P$, however, it is less for the converted shear wave. The angle of incidence which is plotted for $P \to S_v$ reflection coefficient is in fact the P-wave angle of incidence.

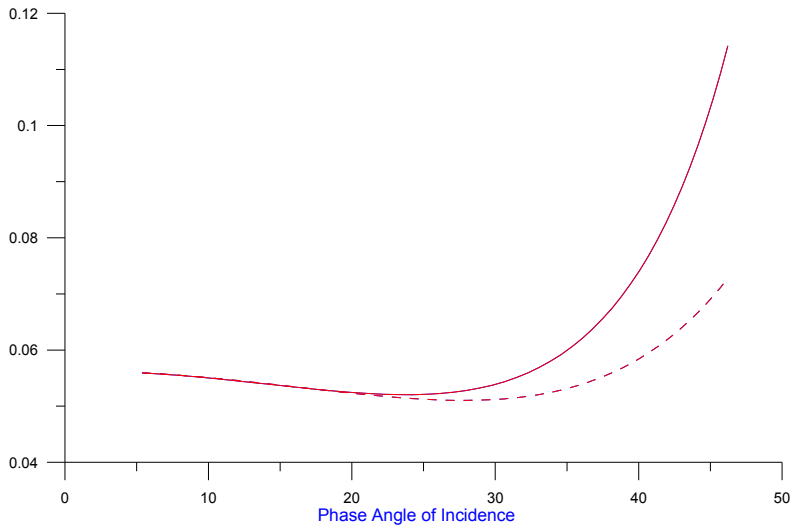


Figure 3.2: Reflection coefficients for $P \to P$ waves from the interface between layer 1 and layer 2. The solid curve shows the exact plane wave and the dashed curve shows plane wave linear approximation. $\Delta\alpha_0 = 9.80392\times10^{-2}$, $\Delta\beta_0 = 9.7561\times10^{-2}$, $\Delta\rho = 1.46699\times10^{-2}$, $\Delta\delta = -5\times10^{-2}$, $\Delta\varepsilon = 4\times10^{-2}$

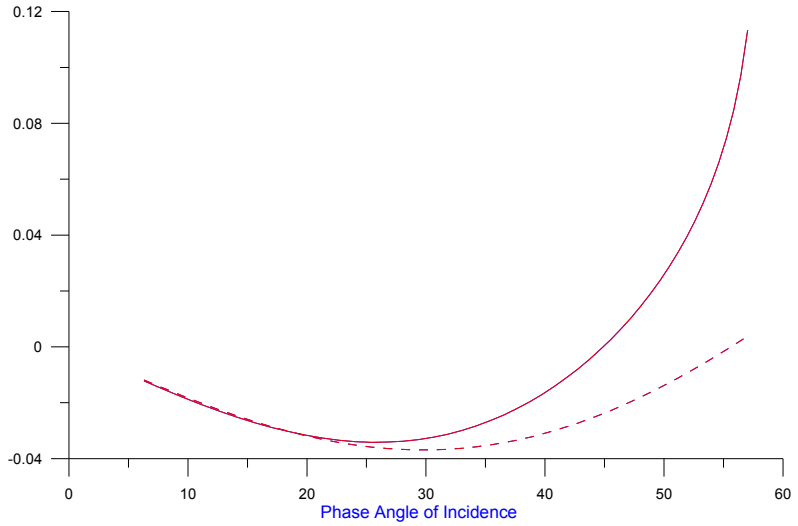


Figure 3.3: Reflection coefficients for $P \to S_v$ waves from the interface between layer 1 and layer 2. The solid curve shows the exact plane wave and the dashed curve shows the plane wave linear approximation. See Figure 3.2 for the contrast in elastic properties.

Figure 3.4 and Figure 3.5 illustrate the exact and approximate reflection coefficients for $P \to P$ and $P \to S_v$ plane waves from the interface between layer 5 and layer 6 respectively. The transmission effect has been included for the entire layers down to the interface for down-going and up-going waves and is shown in blue colour. The approximate solution matches the exact solution up to 30 degree, and after that, it starts to diverge from the exact solution. The approximation in $P \to S_v$ case is even better and it extends to 35 degrees. The transmission effect is quite obvious and it reaches a maximum 5 percent difference compared to the only reflection coefficient (red curves).

Figure 3.6 and Figure 3.7 illustrate the exact and approximate reflection coefficients for $P \to P$ and $P \to S_v$ plane waves from the interface between layer 9 and layer 10. The transmission effect has been included for the entire layers down to the interface for down-going and up-going waves and is shown in blue colour. The approximation is fairly good over the entire range of the incidence angles. The transmission effect is about 10 percent and neglecting it could results in error in effective reflection coefficient. Transmission effect has caused an increase in the absolute magnitude of the reflection coefficient which is not a very uncommon phenomenon. This will happen when the transmission coefficient is bigger than unity.

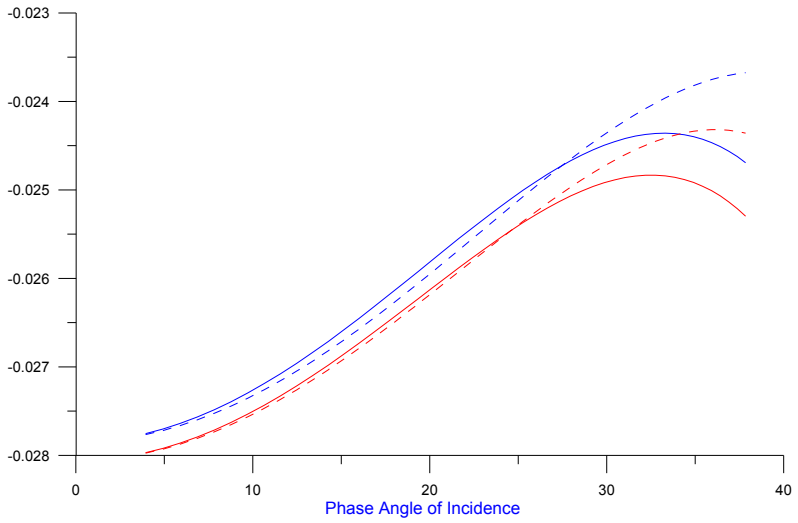


Figure 3.4: Reflection coefficients for $P \to P$ waves from the interface between layer 5 and layer 6. The solid curves show the exact plane wave and the dashed curves show the plane wave linear approximation. Red curves show reflection coefficient without transmission effect while the blue curves correspond to including transmission effect, $\Delta\alpha_0 = \text{-}4.62725\times10^{-2}$, $\Delta\beta_0 = \text{-}4.12371\times10^{-2}$, $\Delta\rho = \text{-}0.985222\times10^{-2}$, $\Delta\delta = \text{-}1\times10^{-2}$, $\Delta\varepsilon = 2\times10^{-2}$.

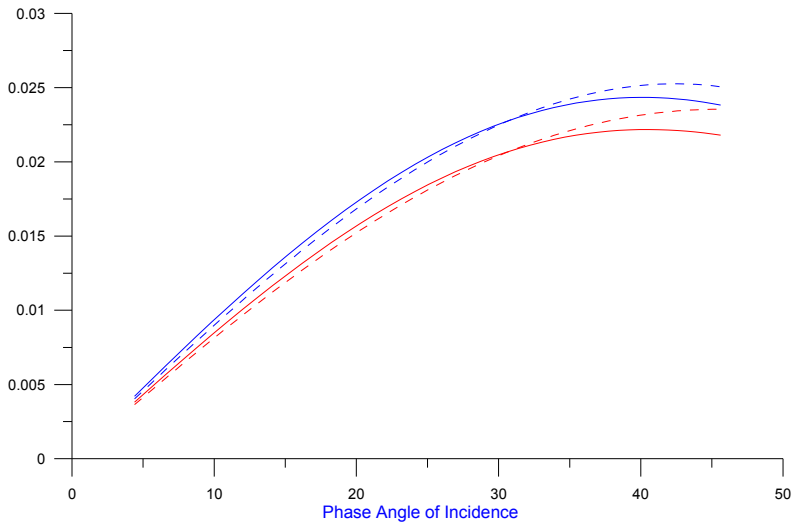


Figure 3.5: Reflection coefficients for $P \to S_v$ waves from the interface between layer 5 and layer 6. See Figure 3.4 for the contrast in elastic properties and identifying the curves.

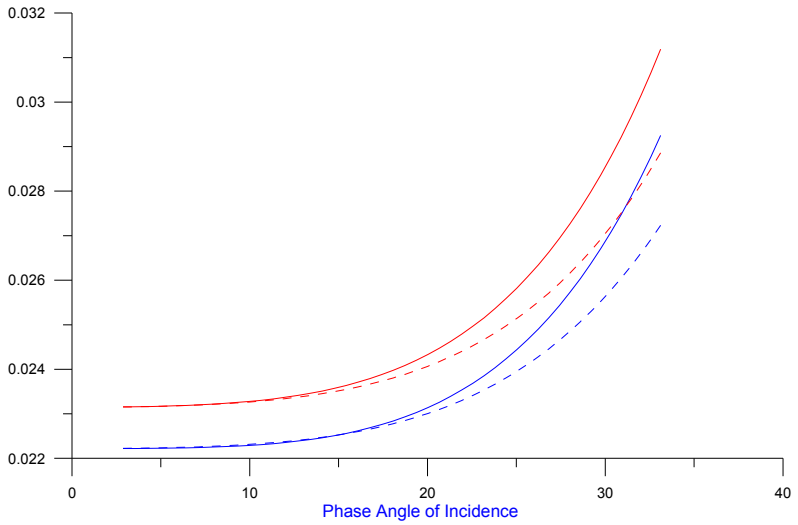


Figure 3.6: Reflection coefficients for $P \to P$ waves from the interface between layer 9 and layer 10. The solid curves show the exact plane wave and the dashed curves show the plane wave linear approximation. The red curves show reflection coefficient without transmission effect while for blue curves, transmission is included. $\Delta\alpha_0 = 3.7037\times10^{-2}$, $\Delta\beta_0 = 1.9802\times10^{-2}$, $\Delta\rho = 0.925926\times10^{-2}$, $\Delta\delta = 3\times10^{-2}$, $\Delta\varepsilon = 4\times10^{-2}$.

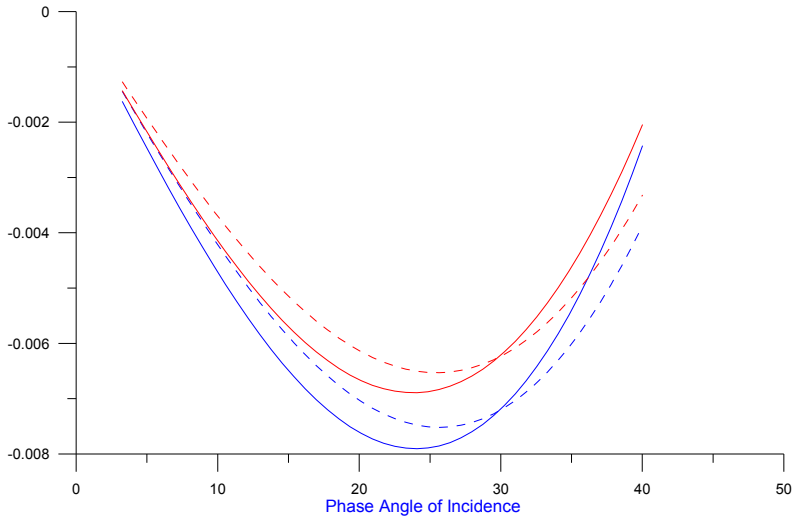


Figure 3.7: Reflection coefficients for $P \to S_v$ waves from the interface between layer 9 and layer 10. See Figure 3.6 for the contrast in elastic properties and indentifying the curves.

3.6 Computation of synthetic seismograms in a layered media

A synthetic seismogram is a time series of reflectivity from a layered earth model which has been convolved with a wavelet. Seismic source signature (wavelet) estimation is a key part in seismic waveform inversion algorithms. The very straightforward statistical deconvolution methods assuming that the signals are stationary and, with an assumption about the phase of the wavelet, only use the amplitude information to estimate the wavelet. On the other hand, the more sophisticated deterministic methods utilize seismic amplitudes as well as sonic log and traveltime information at the location of a well by correlating and calibration of seismic data with well logs, see White and Simm (2003) for more detail. Convolution of a discrete reflectivity series with a wavelet can be written in the time domain as:

$$U(t) = (r * w)(t) = \sum_{i=0}^{n-1} r_i w(t - \tau_i) \ , \tag{3.44}$$

where t is the discrete time sampling and τ_i is traveltime from the source to the receiver for a particular ray path i, n is the number of ray paths, τ_i has to be an integer number of sampling interval Δt, and U is the amplitude of the seismic trace. Geometrical spreading term is ignored as in chapter five we assume seismic data has been corrected for this effect.

Synthetic seismic computation is a single trace based algorithm and the traveltime from all the reflectors for a specific source – receiver offset has to be computed. Convolution in the time domain is computationally intensive and it is usually computed in the frequency domain by Fourier transform. A fast Fourier transform (FFT) algorithm (Press et al., 2002) is best suited to transform wavelet to frequency domain but it is less desirable to transform the reflectivity series because of irregularity in time. One way to suppress irregularity is to interpolate reflectivity in a regular interval using Lagrange interpolation scheme (Gunning and Glinsky, 2004) based on the nearest four samples to the time of a spike. The other way is to use a discrete Fourier transform (DFT) as:

$$U(f) = \sum_{i=0}^{n-1} r_i w(f) e^{j2\pi f \tau_i} \tag{3.45}$$

where f is frequency. This method is effective for the limited number of reflectivity series with which we are dealing in this thesis. The maximum frequency could be the Nyquist frequency but it is usually truncated at the bandwidth of the amplitude spectrum without introducing too much error in the synthetic seismic traces. By inverse Fourier transform seismic traces will be transformed to the time domain.

Synthetic seismic traces are computed for the model in table 2.1 using both exact and approximate reflection and transmission coefficients of plane waves. 96 traces at 25 metre interval over 4 second with a sampling interval 2 ms for compressional waves and converted shear waves are modelled. In all of the synthetic seismic traces transmission effect has been considered. A minimum phase wavelet with dominant frequency 25 Hz is used in the convolution model (Figure 3.8).

The first two panels from the left in Figure 3.9 illustrate the farthest traces at offset 2550 metre for compressional and converted shear waves. The blue traces are exact plane wave and red traces show approximate plane waves. The Rüger approximation for a shallow reflector in converted shear waves panel is quite different from the plane waves with exact R/T coefficients. Figure 3.3 demonstrates this difference well where the angle of incidence is high. The Rüger approximation for P-waves is quite similar to the plane waves trace with exact R/T coefficients. Panel 3 and panel 4 illustrate the primary wavefield for both compressional and converted shear waves for the farthest offset of 2550 m. and near offset at 200 m. The change in amplitude is very obvious.

Figure 3.10 illustrates the synthetic compressional wavefield over 96 traces separated at 25 metre intervals. The left panel shows the plane wave with exact R/T coefficients while the middle one shows Rüger approximation. The right panel shows the difference in amplitude between the two. Except for large angles of incidents, the approximation is close to the plane waves with exact R/T coefficients. All traces in three panels are displayed with the same gain. Figure 3.11 illustrates the synthetic converted shear wavefield over the same geometry as in Figure 3.10. Because of the conversion point, the angle of incidence is much larger in $P \to S_v$ than in $P \to P$ reflections; hence Rüger approximation is a good approximation of the exact plane waves with exact R/T

coefficients. In Figure 3.12 we have illustrated the complete primary wavefield of compressional and converted shear waves for the same geometry as in the previous figure. The residual of the plane waves with exact and approximate R/T coefficients for converted shear waves is bigger than for P-wave reflections at large angles of incidence. In the chapter six a resample of 48 traces (with trace interval of 50 metre) from each of the P-waves and combined P-waves and SV-waves shot gathers which have been computed using Rüger approximation will be considered as observed seismic data for parameter estimation.

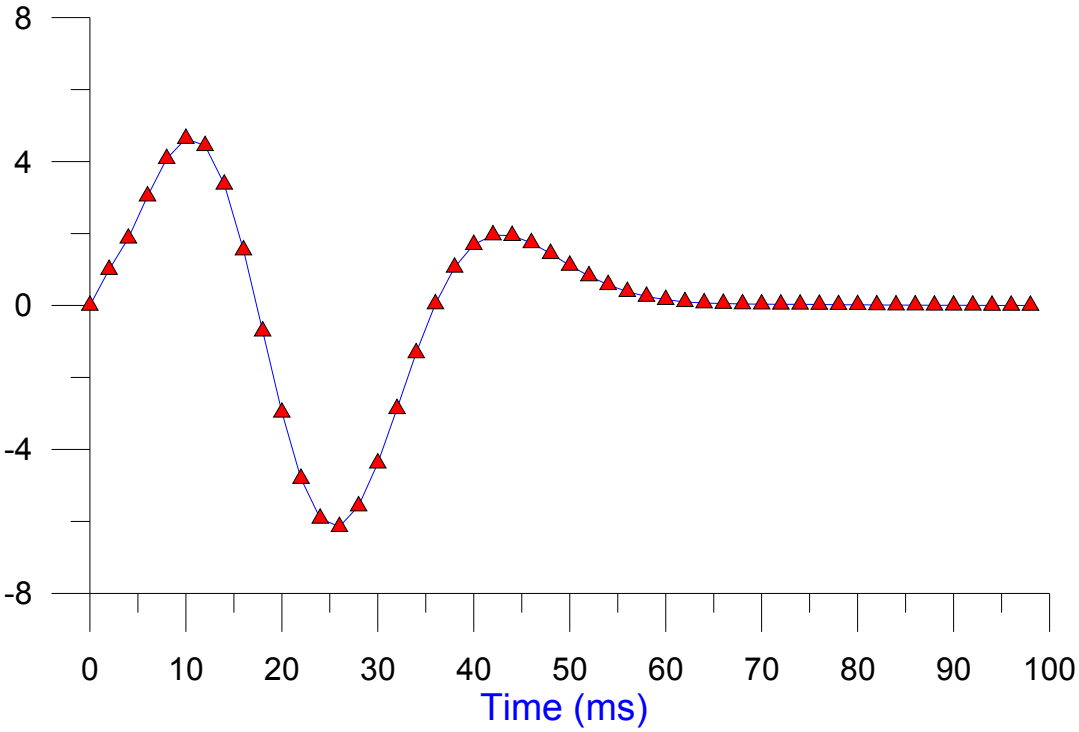


Figure 3.8: A minimum phase wavelet with a 25 Hz dominant frequency and 2ms sampling interval. I use this wavelet whenever we need to compute a synthetic trace.

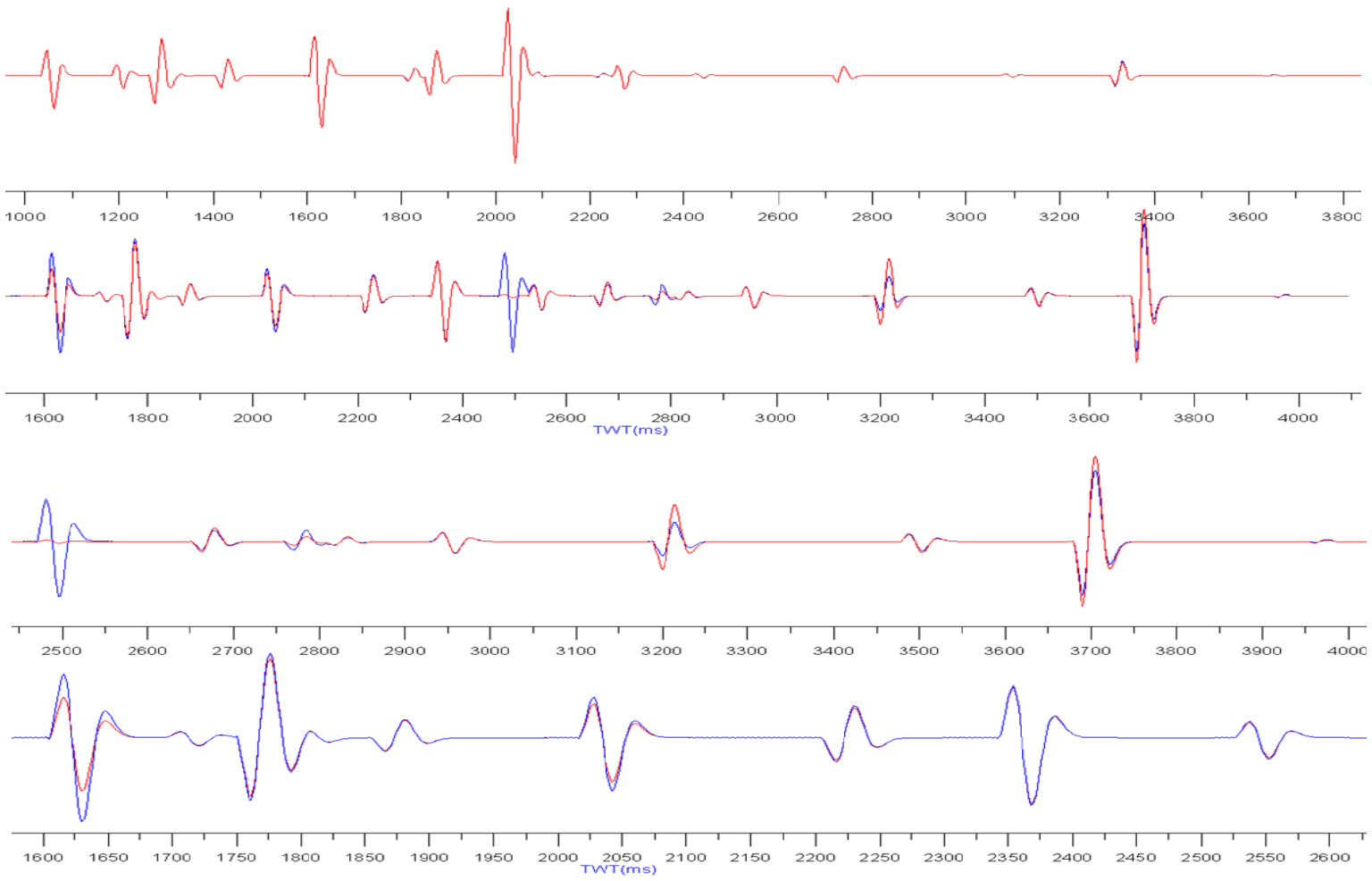


Figure 3.9: Exact (blue) and approximate (red) plane wave of single traces. From top: The first and second traces show near (offset = 200 m) and far (offset = 2550 m) offset traces of mixed compressional and converted shear waves. the third and fourth traces show converted shear waves and compressional waves separately for the farthest offset at 2550 meter.

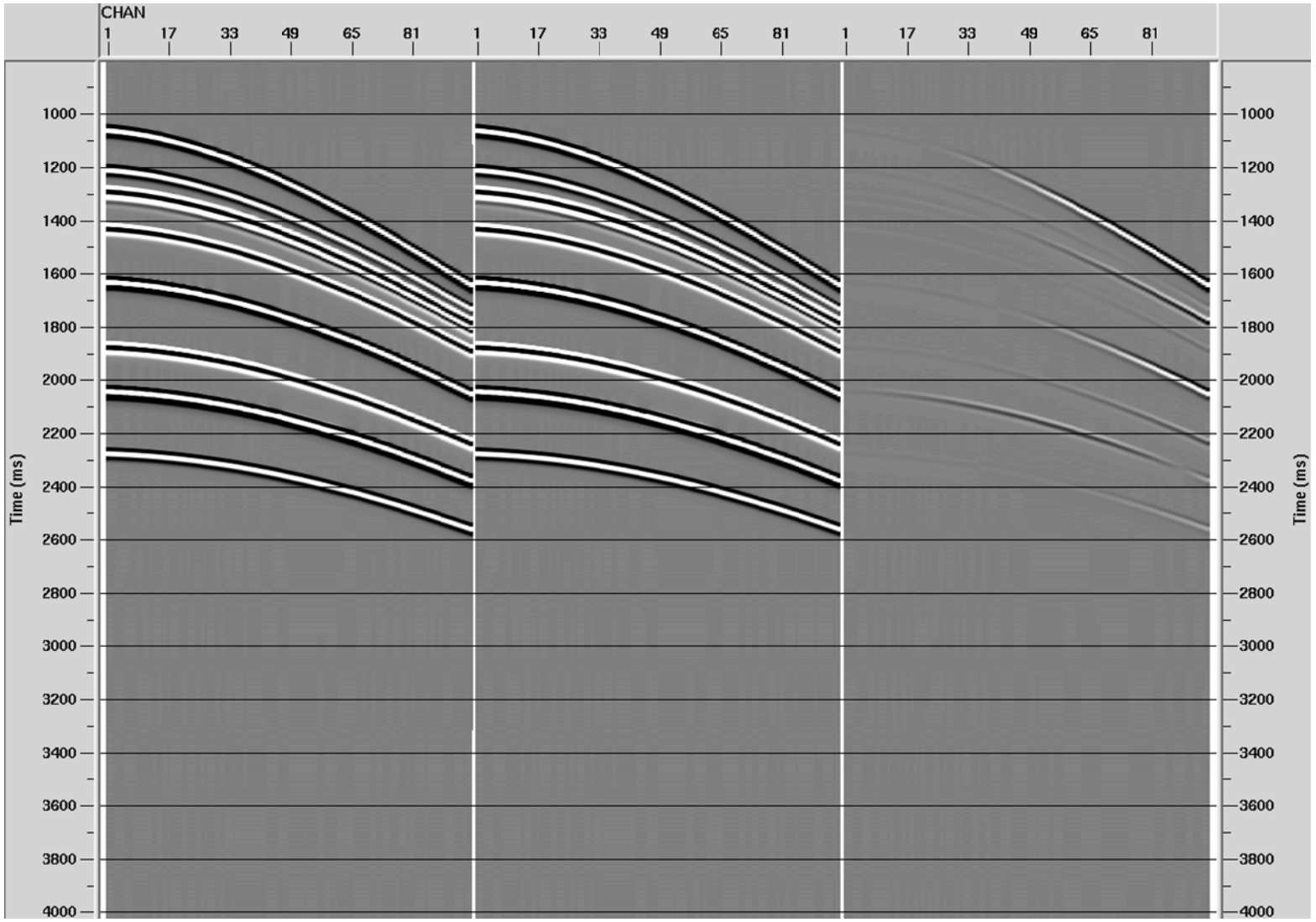


Figure 3.10: Exact (left) and approximate (middle) compressional plane waves. The right panel illustrate the difference between exact and approximate wavefield.

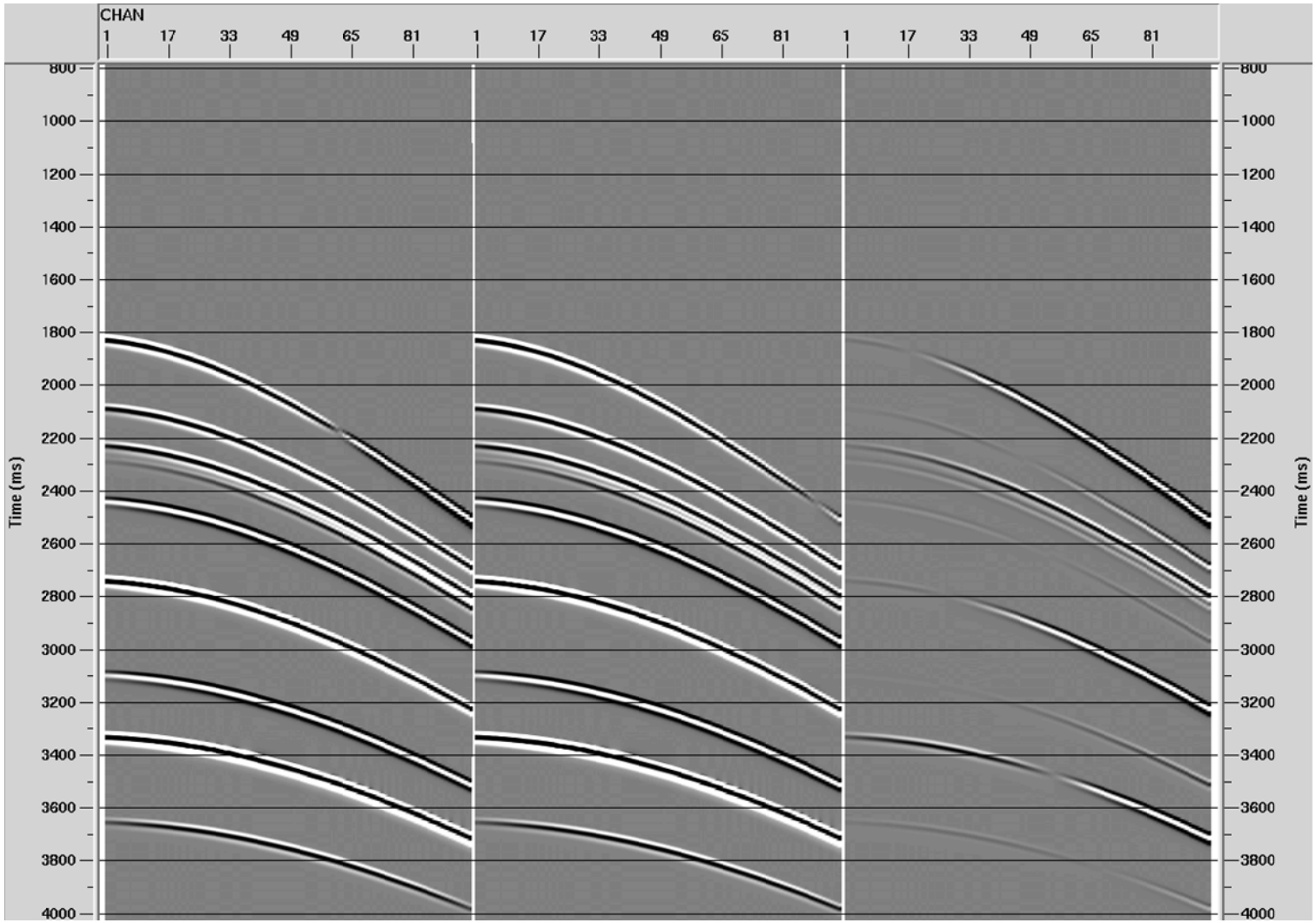


Figure 3.11: Exact (left) and approximate (middle) converted shear plane waves. The right panel illustrate the difference between exact and approximate wavefield.

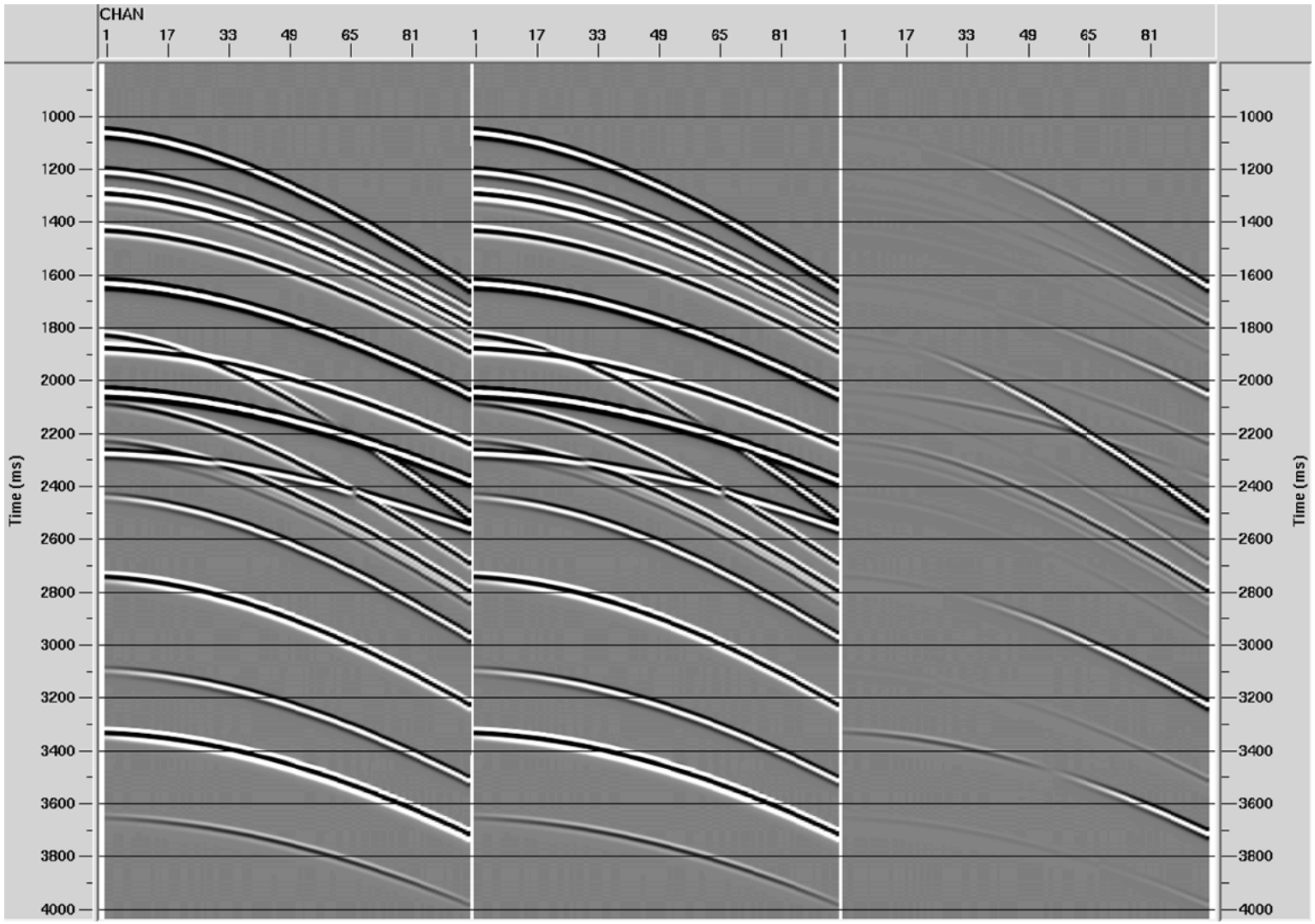


Figure 3.12: Primary and converted shear wavefield of exact (left) and approximate (middle) plane waves. The right panel illustrate the difference between exact and approximate wavefield

References

Aki, K., and P. G. Richards, 1980, Quantitative Seismology: Theory and methods: W. H. Freeman.

Bamford, D., and S. Crampin, 1977, Seismic anisotropy - the state of the art: Geophys. J. R. astr. Soc, **49**, 1-8.

Banik, N. C., 1987, An effective anisotropy parameter in transversely isotropic media: Geophysics, **52**, 1654-1664.

Buchwald, V. T., 1959, Elastic Waves in Anisotropic Media: Proc. R. Soc. A., **253**, 563-580.

Ĉervenŷ, V., 2001, Seismic ray theory: Camridge Unversity Press.

Daley, P. F., and F. Hron, 1977, Reflection and transmission coefficients for transversely isotropic media: Bull. Seismic Soc. Am., **67**, 661-675.

Graebner, M., 1992, Plane-wave reflection and transmission coefficients for a transversley isotropic solid: Geophysics, **57**, 1512-1519.

Gunning, J., and M. E. Glinsky, 2004, Delivery: an open-source model-based Bayesian seismic inversion program: Computers & Geosciences, **30**, 619-636.

Jílek, P., 2002, Converted PS-wave reflection coefficients in weakly anisotropic media, *in* I. Pšenčik and V. Červený, eds., Seismic waves in laterally inhomogeneous media: Birkhäuser Verlag, 1527-1562.

Keith, C. M., and S. Crampin, 1977, Seismic body waves in anisotropic media: reflection and refraction at a plane interface: Geophys. J. R. astr. Soc, **49**, 181-208.

Musgrave, M. J. P., 1954a, On the propagation of elastic waves in aeolotropic media. I. General principles: Proc. R. Soc. A., **226**, 339-355.

Musgrave, M. J. P., 1954b, On the propagation of elastic waves in aeolotropic media. II. Media of hexagonal symmetry: Proc. R. Soc. A., **226**, 356-366.

Musgrave, M. J. P., 1959, The propagation of elastic waves in crystals and other anisotropic media: Rep. Prog. Phys, **22**, 74-96.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 2002, Numerical recipes in C++: the art of scientific computing: Cambridge University Press.

Pšenčik, I., and V. Vavryčuk, 1998, Weak contrast PP wave displacement R/T coefficients in weakly anisotropic elastic media: Pure and Applied Geophysics, **151**, 699-718.

Rüger, A., 1996, Reflection coefficients and azimuthal AVO analysis in anisotropic media: PhD, Colorado School of Mines.

Rüger, A., 1998, Variation of P-wave reflectivity with offset and azimuth in anisotropic media: Geophysics, **63**, 935-947.

Shuey, R. T., 1985, A simplification of the Zeoppritz equations: Geophysics, **50**, 609-614.

Slawinsky, M. A., 2003, Seismic waves and rays in elastic media: Elsevier Science Ltd.

Stovas, A., and B. Ursin, 2003, Reflection and Transmission responses of layered transversely isotropic viscoelastic media: Geophysical Prospecting, **51**, 447-477.

Thomsen, L., 1986a, Weak elastic anisotropy: Geophysics, **51**, 1954-1966.

Thomsen, L., 1990, Poisson was not a geophysicist!: The Leading Edge, **9**, 27-29.

Thomsen, L., 1993, Weak anisotropic reflections, *in* J. P. Castagna and M. E. Backus, eds., Offset-dependent reflectivity-Theory and practice of AVO analysis: Society of Exploration Geophysicists, 1, 103-111.

Tsvankin, I., 1996, P-wave signatures and notation for transversely isotropic media: An overview: Geophysics, **61**, 467-483.

Varyč, V., and I. Pšenčik, 1998, PP-wave reflection coefficients in weakly anisotropic elastic media: Geophysics, **63**, 2129-2141.

Waters, K. H., 1981, Reflection Seismology: a tool for energy resource exploration: John Wiley & Sons Inc.

White, J. E., 1983, Underground sound : application of seismic waves: Elsevier.

White, R., and R. Simm, 2003, Tutoria: Good practice in well ties: First Break, **21**, 75-83.

Zoeppritz, K., 1919, Erdbedbenwellen, on the reflection and penetration of seismic waves through unstable layers.: Göttingen Nachrichten, **1(VII B)**, 66-84.

Chapter 4

A review of parameter estimation

Introduction

Understanding the structure and the physical properties of earth's subsurface is one of the ultimate goals of geophysical exploration. These physical properties could be the elastic or other properties such as electrical, gravitational or electromagnetic. A mathematical model relates the potential fields, created naturally or by interaction with an external field, to their physical properties. By a finite number of measurements which may be inaccurate or even incomplete, we aim to understand the system under study. Unfortunately, most of the physical problems are nonlinear in nature, mathematical models are approximate and data are insufficient, therefore, a prior knowledge of the system is always required. Inverse modelling provides a framework to use all the information to realize the system. In this chapter we review first the conventional constrained nonlinear least square minimization method which is very popular in geophysical inversion and provide an inverse model to estimate the elastic properties in a system of transversely isotropic layered media. In the next section we will concentrate mainly on unconstrained minimization methods which have been used widely in this thesis. During the review of these methods we bring a few examples of Thomsen's anisotropy parameters estimation, to emphasize the efficiency of different unconstrained minimization algorithms.

4.1 The anatomy of inverse problems

Any geophysical inverse problem consists of at least three components; a model parameter vector which consist of a set of unknown physical parameters that are to be estimated from a model space, a set of measured data which are usually contaminated with noise, and the physics of the problem which map model to data space. The physics is also called the forward operation. The process of estimation of the model parameters from the measured data in the geophysical community is usually referred to as the inverse problem or parameter estimation. An earth model has infinitely many degrees of freedom

and the number of measurements is always finite, so mapping from a finite data space to an infinite model space using the inverse operator is always not unique even with noise free data. Moreover, the physics governing geophysical problems is intrinsically nonlinear and this worsens the nonuniqueness. This makes the inverse problem to be considered as an inference or estimation problem. Due to nonlinearity as well as parameterization and regularization, the estimated parameter model may not be exactly the true model, so error estimation needs to be considered. This part is called the appraisal problem (Scales and Snieder, 2000). A schematic of the anatomy of an inverse problem is illustrated in Figure 4.1.

For linear problems, the theory of error estimation is well developed and estimators, such as the model resolution matrix and covariance matrix, quantify the error propagation in model space due to errors in data and confidence level in the model parameter (Jackson, 1972; Menke, 1989; Berryman, 1994; Tarantola, 2005). There is no solid theory to quantify errors in high dimensional nonlinear problems, however, using linearization scheme such as a first order Taylor approximation, one can linearize the forward problem around a prior model and use the same analysis for linear problems (Tarantola, 2005). For small scale nonlinear inverse problems, error estimation would be possible by sampling from the posterior model space using computationally intensive algorithms such as Markov Chain Monte Carlo, and putting the prior probability and likelihood function in a Bayesian framework (Scales and Tenorio, 2001). Posterior model space is the product of a likelihood function and prior probability density function (pdf) and has a narrower distribution than the prior pdf which is scaled by the marginal distribution function.

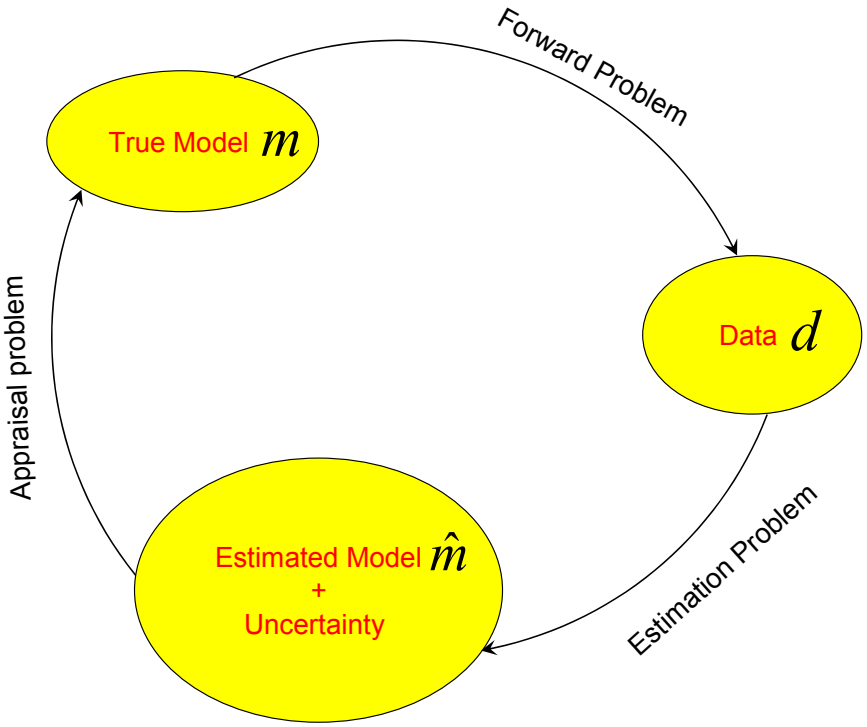


Figure 4.1: A schematic view of an inverse problem (After Scales and Snieder (2000)).

4.2 Nonlinear least square optimization methods

Least square algorithms, are by far, the most widely used minimization algorithms in linear and nonlinear geophysical problems. They are easy to implement and are robust enough, but they may not converge if the prior model is too far from the solution. Gauss-Newton and Levenberg-Marquardt (Nocedal and Wright, 1999) are two varieties of least square method which will be reviewed here.

Any inverse problem tries to optimize an objective function of the sum of all the residual errors. The shape of the objective function has the greatest importance and for each inverse problem it may be different. We may include structural information as well as data residual (Oldenburg and Li, 2005). A general form of an objective function for unconstrained minimization problems may be written as:

$$\varphi = \varphi_d + \beta\varphi_m \quad . \tag{4.1}$$

where

$$\begin{aligned} \phi_d &= e^T e \,, \\ e &= d_0 - d \,, \end{aligned} \tag{4.2}$$

$$d = Gm \,, \quad Linear\ problem \quad , \tag{4.3}$$

$$d = g(m), \quad Nonlinear\ problem \tag{4.4}$$

where m is the model parameter, d is the synthetic data, d_0 is the measured data, G is the forward model, e is the residual error, and T is the transpose operator. β is a stabilizer and generally is called the regularization parameter. β can be constant $(\beta \geq 0)$ but it may be decreased monotonically as iteration number increases. As an alternative, the value of β can be estimated by root finding of the objective function (Nocedal and Wright, 1999). φ_m is the model structure and can simply be the discrepancy from a prior model (Tarantola, 2005),

$$\varphi_m = \left(m_0 - m\right)^T \left(m_0 - m\right) \quad . \tag{4.5}$$

Depending on the specific geophysical problem, φ_m can be more sophisticated (Oldenburg and Li, 2005) . In underdetermined systems (4.3), where the number of model parameters is far more than the number of independent data, φ_m has an important effect on reducing the nonuniqueness. Underdetermined systems arise frequently in 3-dimensional discretized parameters such as velocity tomography, geoelectrical resistivity tomography, and other potential field problems.

Uncertainty should be incorporated in the model and data space of the objective function through the covariance matrices (Tarantola, 2005). Hence, the equations (4.2) and (4.5) should be rewritten as:

$$\varphi_d = e^T C_D^{-1} e \quad , \tag{4.6}$$

$$\varphi_m = \left(m_0 - m\right)^T C_M^{-1} \left(m_0 - m\right) \quad . \tag{4.7}$$

where C_D and C_M are covariance matrices of the data and model parameter respectively. These matrices are generally diagonal and are simply the variances (assuming there is no

correlation between model parameters). If, for example, data are traveltimes, then the square of the errors in picking traveltime can be considered as the data variances. Similarly if the model parameters are velocities, we can use the square of half of the difference between upper and lower limits, as the model variance for each model parameter.

We have discussed the objective function from the consideration of a general inverse theory. Now we look at the objective function from a maximum likelihood point of view. Least square minimization is based on the assumption of a Gaussian distribution of both model parameter and data. If we assume that prior information of model parameters has the mean m_0 and covariance matrix C_M, with the assumption of the Gaussian distribution (in fact for many geophysical problem, where the model parameters are positive, the distribution is log normal) the prior probability density function is defined as:

$$\rho_M = \frac{1}{\left((2\pi)^M |C_M|\right)^{1/2}} \exp\left[-\frac{1}{2}(m-m_0)^T C_M^{-1}(m-m_0)\right] . \tag{4.8}$$

M is the number of model parameters and $|C_M|$ is the determinant of C_M. By incorporating regularization parameter and a smoother matrix, Ulrych et al. (2001) give a different pdf for the prior model. A Gaussian probability density function for data with the mean d and covariance matrix C_D is defined as:

$$\rho_D = \frac{1}{\left((2\pi)^N |C_D|\right)^{1/2}} \exp\left[-\frac{1}{2}(d-g(m))^T C_D^{-1}(d-g(m))\right] . \tag{4.9}$$

where N is the number of data and $|C_D|$ is the determinant of C_D. Using the forward model (4.3) relating data d and model parameter m and putting all the probability density functions in terms of Bayes theorem the posterior density function σ_M can be written as (Ulrych, Sacchi and Woodbury, 2001; Tarantola, 2005):

$$\sigma_M = \frac{\rho_D \rho_M}{\int \rho_D \rho_M dm} , \tag{4.10}$$

$$\sigma_M = const. \exp\left(-\frac{1}{2}\varphi(m)\right) . \tag{4.11}$$

where φ is the objective function (4.1). If the forward problem is linear or weakly nonlinear or if the objective function is quadratic, the posterior density function σ_M is Gaussian. The further the forward problem is away from linearity or the objective function is from being quadratic, the further σ_M will be from Gaussian distribution.

In linear problems, the mean of the posterior density function is the solution, however, in nonlinear systems we have to explore the whole distribution to estimate the empirical mean of the posterior. Sampling methods such as Markov Chain Monte Carlo, are very efficient in low dimensional problems in estimating the mean of the posterior but in high dimensional problems it could be computationally intensive. By assuming a weak nonlinearity, one tries to find the solution where the posterior pdf is a maximum (maximum likelihood point). Maximizing the posterior density function means minimizing the objective function, so in the weakly nonlinear systems the mode of the posterior probability density function can be considered as a solution.

In the inverse problems, where the objective function is quadratic, hence the posterior density function is Gaussian; there should be a point $\tilde{m}$ with covariance matrix $\tilde{C}_M$ such that the posterior density function (4.11) can be expressed as (Tarantola, 2005):

$$\sigma_M = cons{t}s.\exp\left(-\frac{1}{2}\left(m-\tilde{m}\right)^T \tilde{C}_M^{-1}\left(m-\tilde{m}\right)\right) . \tag{4.12}$$

Maximizing (4.12) with respect to *m* gives the following expression for the mean and covariance matrix of the posterior density function, assuming the exact theory (4.12), is given by Tarantola (2005) and Menke (1989) as:

$$\tilde{m} = m_0 + C_M G^T \left(G C_M G^T + C_D\right)^{-1}\left(d_{obs} - G m_0\right) , \tag{4.13}$$

$$\tilde{C}_M = C_M - C_M G^T \left(G C_M G^T + C_D\right)^{-1} G C_M . \tag{4.14}$$

So far we have been mostly concerned about the linear least square problems from both a general inverse theory and maximum likelihood point of view. Now we will be talking about linearizing the nonlinear least square methods and problems associated with rank deficiency and ill-conditioning and stabilizing them using singular value

decomposition (SVD) and Levenberg-Marquardt algorithms for regularization. We will estimate the solution of a nonlinear inverse problem by looking at the iterative methods.

Let us assume that m^0 is an initial estimate of m and d^0 is the nonlinear equation (4.3) evaluated at m^0, a first order Taylor expansion of $d = g(m)$ around m^0 can be expressed as (Lines and Treitel, 1984):

$$d = d^0 + \sum_{j=1}^{M} \left.\frac{\partial d}{\partial m_j}\right|_{m=m_0} \left(m_j - m_j^0\right) . \tag{4.15}$$

In matrix notation, (4.15) can be expressed as:

$$d = d^0 + J\Delta m , \tag{4.16}$$

where J is the matrix of partial derivative or Jacobian matrix ($N \times M$, where N is the number of data and M is the number of the model parameters) and $\Delta m = m_j - m_j^0$. The error vector *e,* the difference between the measured data *y* and computed data *d* (simply data residual) can be written as:

$$e = y - d . \tag{4.17}$$

Substitution of (4.16) in (4.17) results in:

$$r = y - d^0 = e + J\Delta m . \tag{4.18}$$

where r is the discrepancy vector. By rearranging (4.18), the error vector e can be expressed as:

$$e = r - J\Delta m . \tag{4.19}$$

Minimization of the objective function (4.1) by substituting the error vector (4.19) and differentiation with respect to the Δm can be expressed as:

$$\left(J^T J + \beta I\right)\Delta m = J^T r \tag{4.20}$$

$$\Delta m = \left(J^T J + \beta I\right)^{-1} J^T r \tag{4.21}$$

This is the Levenberg-Marquardt (Marquardt, 1963) algorithm. Due to the β in (4.21) Levenberg-Marquardt algorithm belongs to the class of constrained optimization problem. If $\beta = 0$, the Levenberg-Marquardt algorithm reduces to Gauss-Newton method. The equation (4.20) can be solved by any linear matrix solvers such as Choleski factorization without computing the inverse matrix in (4.21). Despite that, the Gauss-

Newton method generally suffers from rank-deficiency of the $J^T J$ matrix, Levenberg-Marquardt algorithm avoids this weakness by including the positive number β. Including β avoids singularity of the eigenvalues of $J^T J$ and will ensure that $J^T J$ is positive definite but this is not sufficient if the residual errors are not a linear function of model parameters. If the residual vector *r* is large, performance of Gauss-Newton or Levenberg-Marquardt algorithms is poor and will likely fail to converge (Nocedal and Wright, 1999). This is the result of neglecting the second derivative term (Hessian matrix) in (4.15), which will be significant if the residuals are large. Sufficiency for minimization of the objective function means that the Hessian matrix has to be positive definite (Fletcher, 1980). By neglecting the second component of the objective function (4.1) and taking its first derivative with respect to model parameter we will get (Lines and Treitel, 1984):

$$\frac{\partial \varphi}{\partial m_j} = 2\sum_{i=1}^{N} \frac{\partial e_i}{\partial m_j} e_i \,, \quad j = 1, 2, ..., M. \,, \tag{4.22}$$

$$A_{jk} = \frac{\partial^2 \varphi}{\partial m_j \partial m_k} = 2\sum_{i=1}^{N} \left(\frac{\partial e_i}{\partial m_j} \frac{\partial e_i}{\partial m_k} + e_i \frac{\partial^2 e_i}{\partial m_j \partial m_k} \right). \tag{4.23}$$

Since

$$\frac{\partial e_i}{\partial m_j} = -\frac{\partial d_i}{\partial m_j} \,. \tag{4.24}$$

The Hessian matrix (4.23) can be written as:

$$A_{jk} = \frac{\partial^2 \varphi}{\partial m_j \partial m_k} = 2\sum_{i=1}^{N} \left(\frac{\partial d_i}{\partial m_j} \frac{\partial d_i}{\partial m_k} + e_i \frac{\partial^2 e_i}{\partial m_j \partial m_k} \right), \tag{4.25}$$

In matrix notation, (4.25) becomes:

$$A = 2J^T J + 2A' \,, \tag{4.26}$$

where $A' = 2e_i \frac{\partial^2 e_i}{\partial m_j \partial m_k}$.

Equation (4.26) shows that Hessian *A* has two components; in a range of model parameters, where, the objective function is quadratic, $J^T J$ will always be positive definite. If the errors are linear function of model parameters, A' vanishes and the Hessian matrix is only a function of the Jacobian matrix $J^T J$, however, if errors are

nonlinear function of model parameters, A' will not be zero and there is no guarantee that the Hessian will be positive definite; hence minimization might not occur.

In large scale problems, where, the number of data and model parameters is large, computation of $J^T J$ and $J^T r$ not only needs large storage but also causes numerical inaccuracies. Singular value decomposition (Golub and Van Loan, 1996) is a method for analyzing rank-deficiency and / or ill-conditioned matrices (Aster et al., 2005). Suppose for any vector g in the range of J the following linear equation needs to be solved:

$$J\Delta m = g \ , \tag{4.27}$$

where J is an N by M ill-conditioned matrix. J can be decomposed in the following way:

$$J = U\Lambda V^T \ , \tag{4.28}$$

where U is an N by N orthogonal matrix whose columns are unit basis vectors spanning the data space, R^n. V is an M by M orthogonal matrix whose columns are basis vectors spanning the model space, R^m. Λ is an N by M diagonal matrix whose nonnegative elements are called singular values. If there are only P nonzero singular values, the M-P zero elements will be discarded and the rest are arranged in a descending order. Because the N-P element of matrix U and N-P element of matrix V are multiplied by zero, (4.28) will simplify into a compact form (4.29). Therefore, Λ is a P by P diagonal matrix with positive values.

$$J = U_p \Lambda_p V_p^T \ . \tag{4.29}$$

The P columns of matrix U_p form an orthonormal basis for $R(J)$, while the N-P columns of U_0 form an orthonormal basis for the null space of $N(J^T)$. Therefore, $N(J^T) + R(J) = R^n$. $N(J^T)$ is usually referred to as data null space. Similarly, because $J^T = V_p \Lambda_p U_p^T$, the P columns of matrix V_p form an orthonormal basis of $R(J^T)$ and M-P columns of matrix V_0 form an orthonormal basis for the null space of $N(J)$, so $N(J) + R(J^T) = R^m$. $N(J)$ is usually called model null space. There are other properties (4.30) and (4.31) of SVD which can be easily proved (see Aster, Borchers and Thurber (2005), such as,

$$JJ^T u_i = \lambda_i^2 u_i \ , \tag{4.30}$$

$$J^T J v_i = \lambda_i^2 v_i \ . \tag{4.31}$$

These equations show that in theory, one can simply use the above equations to compute the eigenvalues and eigenvectors JJ^T and $J^T J$, but there are more efficient algorithms to do SVD (Golub and Van Loan, 1996; Press et al., 2002). It should be mentioned, however, there are *M-P* null vectors in matrix *V*, but this reduces the dimensionality of the inverse problem and in practice the number of *P* eigenvalues (λ) are kept at the number of model parameters (*M*).

Let us write down the least square problem (4.27) in terms of singular value decomposition by replacing (4.28) in (4.27):

$$U\Lambda V^T \Delta m = g \ , \tag{4.32}$$

Now

$$\Delta m = J^{\dagger} g \ . \tag{4.33}$$

where

$$J^{\dagger} = V\Lambda^{-1} U^T \ , \tag{4.34}$$

is called the generalised inverse.

We can expand (4.33) as a linear combination of columns of basis vectors:

$$\Delta m = \sum_{i=1}^{M} \frac{v_i \left(u_i^T g \right)}{\lambda} \ . \tag{4.35}$$

Very small value of the eigenvalues in the denominator will give a very large weight to the corresponding basis vectors of model space (v_i) and they dominate the solution. In the case of random noise *g* has a nonzero projection onto each component of basis vector of data space (U_i). In the worst case, a generalised inverse is just a noise amplifier and the solution is useless (Aster, Borchers and Thurber, 2005).

The condition number of the matrix *J* is a measure of instability of the generalized inverse. It is the ratio of the largest eigenvalue of diagonal matrix Λ to the smallest one. A large condition number is an index of instability. There are several ways to reduce the condition number. Truncated singular value decomposition (TSVD) is a way to discard the small eigenvalues, but this result in lowering the dimensionality of the problem and

we have to sacrifice some model parameters for the sake of stability. Another way of stabilizing the generalized inverse solution is to compute the singular values of an ill-conditioned matrix which has been stabilized by Levenberg-Marquardt algorithm.

Let us rewrite the normal equation (4.21) in terms of SVD. Substituting (4.28) in (4.21) and by the fact that $V^T V = I$, the generalized inverse solution (see Lines and Treitel (1984) for more detail) can be written as:

$$\Delta m = V \frac{\Lambda}{\Lambda^2 + \beta} U^T g \quad . \tag{4.36}$$

By rewriting the (4.36) in terms of eigenvalues of the diagonal matrix Λ, the generalized inverse solution is as follow:

$$\Delta m = V diag\left(\frac{\lambda_i}{\lambda_i^2 + \beta} \right) U^T g \quad . \tag{4.37}$$

Even in the case of small eigenvalues, the regularization parameter (β) avoids the instability in the solution of the generalized inverse.

We have so far talked about finding the solution of either linear or nonlinear inverse problems using singular value decomposition. In theory, the solution of exact linear inverse problem could be achieved in one step, however, due to noise the objective function is not quite quadratic, so SVD should be repeated until a satisfactory solution is found. In nonlinear inverse problems we rely on linearizing the forward model around an initial model parameter and by successive updating of the model parameter we hope to get to a solution. Due to nonlinearity of the forward model we may converge to a minimum which may not necessary be the global minimum. This process has to be an iterative scheme and as such we update the initial model iteratively until we reach a satisfactory solution. In general, whenever the gradient vector of the objective function (4.1) or forward model is zero, iteration should be stopped. Whenever the structural constrained (the second term) of the objective function (4.1) is neglected, the value of the objective function itself could be used as a criterion to stop the iteration process. Sequence (4.38) summarizes the iterative process (Sen and Stofa , (1995) after some modification).

$\varepsilon = tolerance$,

Loop over iteration $i = 1, \ldots n$

$d_i = g(m_i)$, synthetic data vector

$e_i = y_i - d_i$, data residual

Compute φ, The objective function (4.38) .

$J = \dfrac{\partial d}{\partial m}$, gradient vector

$\Delta m_i =$ Computed by SVD, model parameter update

$m_i = m_i + \Delta m_i$, new model parameter

If $(J = 0)$ or $(\varphi < \varepsilon)$ or $i = n$ *Exit*

Continue

4.3 Least square inversion of traveltimes in a multi layered medium

In a layered, transversely isotropic medium (Figure 4.2), traveltimes (either P-waves or SV-waves) are a function of vertical compressional and shear wave velocities, Thomsen's anisotropy parameters (δ, ε), and layer thickness for a given offset and reflector. These properties which, hereafter, will be called model parameters are assumed to be constant laterally. For a specific offset and a stack of $M+1$ layers, M traveltimes will be recorded (assuming the thickness of each layer compared to the wavelength is large enough). As each ray travels through all the layers down to the reflector, each layer has already been hit by elastic waves. The lowest layer which is considered as reflector has only been hit once; while the top layer has been hit by the number of ray paths and the rest are in between. This way of considering the impact of each ray on each layer down to the reflector is different from layer stripping inversion, where each ray path is treated individually and the effects of other ray paths from the upper reflectors are neglected. This will improve the robustness of the upper layer parameters. Simultaneously incorporating all the information is more complicated in terms of configuring a nonlinear least square method.

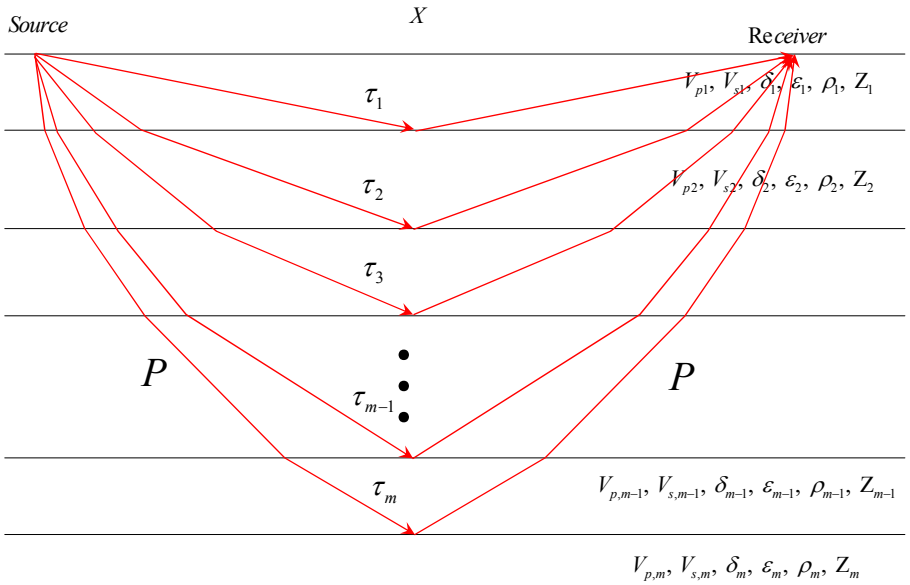


Figure 4.2: Reflection traveltimes of compressional waves in a layered medium.

Following the notation used in previous sections, we express traveltimes by the vector d_j^i, where index i denotes the traveltime for a specific offsets and j refers to the traveltime from each reflector. The model parameter is expressed as m_j^k, where index j refers to layer number and k applies to each model parameter from different type (such as V_p, V_s, *etc*). The matrix of partial derivatives (Jacobian matrix) and its transpose are expressed in (4.39) and (4.40) respectively. The Jacobian matrix is a blocky sparse matrix composed of lower triangular blocks of M by M. The first column of each block is the derivative of the traveltime from a specific offset with respect to a model parameter from the first layer (top layer). The second column belongs to a model parameter in the second layer. The first element of this column is zero because the ray path reflection from the top of the second layer has no effect on the model parameters in the second layer. In the last column only the last element is nonzero, because only one ray path has contributed to the corresponding layer. The vertical columns belong to traveltimes from different offsets,

$$
J=\begin{bmatrix}
\frac{\partial d_1^1}{\partial m_1^1} & 0 & & 0 & \frac{\partial d_1^1}{\partial m_1^2} & 0 & & 0 & \frac{\partial d_1^1}{\partial m_1^k} & 0 & & 0 \\
\frac{\partial d_2^1}{\partial m_1^1} & \frac{\partial d_2^1}{\partial m_2^1} & & 0 & \frac{\partial d_2^1}{\partial m_1^2} & \frac{\partial d_2^1}{\partial m_2^2} & & 0 & \frac{\partial d_2^1}{\partial m_1^k} & \frac{\partial d_2^1}{\partial m_2^k} & & 0 \\
 & \vdots & & & & \vdots & & & & \vdots & & \\
\frac{\partial d_M^1}{\partial m_1^1} & \frac{\partial d_M^1}{\partial m_2^1} & \cdots & \frac{\partial d_M^1}{\partial m_M^1} & \frac{\partial d_M^1}{\partial m_1^2} & \frac{\partial d_M^1}{\partial m_2^2} & \cdots & \frac{\partial d_M^1}{\partial m_M^2} & \frac{\partial d_M^1}{\partial m_1^k} & \frac{\partial d_M^1}{\partial m_2^k} & \cdots & \frac{\partial d_M^1}{\partial m_M^k} \\
\frac{\partial d_1^2}{\partial m_1^1} & 0 & & 0 & \frac{\partial d_1^2}{\partial m_1^2} & 0 & & 0 & \frac{\partial d_1^2}{\partial m_1^k} & 0 & & 0 \\
\frac{\partial d_2^2}{\partial m_1^1} & \frac{\partial d_2^2}{\partial m_2^1} & & 0 & \frac{\partial d_2^2}{\partial m_1^2} & \frac{\partial d_2^2}{\partial m_2^2} & & 0 & \frac{\partial d_2^2}{\partial m_1^k} & \frac{\partial d_2^2}{\partial m_2^k} & & 0 \\
 & \vdots & & & & \vdots & & & & \vdots & & \\
\frac{\partial d_M^2}{\partial m_1^1} & \frac{\partial d_M^2}{\partial m_2^1} & \cdots & \frac{\partial d_M^2}{\partial m_M^1} & \frac{\partial d_M^2}{\partial m_1^2} & \frac{\partial d_M^2}{\partial m_2^2} & \cdots & \frac{\partial d_M^2}{\partial m_M^2} & \frac{\partial d_M^2}{\partial m_1^k} & \frac{\partial d_M^2}{\partial m_2^k} & \cdots & \frac{\partial d_M^2}{\partial m_M^k} \\
 & \vdots & & & & & & & & & & \\
\frac{\partial d_1^N}{\partial m_1^1} & 0 & & 0 & \frac{\partial d_1^N}{\partial m_1^2} & 0 & & 0 & \frac{\partial d_1^N}{\partial m_1^k} & 0 & & 0 \\
\frac{\partial d_2^N}{\partial m_1^1} & \frac{\partial d_2^N}{\partial m_2^1} & & 0 & \frac{\partial d_2^N}{\partial m_1^2} & \frac{\partial d_2^N}{\partial m_2^2} & & 0 & \frac{\partial d_2^N}{\partial m_1^k} & \frac{\partial d_2^N}{\partial m_2^k} & & 0 \\
 & \vdots & & & & \vdots & & & & \vdots & & \\
\frac{\partial d_M^N}{\partial m_1^1} & \frac{\partial d_M^N}{\partial m_2^1} & \cdots & \frac{\partial d_M^N}{\partial m_M^1} & \frac{\partial d_M^N}{\partial m_1^2} & \frac{\partial d_M^N}{\partial m_2^2} & \cdots & \frac{\partial d_M^N}{\partial m_M^k} & \frac{\partial d_M^N}{\partial m_1^k} & \frac{\partial d_M^N}{\partial m_2^k} & \cdots & \frac{\partial d_M^N}{\partial m_M^k}
\end{bmatrix}
$$

(4.39) .

while the horizontal blocks belong to different type of the model parameters. The Jacobian matrix (J) is a rectangular $(N\times M, M\times K)$ sparse matrix, with $KNM(M-1)/2$ null elements. The sparsity of the Jacobian matrix is $\frac{M-1}{2M}$, which, in the limit, reaches 50 percent when *M* goes to infinity.

$$J^T = \begin{bmatrix}
\frac{\partial d_1^1}{\partial m_1^1} & \frac{\partial d_2^1}{\partial m_1^1} & & \frac{\partial d_M^1}{\partial m_1^1} & \frac{\partial d_1^2}{\partial m_1^1} & \frac{\partial d_2^2}{\partial m_1^1} & & \frac{\partial d_M^2}{\partial m_1^1} & \frac{\partial d_1^N}{\partial m_1^1} & \frac{\partial d_2^N}{\partial m_1^1} & & \frac{\partial d_M^N}{\partial m_1^1} \\
0 & \frac{\partial d_2^1}{\partial m_2^1} & \cdots & \frac{\partial d_M^1}{\partial m_2^1} & 0 & \frac{\partial d_2^2}{\partial m_2^1} & \cdots & \frac{\partial d_M^2}{\partial m_2^1} & \cdots 0 & \frac{\partial d_2^N}{\partial m_2^1} & \cdots & \frac{\partial d_M^N}{\partial m_2^1} \\
0 & 0 & & \vdots & 0 & 0 & & \vdots & 0 & 0 & & \vdots \\
0 & 0 & & \frac{\partial d_M^1}{\partial m_M^1} & 0 & 0 & & \frac{\partial d_M^2}{\partial m_M^1} & 0 & 0 & & \frac{\partial d_M^N}{\partial m_M^1} \\
\frac{\partial d_1^1}{\partial m_1^2} & \frac{\partial d_2^1}{\partial m_1^2} & & \frac{\partial d_M^1}{\partial m_1^2} & \frac{\partial d_1^2}{\partial m_1^2} & \frac{\partial d_2^2}{\partial m_1^2} & & \frac{\partial d_M^2}{\partial m_1^2} & \frac{\partial d_1^N}{\partial m_1^2} & \frac{\partial d_2^N}{\partial m_1^2} & & \frac{\partial d_M^N}{\partial m_1^2} \\
0 & \frac{\partial d_2^1}{\partial m_2^2} & \cdots & \frac{\partial d_M^1}{\partial m_2^2} & 0 & \frac{\partial d_2^2}{\partial m_2^2} & \cdots & \frac{\partial d_M^2}{\partial m_2^2} & \cdots 0 & \frac{\partial d_2^N}{\partial m_2^2} & \cdots & \frac{\partial d_M^N}{\partial m_2^2} \\
0 & 0 & & \vdots & 0 & 0 & & \vdots & 0 & 0 & & \vdots \\
0 & 0 & & \frac{\partial d_M^1}{\partial m_M^2} & 0 & 0 & & \frac{\partial d_M^2}{\partial m_M^2} & 0 & 0 & & \frac{\partial d_M^N}{\partial m_M^2} \\
 & & & & & & \vdots & & & & & \\
\frac{\partial d_1^1}{\partial m_1^k} & \frac{\partial d_2^1}{\partial m_1^k} & & \frac{\partial d_M^1}{\partial m_1^k} & \frac{\partial d_1^2}{\partial m_1^k} & \frac{\partial d_2^2}{\partial m_1^k} & & \frac{\partial d_M^2}{\partial m_1^k} & \frac{\partial d_1^N}{\partial m_1^k} & \frac{\partial d_2^N}{\partial m_1^k} & & \frac{\partial d_M^N}{\partial m_1^k} \\
0 & \frac{\partial d_2^1}{\partial m_2^k} & \cdots & \frac{\partial d_M^1}{\partial m_2^k} & 0 & \frac{\partial d_2^2}{\partial m_2^k} & \cdots & \frac{\partial d_M^2}{\partial m_2^k} & \cdots 0 & \frac{\partial d_2^N}{\partial m_2^k} & \cdots & \frac{\partial d_M^N}{\partial m_2^k} \\
0 & 0 & & \vdots & 0 & 0 & & \vdots & 0 & 0 & & \vdots \\
0 & 0 & & \frac{\partial d_M^1}{\partial m_M^k} & 0 & 0 & & \frac{\partial d_M^2}{\partial m_M^k} & 0 & 0 & & \frac{\partial d_M^N}{\partial m_M^k}
\end{bmatrix}$$

(4.40) .

Vector g (4.41) is a blocky vector, composed of N sub vectors each with the size M elements. The index i in g_j^i refers to a specific offset, while index j denotes the ray path reflecting from the top of each layer. The solution vector Δm (4.42) is composed of k sub vectors m_j, each of size M elements. k refers to the number of different types of

parameters to be estimated. Equations (4.39) - (4.42) form a nonlinear least square approach to the traveltime inversion problem in a multi layered media.

$$g^T = \begin{bmatrix} g_1^1 & g_2^1 & \cdots & g_M^1 & g_1^2 & g_2^2 & \cdots & g_M^2 & \cdots & g_1^N & g_2^N & \cdots & g_M^N \end{bmatrix}$$

(4.41) ,

$$\Delta m^T = \begin{bmatrix} m_1^1 & m_2^1 & \cdots & m_M^1 & m_1^2 & m_2^2 & \cdots & m_M^2 & \cdots & m_1^k & m_2^k & \cdots & m_M^k \end{bmatrix}$$

(4.42) .

4.4 Unconstrained minimization using iterative gradient methods

There are two major groups of unconstrained algorithms for multidimensional minimization using the first derivatives; conjugate gradient and variable metric methods. Conjugate gradient is well suited for large scale problems with storage of the order of M model parameters in which the second derivatives do not necessarily need to be found. Variable metric methods instead use the storage of the order of M^2 and second derivatives need to be calculated. Due to implementing the Hessian matrix, variable metric is faster than the general form of the conjugate gradient but this limits the Hessian based minimization algorithms for problems where the prior information is close enough to the solution. In this section we review the multidimensional line minimization which will be followed by a review of conjugate gradient and variable metric method.

Gradient based minimization algorithms use a descent direction h of the objective function f to find a point where the objective function is minimum. Direct search and gradient based methods are two major categories in line search method. Direct search methods like golden section or bisection by successive bracketing find the minimum of a function. They are robust but require more function evaluation. On the other hand, gradient based line search algorithm try to find the minimum of $\varphi(\alpha) = f(x + \alpha h)$ by successive evaluation of φ at different α (step length). In this section, for consistency with mathematical literature, we use x as a variable representing the model parameter m which we are dealing with.

In the linear problems, where the objective function is quadratic, φ can be minimized directly for the α along a descent direction. If the objective function is far from being quadratic, α needs to be computed using iterative line search methods. Inexact line search methods based on Armijo or Wolf conditions (Nocedal and Wright, 1999) can effectively be used for minimization algorithms such as variable metric or Newton method. Steepest descent and conjugate gradient algorithms need a more accurate line search or exact line search such as the Brent method (Brent, 1973). In practice, we initially bracket the minimum to a reasonable interval using, for example, golden section and then use either one of the inexact or inverse quadratic search methods

to find the minimum of φ. This approach is more suitable for nonlinear problems where the objective function is not quadratic.

4.4.1 Golden section search method

The Golden section search is a direct search method which, by successive bracketing, eliminates regions which fall outside a triple point sequence in which the objective function value is larger than the triple points. The amount of jump to a bracket is equal to the golden ratio τ. Suppose a function has a minimum in the range of 0-1. We choose two symmetric points from this range where their distances from either points is equal to τ (Figure 4.3 a). In the first step we eliminate the right end point assuming that the function is larger than other points. Now, the range is limited to τ (Figure 4.3 b). We draw a point with the distance τ from the right end point. This distance is in fact τ^2 in the range of 0-1. To retain the symmetry the inner point has to be symmetric to either end point, hence $\tau^2 = 1 - \tau$. This is a quadratic function (4.43) and the positive solution of τ is golden ratio.

$$\begin{aligned} &\tau^2 + \tau - 1 = 0 \\ &\tau = \frac{-1+\sqrt{5}}{2} = 0.61803... \end{aligned} \quad . \tag{4.43}$$

Figure 4.4 illustrates an example of golden section search bracketing (Ravindran et al., 2006). Function f is bracketed initially at two end points 1 and 2. The first two evaluations are located at points 3 and 4. Since the function f at point 4 is smaller than in point 3 and also point 4 is smaller than point 3, the region beyond point 3 is eliminated. In the next step point 5 will be chosen based on golden ratio τ measured from point 3. The function f at 5 is greater than point at 4, so the region before point 5 will be eliminated and so on. A general criteria to stop a golden section search might be to stop whenever the difference between two successive bracketing is not greater than 10^{-4} for single precision and 3×10^{-8} for double precision computation (Press, Teukolsky, Vetterling and Flannery, 2002). The Convergence rate of a golden section search algorithm is linear and linear improvement in bracketing requires linearly increasing function evaluation.

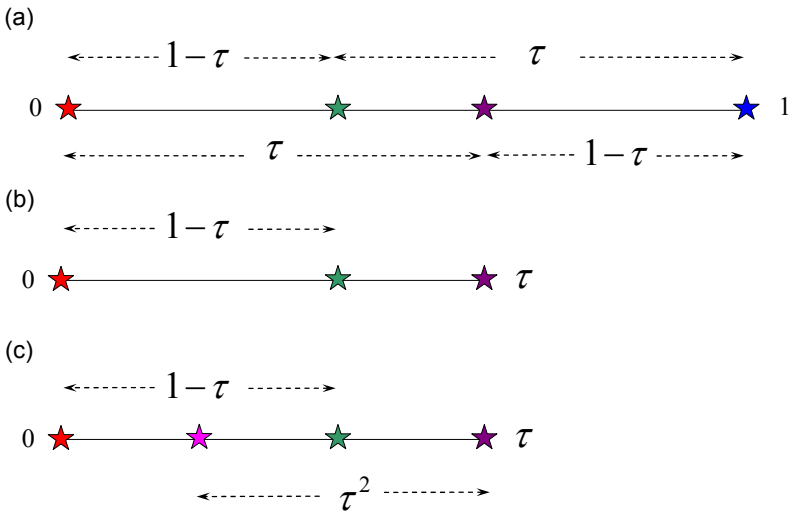


Figure 4.3: Golden ratio search

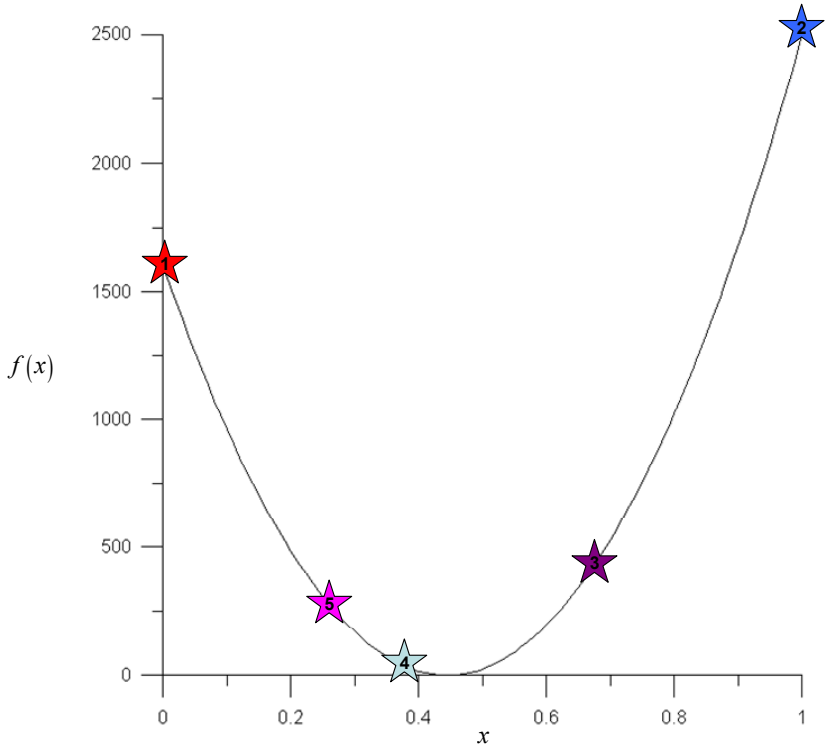


Figure 4.4: Golden section search bracketing $f(x)=(40-90x)^2$

4.4.2 Inexact line search methods

In previous section we mentioned the purpose of the line search. We are looking for a global value of step length α to minimize an objective function f at the minimum cost.

$$\varphi(\alpha) = f(x_k + \alpha h_k) \ . \tag{4.44}$$

The direct search methods such as golden section search, however, are robust to deal even with non-quadratic objective function but require a lot of function evaluations.

In linear problems, where (4.44) is a quadratic one-dimensional minimizer of α along the $x_k + \alpha h_k$, taking the directional derivative of f with respect to α and equating it to zero, results in the following expression (Nocedal and Wright, 1999).

$$\alpha_k = -\frac{\nabla f_k^T h_k}{h_k^T A h_k} \ . \tag{4.45}$$

where h is a search direction and H can be the Hessian matrix for nonlinear problem or could simply be the coefficients of a system of linear equations in linear problems. Nonlinear objective functions need an iterative, either inexact or exact line search, to estimate α. In the following, we briefly review some of the most popular inexact line search methods based on Armijo and Wolfe conditions. Wolfe conditions are the most popular inexact line search method which is used in unconstrained minimization algorithms specifically in the variable metric methods where an exact line search minimizer is not required.

Armijo is a popular inexact line search method where it looks for a step length α that sufficiently decreases the objective function f by satisfying the following inequality:

$$f(x_k + \alpha h_k) \le f(x_k) + c_1 \alpha \nabla f_k^T h_k \ . \tag{4.46}$$

The right side of inequality (4.46) is a linear function $l(\alpha)$ of step length. Inequality (4.46) is usually known as the Armijo condition or rule. $0 < c_1 < 1$, h_k is a descent direction, and $\nabla f_k^T h_k$ is the directional derivative. In practice, $c_1 = 10^{-4}$. Figure 4.5 illustrates the line of sufficient decrease; every value of step length which lies below the objective function $\varphi(\alpha)$ can be a candidate for the line search minimization. As can be

seen from Figure 4.5, there are locations that satisfy Armijo condition, although they are quite far from a minimizer. Armijo condition is not sufficient and other conditions need to be imposed. Curvature of the objective function along with the line of sufficient decrease, collectively, can be used to restrict step length according to:

$$\begin{aligned} f\left(x_k+\alpha h_k\right) &\le f\left(x_k\right)+c_1\alpha\nabla f_k^T h_k \\ \nabla f\left(x_k+\alpha h_k\right)^T h_k &\ge c_2\nabla f_k^T h_k \end{aligned} \quad . \tag{4.47}$$

The inequalities in (4.47) are collectively called the Wolfe conditions. The left hand side of the second inequality of Wolfe conditions is the derivative of the objective function with respect to each step length. $c_1 < c_2 < 1$. Wolfe conditions simply state that the slope of the objective function has to be equal to or greater than a reduced slope of the objective function at a location where the step length is zero. Figure 4.5 shows the areas where the step length satisfies the Wolfe conditions. Wolfe conditions have restricted the areas where step length could be a candidate for line search minimization. The value of c_2 varies, based upon the minimization algorithms which produced the search direction vector h. For the Newton and variable metric algorithms it should be 0.1 while, for nonlinear conjugate gradient it should be 0.9 (Nocedal and Wright, 1999). As it can been seen in Figure 4.5, there are areas where, the curvature is strongly positive and the related step length is quite far from a local minimizer , however it satisfies the Wolfe conditions. By imposing some constraints on curvatures, the step length can be forced closer to the critical point of the objective function.

$$\begin{aligned} f\left(x_k+\alpha h_k\right) &\le f\left(x_k\right)+c_1\alpha\nabla f_k^T h_k \\ \left|\nabla f\left(x_k+\alpha h_k\right)^T h_k\right| &\le c_2\left|\nabla f_k^T h_k\right| \end{aligned} \quad . \tag{4.48}$$

The inequalities in (4.48) are called strong Wolfe conditions. Practical implementation of methods satisfying Wolfe conditions need to find an initial guess for the step length and also its upper and lower limits. This is usually treated as a two step process. In the first step α has to be bracketed to a lower and upper limit by any method such as parabolic interpolation, golden section search, bisection etc. In the second step after finding a suitable range, α is iteratively updated using an interpolation scheme. This includes updating the upper and lower bounds while satisfying the Wolfe conditions.

A practical implementation of line search utilizing Wolfe conditions is given by (Moré and Thuente, 1994; Nocedal and Wright, 1999).

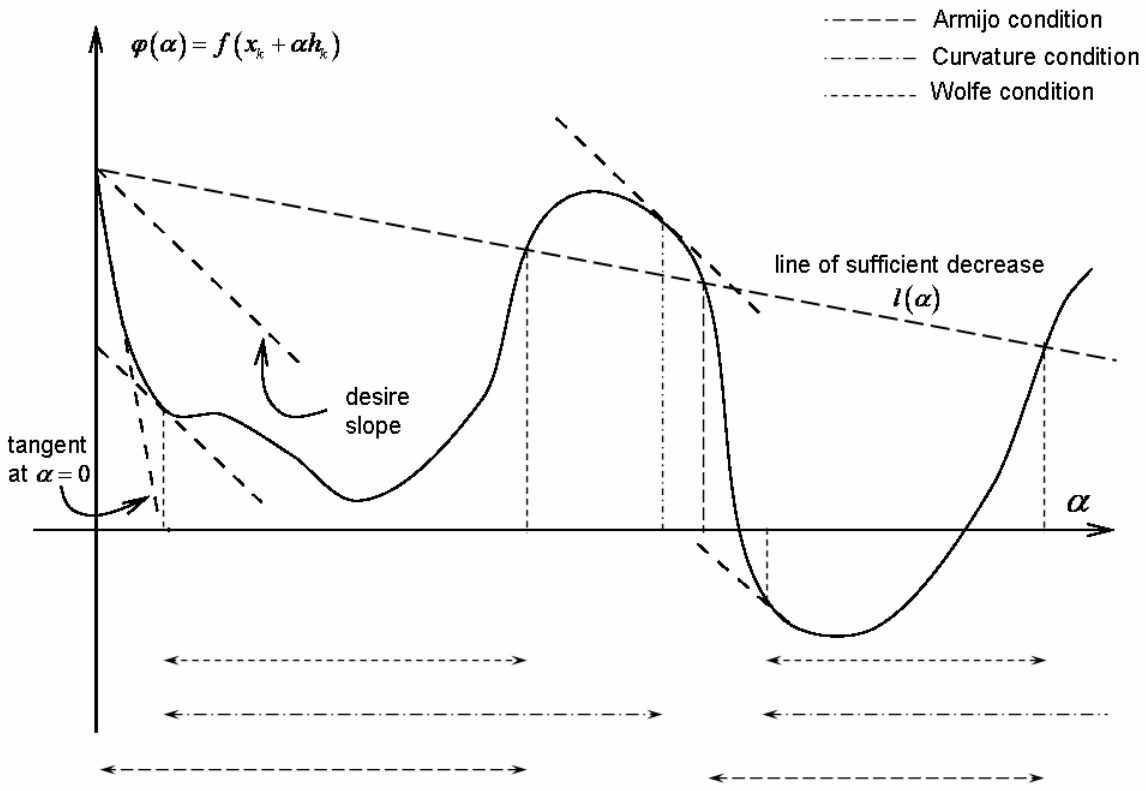


Figure 4.5: Line of sufficient decrease and Wolfe conditions for the objective function (Nocedal and Wright (1999) after some modifications).

4.4.3 Exact line search methods

In situations where the multidimensional minimization algorithms such as nonlinear conjugate gradient need a step length very close to the minimum of (4.44), inexact line search methods may not be very efficient and instead an exact line search minimizer should be considered. Exact line search algorithms need extra computation and line search methods with linear convergence rate such as Golden section search are not desirable. Brent's method (Brent, 1973) is a popular but complicated one-dimensional minimization method with superlinear convergence rate. It only uses the function

evaluation and don't utilise first derivatives which would impose extra function evaluation. There are exact line search methods where, first derivatives are utilized. One of the early works is assigned to William C. Davidon in 1959, who introduced the variable metric method (Davidon, 1991). Brent's method combines successive inverse parabolic interpolation and Golden section search method (whenever, the function is not cooperating) to find an approximate minimum of the objective function.

Lagrange interpolation for any triplet points a, b, and c, where c is located between a and b, can be expressed in the following form.

$$f(x)=f(a)\frac{(x-b)(x-c)}{(a-b)(a-c)}+f(b)\frac{(x-c)(x-a)}{(b-c)(b-a)}+f(c)\frac{(x-a)(x-b)}{(c-a)(c-b)}\ . \qquad (4.49)$$

After equating the first derivative of (4.49) to zero and some algebraic manipulation the extremum can be expressed as:

$$x=b-\frac{1}{2}\frac{(b-a)^2\left[f(b)-f(c)\right]-(b-c)^2\left[f(b)-f(a)\right]}{(b-a)\left[f(b)-f(c)\right]-(b-c)\left[f(b)-f(a)\right]}\ . \qquad (4.50)$$

We express $e=x-b$. At each step there are six points *a, b, u, v, w,* and *x* which are not all distinct and their position changes during the algorithm. Assuming there is a minimum in the bracketed range of $[a,b]$, apart from u, the rest of the variables are initialized using the Golden ratio:

$$v=w=x=a+(0.5)\left(3-\sqrt{5}\right)(b-a),\quad (0.5)\left(3-\sqrt{5}\right)=0.381966... \qquad (4.51)$$

As the process starts, *x* will be calculated using (4.50) as the minimum of the fitted parabola to the objective function; *w* is the point with next least value of the objective function; *v* is the previous value of *w*; and *u* is the last point where, the objective function is evaluated. In the next iteration, if the denominator of (4.50) is equal to zero, if $x\notin(a,b)$, $|e|\le tol$, where e is calculated in the previous iteration, or $|x-b|\ge\frac{1}{2}|e|$, then the process alternates to the Golden section search method. Let $m=\frac{a+b}{2}$ be the mid point of (a,b). The next value of *u* is:

$$u=\begin{Bmatrix}0.618x+0.381966a & x\ge m\\ 0.618x+0.381966b & x<m\end{Bmatrix} \qquad (4.52)$$

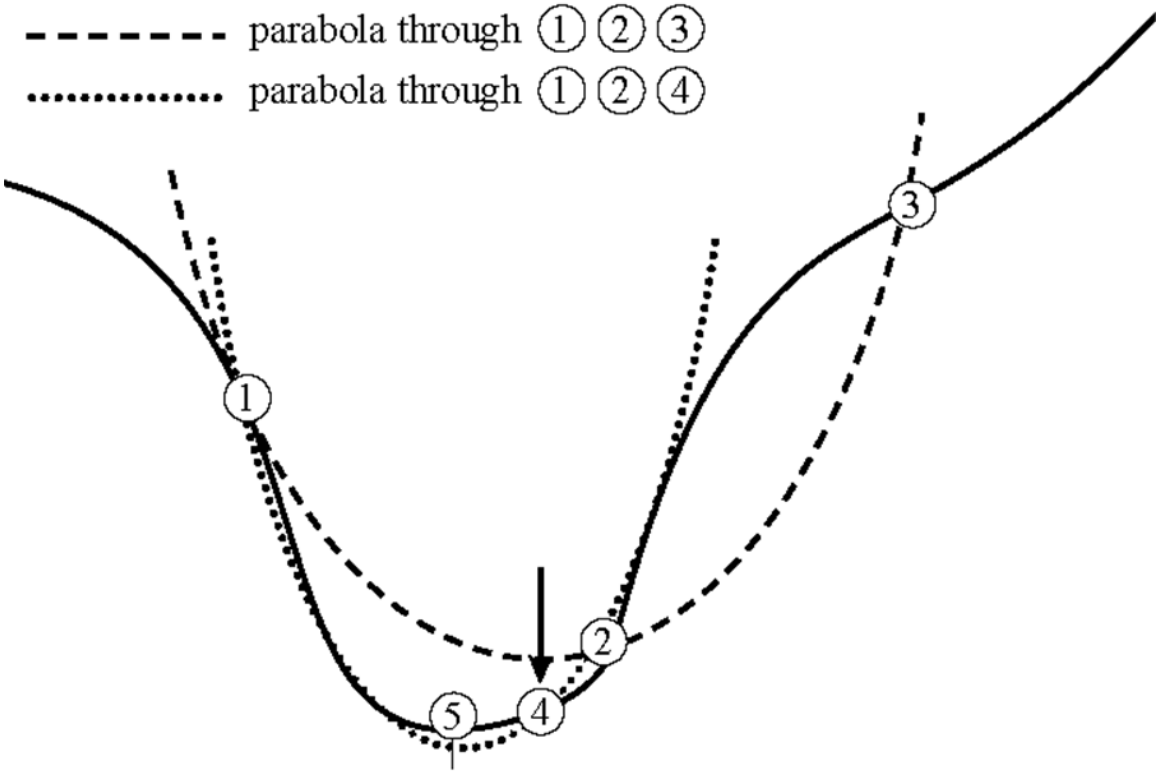


Figure 4.6: Successive fitting a parabola with the objective function. Points 1 and 3 has already been bracketed, while point 2 initially has been located using the golden ratio. The first parabola is fitted in the point 1, 2, and 3. The minimum of the first parabola (4) is considered as the next point along with points 1 and 2 for another parabolic fitting. The minimum of the second parabola (5) is close to minimum of the objective function (After Press, et al. (2002)).

If the Golden section is used or

$$u = x + e \ , \tag{4.53}$$

if the inverse parabolic interpolation method is used. A typical termination of Brent's method will be if the interval of (a,b) is less than $2 \times x \times tol$, where the tol should not be less than the square root of floating point machine precision. In this thesis, we use Brent's method based on the algorithm given by (Press, Teukolsky, Vetterling and Flannery, 2002) whenever nonlinear conjugate gradient is used.

4.4.4 The Method of steepest descent

Let us approximate a nonlinear function f with the Taylor series of the second order in vicinity of the point $x_0 = m_0$ as a quadratic form.

$$f(x) = \frac{1}{2} x^T A x - b^T x + c \quad . \tag{4.54}$$

where $c = f(m)$, $b = -\nabla f\big|_{x_0}$, and $A_{ij} = \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{x_0}$ is a symmetric matrix. Differentiating (4.54) with respect to x results in:

$$\begin{aligned} f'(x) &= \frac{1}{2} A^T x + \frac{1}{2} A x - b \\ &= Ax - b \end{aligned} \quad . \tag{4.55}$$

If A is positive definite then (4.56) will be the solution of $f(x)$ as:

$$Ax = b \tag{4.56}$$

A matrix A is called positive definite if for any vector x, $x^T A x > 0$. Linear problems can also be expressed as a quadratic form for minimization; however, A_{ij}, which are the coefficients of a set of linear equations, may not be symmetric or even positive definite. If A is not positive definite, then

$$\frac{1}{2}\left(A^T + A\right) x = b \quad , \tag{4.57}$$

could be a solution of (4.54). $\left(A^T + A\right)$ is now symmetric.

In the method of steepest descent we start at an initial point x_i and move in the direction of steepest descent:

$$\begin{aligned} r_i &= -f'(x_i) \\ &= b - A x_i \end{aligned} \quad . \tag{4.58}$$

The amount of jump is $\alpha_i r_i$, where α is the step length. We take the next step perpendicular to the previous direction. The model parameter x will be iteratively updated according to:

$$x_{i+1} = x_i + \alpha_i r_i \quad . \tag{4.59}$$

The updating scheme (4.59), in a probabilistic point of view for a steepest descent algorithm, is given by (Tarantola, 2005) as:

$$x_{i+1} = x_i + \alpha_i C_M \nabla_f^T C_D^{-1} e_i \ , \tag{4.60}$$

where C_M and C_D are covariance matrices of model parameters and data respectively. $e = y - d$ is the residual vector. In this thesis wherever we use any unconstrained minimization updating scheme we use (4.60) rather than generic equation (4.59).

Assuming the function f is quadratic, The step length α can be found by taking the directional derivative of $f\left(x_{i+1}\right)$ with respect to α and the orthogonality of successive steepest descent directions (Shewchuk, 1994) , $r_{i+1}^T r_i = 0$ as:

$$\alpha_i = \frac{r_i^T r_i}{r_i^T H r_i} \ . \tag{4.61}$$

For nonlinear functions where the assumption of quadratic behavior is not well satisfied, the step length α should be found by any of the inexact or exact line search methods. The method of steepest descent will become very slow if it goes along a narrow valley and it takes a lot of small steps to reach the solution even if the objective function is a perfect quadratic function. Figure 4.7 shows the convergence of a linear set of equations of two parameters. Steepest descent converges to the minimum by more steps than the number of parameters.

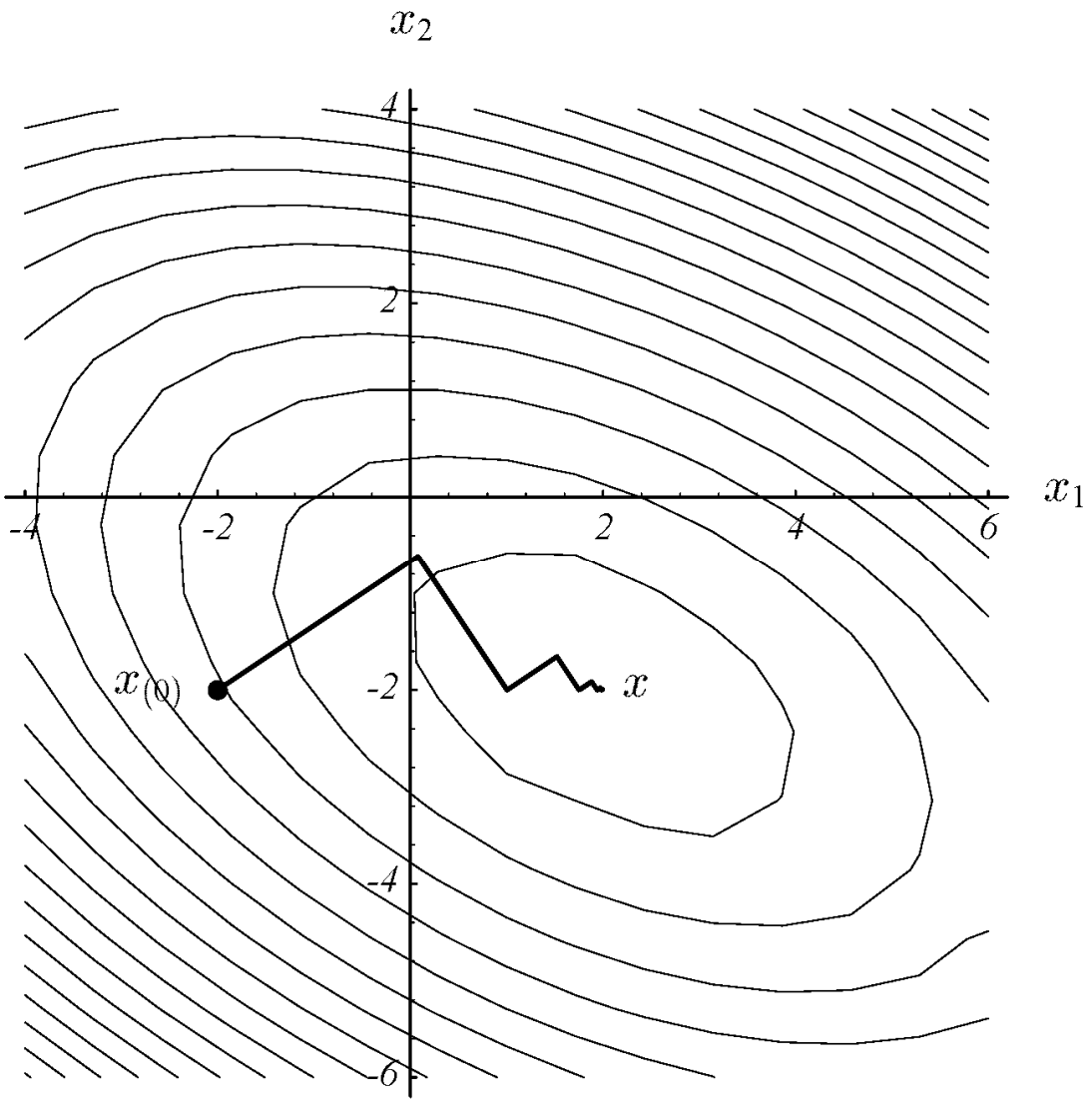


Figure 4.7: Minimization of a two parameter linear equation using the method of steepest descent. $A = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$, $b = \begin{bmatrix} 2 \\ -8 \end{bmatrix}$, and $c = 0$. It takes several small steps to reach the solution (Shewchuk, 1994).

4.4.5 Conjugate gradient algorithm

The conjugate gradient (CG) algorithm was proposed by Stiefel and Hestenes in 1950 separately (Hestenes and Stiefel, 1952) for linear problems but it remained without attention for about 20 years. Conjugate gradient has been considered as one of top 10 algorithms with greatest influence on development and practices of science and engineering in 20th century (Sullivan and Dongara, 2000). It has been particularly important for large scale minimization problems. The first conjugate gradient algorithm for nonlinear problems has been proposed by Fletcher and Reeves (1964). Preconditioning of conjugate gradient algorithms either linear or nonlinear has led to the fast convergence of this algorithm (Van der Vorst, 2000).

One of the remarkable properties of conjugate gradient is to produce *n* conjugate directions along which the objective function can be successively minimized, in *n* steps. A set of non-zero vectors h are said to be conjugate if for a positive definite matrix A :

$$h_i^T A h_j = 0, \quad i \neq j \ . \tag{4.62}$$

The question is, how to get the conjugate directions? There are many ways to do so. The eigenvectors of matrix A are mutually orthonormal and also conjugate to A but this approach is not suitable for large scale problem where it will be computationally expensive to calculate the eigenvectors. A-Orthogonalization (Conjugate Gram-Schmidt) process (Hestenes and Stiefel, 1952; Shewchuk, 1994) as well as conventional or modified Gram-Schmidt (Golub and Van Loan, 1996) processes can produce orthonormal vectors with less computation. The *A* term in *A*-Orthogonalization refers to a positive definite matrix *A*. Using a set of *n* linear independent vectors u_i, *n* conjugate vectors h_i can be constructed using the conjugate Gram-Schmidt orthogonalization as:

$$h_i = u_i + \sum_{k=0}^{i-1} \beta_{ik} h_k \ , \tag{4.63}$$

$$\beta_{ik} = -\frac{u_i^T A h_k}{h_k^T A h_k}, \quad i > k \ . \tag{4.64}$$

where $h_0 = u_0$, β_{ik} are the projection of u over h, and u_i could be a set of linear independent random vectors and h_i are conjugate directions. This process subtracts out

any vector which is not conjugate to the previous h vectors. One disadvantage of Gram-Schmidt orthogonalization is the need to store all the previous conjugate directions.

In the method of conjugate direction the solution vector x_{i+1} can be updated as the sum of previous vectors and a jump in the direction of conjugate vectors as:

$$x_{i+1} = x_i + \alpha_i h_i \ . \tag{4.65}$$

The step length α can be found by minimizing the objective function f along the search direction. Equating the first derivative of $f(x_{i+1})$ with respect to α to zero, results is (Nocedal and Wright, 1999):

$$\alpha_i = -\frac{r_i^T h_i}{h_i^T A h_i} \ . \tag{4.66}$$

where r_i is the previous residual vector. The residual vector r_{i+1} can be computed using a combination of previous residual vector r_i and the new search direction h_{i+1} , but after few iterations it will suffer from floating point roundoff error and should be refreshed by the x vector.

$$r_{i+1} = r_i + \alpha_i A h_i \ . \tag{4.67}$$

The method of conjugate gradient is the same as the method of conjugate directions except the linear independent vector u is replaced by the residual direction r. Equation (4.67) shows that each residual vector is a linear combination of the previous residual vector and Ah_{i+1}. Vector $h_{i+1} \in D_{i+2}$, where D_i is an i-dimensional subspace span $\{h_0, h_1, ..., h_{i-1}\}$. This implies that subspace D_{i+2} is formed from the union of previous subspace D_{i+1} and AD_{i+1} (Shewchuk, 1994).

$$\begin{aligned} D_i &= span\{d_0, Ad_0, A^2 d_0, ..., A^{i-1} d_0\} \\ &= span\{r_0, \ Ar_0, \ A^2 r_0, ..., \ A^{i-1} r_0\} \end{aligned} \ . \tag{4.68}$$

Subspace (4.68) is called Krylov subspace. It is a subspace which is spanned by a Krylov matrix D_i. Because AD_{i+1} is in the subspace D_{i+2} and residual vector r_{i+1} is orthogonal to D_{i+2}, hence r_{i+1} is conjugate to D_{i+1}. This fact results in a nice property in the conjugate Gram-Schmidt orthogonalization where r_{i+1} is conjugate to the previous search

directions and all projection operations in (4.63) and (4.64) will vanish except for h_{i+1}. The new scalars α and β can be expressed as (Kelley, 1995; Nocedal and Wright, 1999):

$$\alpha_i = \frac{r_i^T r_i}{h_i^T A h_i} \ , \tag{4.69}$$

$$\beta_{i+1} = \frac{r_{i+1}^T r_{i+1}}{r_i^T r_i} \ . \tag{4.70}$$

The search vector h_{i+1} can now be updated using the linear combination of current residual vector r_{i+1} and $\beta_{i+1} h_i$ as:

$$h_{i+1} = -r_{i+1} + \beta_{i+1} h_i \ . \tag{4.71}$$

where r_{i+1} can be computed using (4.67).

The linear conjugate gradient algorithm can now be summarized as (Nocedal and Wright, 1999):

Given a prior model $x_0 = m_0$ and convergence tolerance ε>0,
Set $r_0 \leftarrow Ax_0 - b,\ h_0 \leftarrow -r_0,\ i \leftarrow 0;$
while $\|r_i\| > \varepsilon$
$\alpha_i = \dfrac{r_i^T r_i}{h_i^T A h_i};$
$x_{i+1} = x_i + \alpha_i h_i;$
$r_{i+1} = r_i + \alpha_i A h_i;$
$\beta_{i+1} = \dfrac{r_{i+1}^T r_{i+1}}{r_i^T r_i};$
$h_{i+1} = -r_{i+1} + \beta_{i+1} h_i;$
$i = i + 1;$
end

. (4.72)

Due to floating point roundoff errors in matrix-vector product, particularly when the matrix A is ill-conditioned, the conjugate vectors will lose their $A-Orthogonality$ and the following inner products are not strictly satisfied in practice:

$$\langle r_i, r_j \rangle = 0,\ \langle h_i, r_j \rangle = 0,\ \langle h_i, A h_j \rangle = 0 \ . \tag{4.73}$$

To ensure that (4.73) equations are strictly satisfied, search directions have to be re-orthogonalized against all the previous directions. This process is called complete re-

orthogonalization and conjugate gradient algorithm will be very slow. Partial re-orthogonalization of current search direction against a number of previous descent vectors has been shown effective to satisfy conjugacy (Roux, 1989). Equation (4.71) should be expressed as:

$$h_{i+1} = -r_{i+1} + \sum_{j=0}^{k} \beta_{i-j} h_{i-j} \quad , \tag{4.74}$$

$$\beta_{i-j} = -\frac{r_{i+1}^T A h_{i-j}}{h_{i-j}^T A h_{i-j}} \quad . \tag{4.75}$$

We have so far considered the conjugate gradient method to minimize linear functions. One of the remarkable aspects of conjugate gradient is, its extension to nonlinear functions. Fletcher and Reeves (1964) by two modifications in (4.72) showed conjugate gradient can be extended to nonlinear functions. Replacing the residual vector r by the negative of the derivative of the nonlinear objective function (gradient vector) $-f'(x)$ and computing the step length α using a line search method rather than equation (4.69). They proposed a new scalar β which does not need to evaluate the matrix vector product (4.75). This algorithm with the following scalar is known as the Fletcher-Reeves (Fletcher and Reeves, 1964) algorithm,

$$\beta_{i+1}^{FR} = \frac{r_{i+1}^T r_{i+1}}{r_i^T r_i} \quad , \tag{4.76}$$

where $r_i = -f'(x_i)$, $r_{i+1} = -f'(x_{i+1})$. There are many other expressions for the scalar β and there is continuing research for new variants of nonlinear conjugate gradient. Perhaps, the Polak-Ribiére method (Polak and Ribiére, 1969) (4.77) by far, is the most favorable variant of nonlinear conjugate gradient . From the other most important variants of nonlinear conjugate gradient, we can recall the methods of Hestenes-Stiefel , Fletcher (abbreviated as CD) (Fletcher, 1980), Dai and Yuan (Dai and Yuan, 1999) and are given in equations (4.78) - (4.80) respectively.

$$\beta_{i+1}^{PR} = \frac{r_{i+1}^T (r_{i+1} - r_i)}{r_i^T r_i} \quad , \tag{4.77}$$

$$\beta_{i+1}^{HS} = \frac{r_{i+1}^T \left(r_{i+1} - r_i \right)}{\left(r_{i+1}^T - r_i \right)^T h_i} \ , \tag{4.78}$$

$$\beta_{i+1}^{CD} = -\frac{r_{i+1}^T r_{i+1}}{h_i^T r_i} \ , \tag{4.79}$$

$$\beta_{i+1}^{DY} = \frac{r_{i+1}^T r_{i+1}}{h_i^T \left(r_{i+1} - r_i \right)} \ . \tag{4.80}$$

A generic nonlinear conjugate gradient algorithm is summarized as (Nocedal and Wright, 1999; Bonnans et al., 2003):

Given a prior model $x_0 = m_0$ and convergence tolerance ε>0,
Evaluate $f_0 = f\left(x_0\right), \nabla f_0 = \nabla f\left(x_0\right)$;
Set $r_0 = h_0 \leftarrow -\nabla f_0, \ \ i \leftarrow 0$;
while $\left\| r_i \right\| > \varepsilon$
  compute α_i (Either exact or inexact line search);
  $x_{i+1} = x_i + \alpha_i h_i$;
  Evaluate $r_{i+1} = -\nabla f_{i+1}\left(x_{i+1}\right)$;
  $\beta_{i+1} = \frac{r_{i+1}^T r_{i+1}}{r_i^T r_i}$;
  if $\left(\beta_{i+1} < 0\right)$
    $h_{i+1} = r_{i+1}$
  else
    $h_{i+1} = r_{i+1} + \beta_{i+1} h_i$;
  $i = i + 1$;
end . (4.81)

Shi and Shen (2006) proposed the following recursive step length for unconstrained minimization problems. The first step length α_0 can be calculated for instance, by Armijo rule. This may reduce the cost of line minimization computation.

$$\alpha_{i+1} = -\frac{\alpha_i r_{i+1}^T h_{i+1}}{\left(r_{i+1} - r_i \right)^T h_i} \ . \tag{4.82}$$

In nonlinear problems it is common to have a negative β, especially for the method of Polak and Ribiére where, the strong Wolfe conditions do not guarantee that h_i is always a descent direction (Nocedal and Wright, 1999). In this case β should be set to zero.

To asses the effectiveness of the different nonlinear conjugate gradient algorithms we have applied methods (4.76)-(4.80) to a stack of anisotropic layers (Table 2.1) to estimate the Thomson anisotropy parameters (δ, ε) from traveltime data. Figure 4.8 illustrate the objective function evolution for the above algorithms. All use the same prior model and use the exact line minimization. 96 traces have been used where the first trace was located at 200 metre from the source and the rest of the traces were kept at 50 metre intervals. It can be seen that only Polak-Ribiére method has converged reasonably but the method of Dai-Yuan and Fletcher-Reeves have failed. The method of Hestenes-Stiefel has a step-wise like convergence pattern and it might converge if we continued to more iterations. Fletcher method has a monotonic decreasing behavior and tends to converge as the iteration goes further.

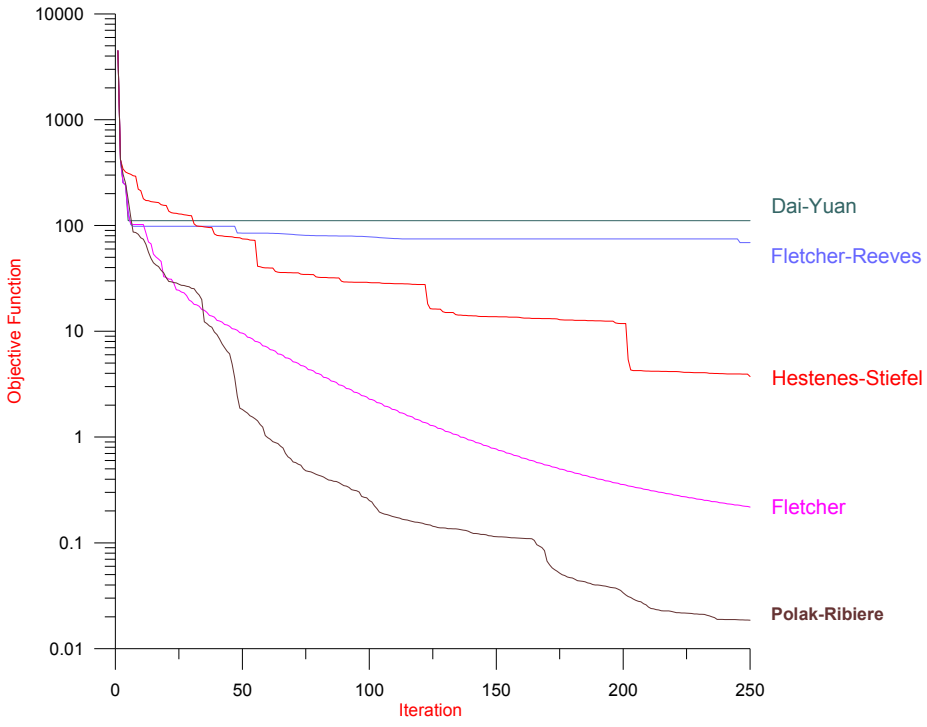


Figure 4.8: Convergence pattern of different nonlinear conjugate gradient methods. Polak-Ribiére method seems superior to other algorithms.

The conjugate directions ensure the convergence of conjugate gradient in *n* steps, where *n* is the number of conjugate directions. Due to floating point roundoff error, conjugacy is not always observed and conjugate gradient may converge in more than n iterations. Lack of conjugacy of search directions is even more serious in nonlinear than in linear problems. Figure 4.8 clearly shows these issues where there are 18 parameters to be estimated. It takes more than 250 iterations to converge to a reasonable error level. In large scale problem or where the forward problem is computationally intensive, continuing the process even by *n* iteration is not economic. If the condition number of the matrix A is high or the eigenvalues are not clustered, the convergence rate will be slow. Preconditioning is a way to reduce the condition number of a matrix and increase the robustness and efficiency of iterative algorithms. Preconditioning is a means to transform the original linear system to another one which has the same solution but is likely to be easier to solve by an iterative solver.

Suppose A is a symmetric positive definite matrix and there exist a preconditioner matrix M which approximates A in each iteration. Matrix M is also symmetric and positive definite and easier to invert. Then, the system of linear equations (4.56) can be solved as:

$$M^{-1}Ax = M^{-1}b \ . \tag{4.83}$$

If the condition number of matrix $M^{-1}A$ is much smaller than the condition number of A or if the eigenvalues of $M^{-1}A$ are more clustered, then (4.83) can be solved iteratively. However, matrix $M^{-1}A$ is no longer symmetric and positive definite. One can use incomplete Choleski factorization to decompose matrix M to a lower triangular and its transposed matrices, to preserve the symmetry. However, we can use *M*-inner product $\langle x, y\rangle_M$ of two vectors in conjugate gradient algorithm instead of their Euclidean inner products $\langle Mx, y\rangle$ (Saad, 2000). Since $M^{-1}A$ is self-adjoint for the *M*-inner product,

$$\langle x, y\rangle_M \equiv \langle Mx, y\rangle = \langle x, My\rangle \ , \tag{4.84}$$

where

$$\langle M^{-1}Ax, y\rangle_M = \langle Ax, y\rangle = \langle x, Ay\rangle = \langle x, M\left(M^{-1}A\right)y\rangle = \langle x, M^{-1}Ay\rangle_M \ . \tag{4.85}$$

The new residual vector for preconditioned system can be written as:

$$z_i = M^{-1}r_i \ . \tag{4.86}$$

where the old residual r is defined in (4.58). Now, the following expressions need to be considered for preconditioned conjugate gradient algorithm.

$$\alpha_i = \frac{\langle z_i, z_i \rangle_M}{\langle M^{-1} A h_i, h_i \rangle_M} , \tag{4.87}$$

$$\beta_i = \frac{\langle z_{i+1}, z_{i+1} \rangle_M}{\langle z_i, z_i \rangle_M} , \tag{4.88}$$

where $z_{i+1} = M^{-1} r_{i+1}$. Since $\langle z_i, z_i \rangle_M = \langle r_i, z_i \rangle$ and $\langle M^{-1} A h_i, h_i \rangle_M = \langle A h_i, h_i \rangle$, the M-inner products do not need to be computed explicitly. The Following is a sequence for preconditioned linear conjugate gradient algorithm given by Saad (2000).

Given a prior model $x_0 = m_0$ and convergence tolerance ε>0,
$i = 0$
Compute $r_0 = b - A x_0, z_0 = M^{-1} r_0,\ h_0 = z_0;$
While $\|r_i\| > \varepsilon$
$\alpha_i = \dfrac{r_i^T z_i}{h_i^T A h_i};$
$x_{i+1} = x_i + \alpha_i h_i;$
$r_{i+1} = r_i - \alpha_i A h_i;$
$z_{i+1} = M^{-1} r_{i+1};$
$\beta_{i+1} = \dfrac{r_{i+1}^T z_{i+1}}{r_i^T z_i};$
$h_{i+1} = z_{i+1} + \beta_{i+1} h_i;$
$i = i + 1;$
end

. (4.89)

The simplest and easily implemented preconditioner is the diagonal preconditioning or Jacobian preconditioning where, the inverse of the diagonal elements of matrix A is used as preconditioner (Kelley, 1995). Sparse Cholesky factorization of matrix A can also be used as a more elaborated preconditioner which is called incomplete Cholesky preconditioner (Nocedal and Wright, 1999). To see more comprehensive preconditioners for linear conjugate gradient problems the reader may refer to Saad (2000).

Nonlinear problems can also be preconditioned; however there is not much about nonlinear preconditioners in the literature. The Hessian matrix could be an effective preconditioner to approximate M if it is positive definite (Shewchuk, 1994). In areas of model space where x_i is quite far from the quadratic region, there is no guarantee that the Hessian matrix is positive definite (Press, Teukolsky, Vetterling and Flannery, 2002) and unfortunately there is not much that we can do. In this case we have to force the Hessian matrix to be positive definite, simply by changing to a unit matrix. If the computation of Hessian matrix is expensive we may only use the diagonal elements and setting the off-diagonal elements to zero. By computing the first derivatives at two locations of model space, we can calculate an approximation to the inverse of the Hessian matrix using the BFGS method numerically. This means we have to compute the inverse of the Hessian matrix in every iteration. In the next section we will review the BFGS method. The sequence (4.89) with few changes can also be implemented for nonlinear problems. The residuals have to be considered as the sequence (4.81) and step length α should also be computed using a line search. β should also be computed according to (4.81). One of the disadvantages of preconditioning is that the model parameters and their derivatives must have almost the same magnitude otherwise they have to be rescaled.

We have implemented a preconditioned nonlinear conjugate gradient on the same example as we did in nonlinear conjugate gradient without preconditioning (Figure 4.8). we used BFGS method to approximate the inverse of the Hessian matrix as a preconditioner with the Polak-Ribiére method. As can be seen in Figure 4.9 there is an outstanding improvement in convergence rate of the objective function. The method of Polak-Ribiére, without preconditioning, has also been shown for comparison. The preconditioning conjugate gradient after 25 iterations has reached the same error level that the conjugate gradient without preconditioning has reached at 250 iterations.

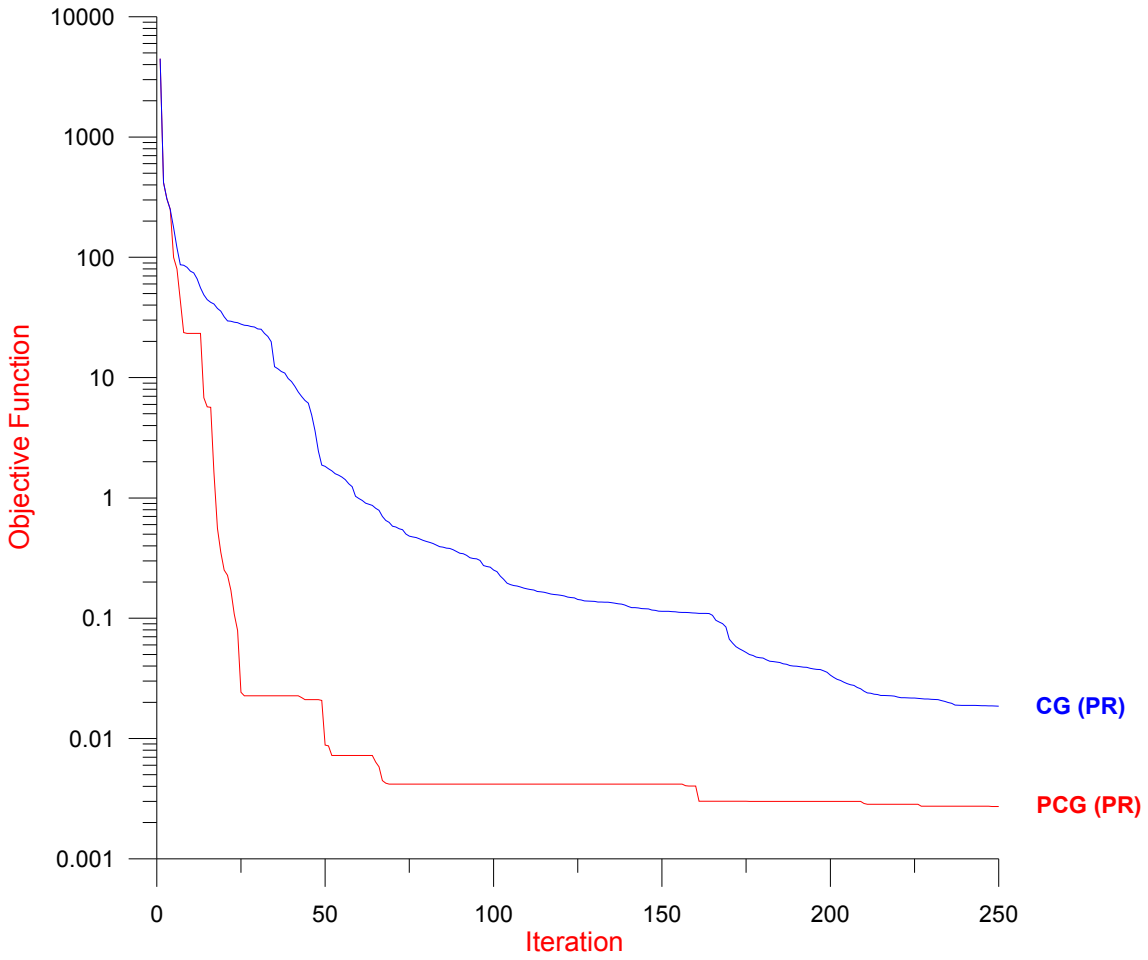


Figure 4.9: Convergence pattern of preconditioned nonlinear conjugate gradient (PCG) using the method of Polak-Ribiére (PR) (red line). The convergence for the conjugate gradient without preconditioning (CG) (blue line) has also been shown for comparison.

4.4.6 Variable metric (quasi-Newton) methods

In mid 1950's C. W. Davidon, a physicist from Argonne National Laboratory, while he was frustrated by slow convergence of the steepest descent method, proposed a new unconstrained minimization algorithm which turned out to be one of the revolutionary ideas in optimization. His algorithm, named variable metric, improved the convergence rate and efficiency of the numerical minimization method remarkably. "*An interesting historical irony is that Davidon's paper (Davidon, 1959) was not accepted for publication; it remained as a technical report for more than thirty years until it appeared in the first issue of the SIAM Journal on Optimization in 1991 (Davidon, 1991)"* (Nocedal and Wright, 1999). The variable metric algorithm was further studied, implemented and popularized by Fletcher and Powell. DFP and BFGS are two variants of variable metric. The first one is called after works of Davidon, Fletcher and Powell and the later is named due to works of Broyden, Fletcher, Goldfarb and Shanno. Variable metric essentially uses the updating scheme, like the Newton method, which implements the matrix of second derivatives, but instead it approximates the Hessian matrix implicitly. It does not calculate the Hessian matrix at every iteration explicitly, but instead it updates the Hessian matrix from the previous and current search directions and gradient information. Variable metric methods need extra storage of the order of n^2 for the Hessian matrix. With the advent of high performance computers where memory is not a big issue, variable metric can be used efficiently for large scale problems. It is more popular than conjugate gradient but more likely to fail if we are too far from the quadratic region and where the assumption of positive definite of the Hessian matrix is violated.

Let us approximate the objective function f by its Taylor expansion in the vicinity of an initial model parameter x_i as:

$$f(x) = f(x_i) + (x - x_i)\nabla f(x_i) + \frac{1}{2}(x - x_i)\, A(x - x_i) \ , \tag{4.90}$$

where A is the matrix of second derivatives. Taking the derivative of (4.90) we get:

$$\nabla f(x) = \nabla f(x_i) + A(x - x_i). \tag{4.91}$$

The minimum of (4.90) can be computed by taking $\nabla f(x) = 0$, so

$$h_i = x - x_i = -A^{-1}\nabla f\left(x_i\right) \ . \tag{4.92}$$

This is the search direction h_i as in Newton method (2.30) and the model will be updated according to:

$$x_{i+1} = x_i + \alpha_i h_i \ , \tag{4.93}$$

where the step length α_i is chosen to satisfy the Wolfe conditions. The only difference from Newton method is that A is used in place of the true Hessian. This is why variable metric is also called as quasi-Newton method. By evaluating (4.90) at the new location x_{i+1} and calculating the derivative as (4.91) at the minimum we get:

$$x - x_{i+1} = -A^{-1}\nabla f\left(x_{i+1}\right) \ . \tag{4.94}$$

Now by subtracting the (4.94) and (4.92) we have:

$$x_{i+1} - x_i = A^{-1}\left(\nabla f_{i+1} - \nabla f_i\right), \tag{4.95}$$

where $\nabla f_i \equiv \nabla f\left(x_i\right)$. Lets $y_i = \nabla f_{i+1} - \nabla f_i$ and $s_i = x_{i+1} - x_i$, now (4.95) can be written as:

$$y_i = A s_i \ . \tag{4.96}$$

This is the secant equation. Multiplying both side of (4.96) by s_i^T we have curvature condition:

$$s_i^T y_i = s_i^T A s_i > 0 \ . \tag{4.97}$$

When f is strongly convex, the inequality (4.97) is satisfied for any two points x_i and x_{i+1}. However, if f is not strongly convex, inequality (4.97) will not be satisfied and some restriction has to be imposed on α in the line search. If Wolfe conditions or strong Wolfe conditions are imposed during the line search the condition (4.97) will be satisfied (Nocedal and Wright, 1999).

By moving from location x_i to x_{i+1} , we will be looking for a new matrix H_{i+1} as if it was actually A^{-1}. We assume that among all symmetric matrices satisfying the secant equation, H_{i+1} is, the one which is closest to the current matrix A^{-1}. We might also think that $H_{i+1} = H_i + Correction$. The amount of correction is given by either DFP or BFGS methods as:

DFP:

$$H_{i+1} = H_i - \frac{H_i y_i y_i^T H_i}{y_i^T H_i y_i} + \frac{s_i s_i^T}{y_i^T y_i} \quad . \tag{4.98}$$

BFGS:

$$H_{i+1} = \left(I - \frac{s_i y_i^T}{y_i^T s_i} \right) H_i \left(I - \frac{y_i s_i^T}{y_i^T s_i} \right) + \frac{s_i s_i^T}{y_i^T s_i} \tag{4.99}$$

DFP is quite effective but very soon, it was replaced by BFGS which is presently considered to be the most effective updating formula among all other variable metric variants. There is no unique solution for the initial inverse of the Hessian matrix, one can calculate the Hessian matrix, for instance, by finite differences at x_0 and compute its inverse. Otherwise we can simply use an identity matrix. The following is a sequence for implementing variable metric with BFGS (Nocedal and Wright, 1999).

Given a prior model $x_0 = m_0$ and convergence tolerance ε>0,
    inverse Hessian approximation H_0
$i \leftarrow 0$;
while $\|\nabla f_i\| > \varepsilon$
    Compute search direction $h_i = -H_i \nabla f_i$;
    Compute step length α by line search
    (Satsify Wolfe or strong Wolfe condition) (4.100)
    $x_{i+1} = x_i + \alpha_i h_i$;
    $s_i = x_{i+1} - x_i$;
    $y_i = \nabla f_{i+1} - \nabla f_i$;
    Compute H_{i+1} by (4.97)
    $i = i + 1$;
end

To asses the performance of variable metric over conjugate gradient methods we estimated the Thomson's anisotropy parameters as in the previous example with the method of BFGS. Variable metric was converging quickly but has been trapped in a plateau (Figure 4.10).

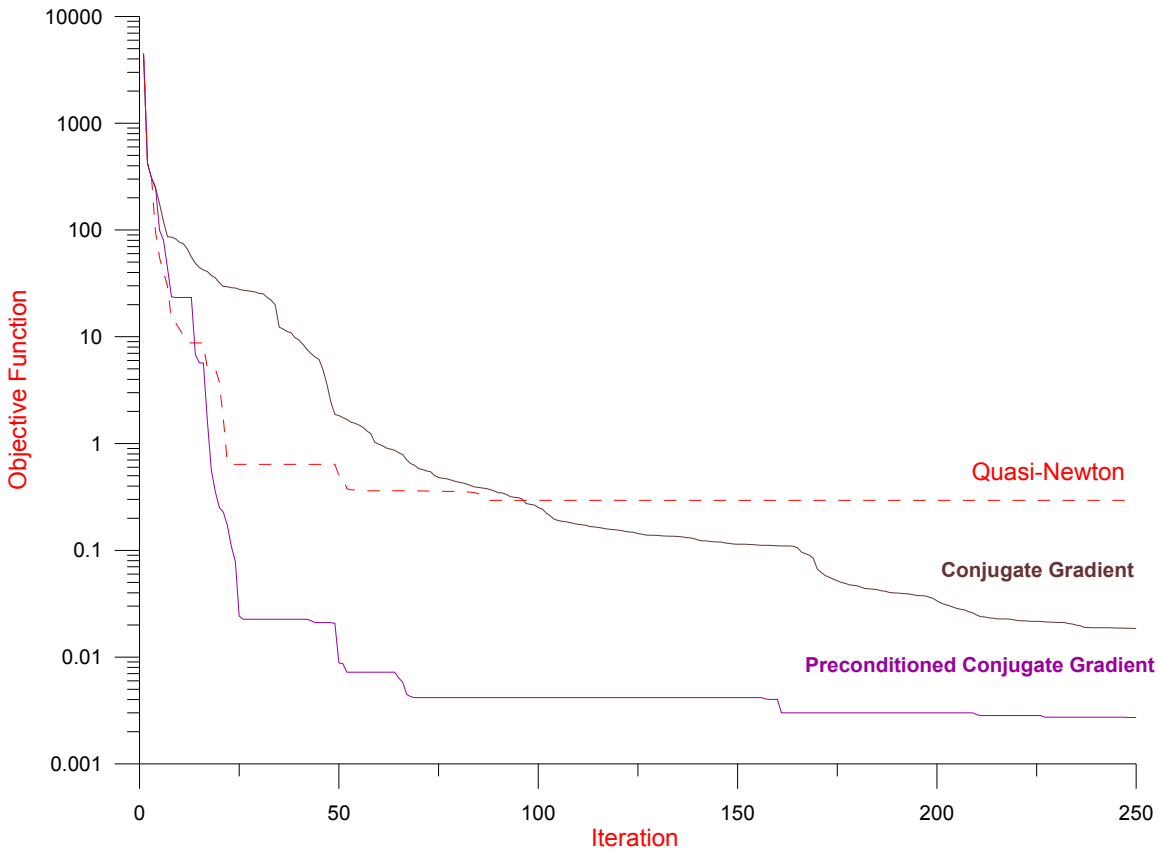


Figure 4.10: Convergence pattern of variable metric using the BFGS method. Preconditioned nonlinear conjugate gradient and conjugate gradient without preconditioning, both using the method of Polak-Ribiére have also been shown for comparison. Variable metric has converged in initial iterations but it has been trapped in a plateau later. In this case one may restart variable metric algorithm by changing the prior model.

References

Aster, R. C., B. Borchers, and C. H. Thurber, 2005, Parameter estimation and inverse problems: Elsevier Academic Press.

Berryman, J. G., 1994, Resolution of iterative inverses in seismic tomography, *in* J. Brown, M. Chu, D. Ellison and R. Plemmons, eds., Proc. Cornelius Lanczos International Centenary Conference: SIAM, 297-299.

Bonnans, J. F., J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, 2003, Numerical optimization: Theoretical and practical aspects: Springer-Verlag.

Brent, R. P., 1973, Algorithms for minimization without derivatives: Prentice-Hall.

Dai, Y. H., and Y. Yuan, 1999, A nonlinear conjugate gradient with a strong Wolfe convergence property: SIAM Journal on Optimization, **10**, 177-182.

Davidon, W. C., 1959, Variable metric method for minimization, Technical Report ANL–5990 (revised): Argonne National Laboratory.

Davidon, W. C., 1991, Variable metric method for minimization: SIAM Journal on Optimization, **1**, 1-17.

Fletcher, R., 1980, Practical methods of optimization, Vol 1, Unconstrained optimization: John Wiley & Sons.

Fletcher, R., and C. M. Reeves, 1964, Function minimization by conjugate gradients: the Computer Journal, **7**, 149-154.

Golub, G. H., and C. F. Van Loan, 1996, Matrix computations: Johns Hopkins University Press.

Hestenes, M. R., and E. Stiefel, 1952, Methods of Conjugate Gradients for Solving Linear Systems: Journal of Research of the National Bureau of Standards, **49**, 409-436.

Jackson, D. D., 1972, Interpretation of inaccurate, insufficient and inconsistent data: Geophys. J. R. astr. Soc, **28**, 97-109.

Kelley, C. T., 1995, Iterative Methods for Linear and Nonlinear Equations: Society for Industrial and Applied Mathematics (SIAM).

Lines, L. R., and S. Treitel, 1984, Tutorial: A review of least-squares inversion and its application to geophysical problems: Geophysical Prospecting, **32**, 159-186.

Marquardt, D. W., 1963, An algorithm for least-squares estimation of nonlinear parameters: Journal of the Society for Industrial and Applied Mathematics, **11**, 431-441.

Menke, W., 1989, Geophysical data analysis: Discrete inverse theory: Academic Press, Inc.

Moré, J. J., and D. J. Thuente, 1994, Line search algorithms with guaranteed sufficient decrease: ACM Transactions on Mathematical Software **20**, 286-307.

Nocedal, J., and S. J. Wright, 1999, Numerical optimization: Springer-Verlag.

Oldenburg, D. W., and Y. Li, 2005, Inversion for applied geophysics: A tutorial, *in* D. K. Butler, ed., Near-Surface Geophysics: Society of Exploration Geophysicists, 89-150.

Polak, E., and G. Ribiére, 1969, Note sur la Convergence de Methodes de Directions Conjugees: Revue Francaise d'lnformatique et de Recherche Operationnelle, **3**, 35-43.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 2002, Numerical recipes in C++: the art of scientific computing: Cambridge University Press.

Ravindran, R., K. M. Ragsdell, and G. V. Reklaitis, 2006, Engineering optimization: Methods and applications: John Wiley & Sons.

Roux, F. X., 1989, Acceleration of the outer conjugate gradient by reorthogonalization for a domain decomposition method for structural analysis problems: Presented at the Proceedings of the 3rd international conference on Supercomputing Association for Computing Machinery.

Saad, Y., 2000, Iterative Methods for Sparse Linear Systems.

Scales, J. A., and R. Snieder, 2000, The anatomy of inverse problems: Geophysics, **65**, 1708-1710.

Scales, J. A., and L. Tenorio, 2001, Prior information and uncertainty in inverse problems: Geophysics, **66**, 389-397.

Sen, M., and P. L. Stofa, 1995, Global optimization methods in geophysical inversion: Elsevier Science B. V.

Shewchuk, J. R., 1994, An introduction to the conjugate gradient method without agonizing pain.

Shi, Z.-J., and J. Shen, 2006, On step-size estimation of line search methods: Applied Mathematics and Computation, **173**, 360-371.

Sullivan, F., and J. S. Dongara, 2000, Guest Editors Introduction to the top 10 algorithms: Computing in Science & Engineering, **2**, 22-23.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: SIAM.

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

Van der Vorst, H. A., 2000, Krylov subspace iteration: Computing in Science & Engineering, **2**, 32-37.

Chapter 5

Unconstrained inverse models for traveltime and amplitude

Introduction

In inverse problems, where the forward model is computationally intensive, it is important to obtain the solution within minimum computation time. Robustness and efficiency of unconstrained minimization algorithms have attracted the numerical analysts to optimize such computationally intensive problems. Unconstrained minimization algorithms often need gradient information to find a direction in order to converge quickly. Unfortunately, it is not always easy to extract analytic partial derivatives. Instead, numerical methods such as finite differences can often be considered as an alternative. Finite differences have a few drawbacks; they are computationally intensive and their accuracy is very much dependent on the interval of differentiation. This feature which is called instability is a serious issue especially when the function is highly nonlinear. Analytic partial derivatives are exact, fast but often lengthy and cumbersome to calculate if possible at all. In this chapter I extract the partial derivatives of traveltimes, reflection and transmission coefficients and verify them with numerically computed ones for a system of transversely isotropic layers. I provide unconstrained inverse models for prestack $P \to P$ and joint $P \to P \,\&\, P \to S_v$ traveltimes in both horizontal and dipping layers. I also formulate an unconstrained minimization algorithm for waveform inversion of both prestack $P \to P$ and joint $P \to P \,\&\, P \to S_v$ seismic amplitudes. This approach is superior to layer stripping method in which traveltimes from lower layers make no contribution to the model parameters in upper layer.

5.1 Partial derivatives of traveltime in a layered TI media

We follow the Dahl and Ursin (1991) approach to compute the partial derivatives of traveltime with respect to model parameters in a layered, either horizontally layered or dipping media. Since we express the traveltimes as a function of ray parameter, the derivative of traveltimes with respect to any model parameter has to consider the derivative of ray parameter with respect to any model parameter too. This will be accomplished by taking the derivative of offset with respect to each model parameters as:

$$\frac{dx}{dm_j} = \left.\frac{\partial x}{\partial m_j}\right|_{p=const.} + \frac{\partial x}{\partial p}\frac{\partial p}{\partial m_j} \quad , \tag{5.1}$$

where m_j is the model parameter for a layer j, x is source-receiver offset , and p is the ray parameter. Since the offset is given for a given ray, the derivative of the offset with respect to any model parameter is zero, $\partial x/\partial m_j = 0$. These result in derivatives of ray parameter with respect to a given model parameter as:

$$\frac{\partial p}{\partial m_j} = -\frac{\left.\dfrac{\partial x}{\partial m_j}\right|_{p=const.}}{\dfrac{\partial x}{\partial p}} \quad , \tag{5.2}$$

where $\partial x/\partial p$ and $\left.\partial x/\partial m_j\right|_{p=const.}$ are given in chapter 2 and appendix A respectively.

Given the equation (5.2), the derivative of traveltime with respect to each model parameter can be expressed as the sum of the derivative of traveltime with respect to a model parameter for a given layer at a constant ray parameter and a derivative term including the effect of change in ray parameter for a ray path as the consequence of change in a model parameter as:

$$\frac{dT}{dm_j} = \left.\frac{\partial T}{\partial m_j}\right|_{p=const.} + \frac{\partial T}{\partial p}\frac{\partial p}{\partial m_j} \quad , \tag{5.3}$$

where $\partial T/\partial p$ and $\left.\partial T/\partial m_j\right|_{p=Const.}$ are given in appendix A.

5.2 Partial derivatives of reflection and transmission coefficients in a TI horizontal layered media

The computation of partial derivative of reflection and transmission (R/T) coefficients with respect to model parameters is more complicated but we still follow the same approach as for the traveltimes. Since at the interface of two elastic media, R/T coefficients are functions of elastic properties of both upper and lower media, any changes in one of these media will change the R/T, hence the derivative. Suppose a plane wave travelling across a system of layered TI media (Figure 5.1), the effective reflection coefficient can be expressed as:

$$\ell = \prod_{i=1}^{2M-3} s_i \quad M \geq 2$$

$$\begin{cases} s_i = t_i & i = 1, M-2 \quad \text{down-going transmission coefficients} \\ s_i = r & i = M-1 \quad \text{reflection coefficient} \\ s_i = t_i' & i = M-2, 1 \quad \text{up-going transmission coefficients} \end{cases} \tag{5.4}$$

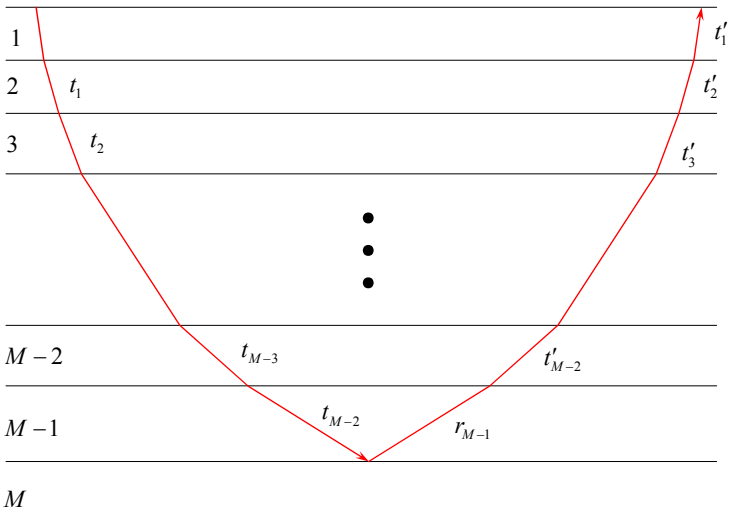


Figure 5.1: Reflection and transmission coefficients of a plane wave across a M layer TI media.

The derivative of reflection coefficient r with respect to a model parameter m_j at the interface of two layers j and $j+1$ for a down going plane wave can be written as:

$$\frac{dr}{dm_j} = \left.\frac{\partial r}{\partial m_j}\right|_{p=const.} + \frac{\partial r}{\partial p}\frac{\partial p}{\partial m_j}, \tag{5.5}$$

$$\frac{dr}{dm_{j+1}} = \left.\frac{\partial r}{\partial m_{j+1}}\right|_{p=const.}, \tag{5.6}$$

where r could be $P \to P$ or $P \to S_v$ reflection coefficient. The partial derivative of reflection coefficient with respect to ray parameter in (5.6) vanishes because the layer $j+1$ has no effect on the ray parameter of the ray reflected from the top of this layer. Equations (5.5) and (5.6) are only valid for model parameters Vp, Vs, δ, and ε. These equations have to be modified for density ρ and layers thickness z according to:

$$\frac{dr}{d\rho_j} = \left.\frac{\partial r}{\partial \rho_j}\right|_{p=const.}, \tag{5.7}$$

$$\frac{dr}{d\rho_{j+1}} = \left.\frac{\partial r}{\partial \rho_{j+1}}\right|_{p=const.}, \tag{5.8}$$

$$\frac{dr}{dz_j} = \frac{\partial r}{\partial p}\frac{\partial p}{\partial z_j}, \tag{5.9}$$

$$\frac{dr}{dz_{j+1}} = 0. \tag{5.10}$$

Since $\partial p/\partial \rho = 0$, (5.7) should only be evaluated in layer j at a constant ray parameter. The reflection coefficient at the interface of two media is not a function of layer thickness, however, the reflection coefficient for a given offset will vary as the thickness varies. This is due to the fact that a change in thickness will affect the ray parameter and hence, the angles of incidence and refraction.

The derivative of the transmission coefficients t with respect to model parameter m_j and m_{j+1} at the interface of two elastic medium j and $j+1$ for a down going compressional plane wave can be expressed as:

$$\frac{dt}{dm_j} = \left.\frac{\partial t}{\partial m_j}\right|_{p=const.} + \frac{\partial t}{\partial p}\frac{\partial p}{\partial m_j}, \tag{5.11}$$

$$\frac{dt}{dm_{j+1}} = \left.\frac{\partial t}{\partial m_{j+1}}\right|_{p=const.} + \frac{\partial t}{\partial p}\frac{\partial p}{\partial m_{j+1}} \quad , \tag{5.12}$$

where model parameter m could be Vp, δ, or ε. Equations (5.11) and (5.12) can be easily extended to the up going compressional plane waves by considering layer $j+1$ at the top of layer j. Since the transmission coefficient is not a function of the vertical shear wave velocity in Rüger's equations for a down going or up going compressional plane wave, (5.11) and (5.12) have to be rewritten as:

$$\frac{dt}{dVs_j} = \frac{\partial t}{\partial p}\frac{\partial p}{\partial Vs_j} \quad , \tag{5.13}$$

$$\frac{dt}{dVs_{j+1}} = \frac{\partial t}{\partial p}\frac{\partial p}{\partial Vs_{j+1}} \quad . \tag{5.14}$$

The above derivatives are in fact numerically very small, because the ray parameter for compressional wave field is a very weak function of shear wave velocity in a transverse isotropic medium.

The partial derivative of transmission coefficient for a down-going or up-going compressional plane wave with respect to density and layer thickness can be written as:

$$\frac{dt}{d\rho_j} = \left.\frac{\partial t}{\partial \rho_j}\right|_{p=const.} , \tag{5.15}$$

$$\frac{dt}{d\rho_{j+1}} = \left.\frac{\partial t}{\partial \rho_{j+1}}\right|_{p=const.} , \tag{5.16}$$

$$\frac{dt}{dz_j} = \frac{\partial t}{\partial p}\frac{\partial p}{\partial z_j} \quad , \tag{5.17}$$

$$\frac{dt}{dz_{j+1}} = \frac{\partial t}{\partial p}\frac{\partial p}{\partial z_{j+1}} \quad . \tag{5.18}$$

For the case of up going waves in equations (5.13)-(5.14), we need to assume that layer $j+1$ is laid on the top of layer j.

Equations (5.5)-(5.10) can be also used for the converted shear wave reflection coefficient in Rüger's equations. To compute the partial derivatives of up going shear

wave transmission coefficients, we can also use equations (5.11) , (5.12) and (5.15)-(5.18). Shear waves transmission coefficients for up going waves can be written as:

$$\frac{dt}{dVs_j} = \left.\frac{\partial t}{\partial Vs_j}\right|_{p=const.} + \frac{\partial t}{\partial p}\frac{\partial p}{\partial Vs_j}, \tag{5.19}$$

$$\frac{dt}{dVs_{j+1}} = \left.\frac{\partial t}{\partial Vs_{j+1}}\right|_{p=const.} + \frac{\partial t}{\partial p}\frac{\partial p}{\partial Vs_{j+1}}. \tag{5.20}$$

The partial derivative of the effective reflection coefficient (5.4) with respect to model parameter m_k can be expressed as:

$$\frac{\partial \ell}{\partial m_k} = \ell \sum_{i=1}^{2M-3} \frac{1}{s_i}\frac{\partial s_i}{\partial m_k}. \tag{5.21}$$

In a stack of layers, where a ray is reflecting from the top of the deep layers (more than three layers down), only the maximum 4 terms of the (5.21) have to be evaluated and the rest are zero. To have a clearer understanding of (5.21), we demonstrate the above statement by an example of a system of 5 layers, where a ray is supposedly reflected from the top of the last layer. The partial derivatives of effective reflection coefficients with respect to model parameter m_j, in each layer, only consist of the following non-zero terms:

$$\begin{aligned}
&\ell = t_1 t_2 t_3 r_4 t_3' t_2' t_1' \\
&j=1: \quad \frac{\partial t_1}{\partial m_1}, \frac{\partial t_1'}{\partial m_1}; \quad \frac{\partial \ell}{\partial m_1} = \left(\frac{1}{t_1}\frac{\partial t_1}{\partial m_1} + \frac{1}{t_1'}\frac{\partial t_1'}{\partial m_1}\right)\ell \\
&j=2: \quad \frac{\partial t_1}{\partial m_2}, \frac{\partial t_2}{\partial m_2}, \frac{\partial t_2'}{\partial m_2}, \frac{\partial t_1'}{\partial m_2}; \quad \frac{\partial \ell}{\partial m_2} = \left(\frac{1}{t_1}\frac{\partial t_1}{\partial m_2} + \frac{1}{t_2}\frac{\partial t_2}{\partial m_2} + \frac{1}{t_2'}\frac{\partial t_2'}{\partial m_2} + \frac{1}{t_1'}\frac{\partial t_1'}{\partial m_2}\right)\ell \\
&j=3: \quad \frac{\partial t_2}{\partial m_3}, \frac{\partial t_3}{\partial m_3}, \frac{\partial t_3'}{\partial m_3}, \frac{\partial t_2'}{\partial m_3}; \quad \frac{\partial \ell}{\partial m_3} = \left(\frac{1}{t_2}\frac{\partial t_2}{\partial m_3} + \frac{1}{t_3}\frac{\partial t_3}{\partial m_3} + \frac{1}{t_3'}\frac{\partial t_3'}{\partial m_3} + \frac{1}{t_2'}\frac{\partial t_2'}{\partial m_3}\right)\ell \\
&j=4: \quad \frac{\partial t_3}{\partial m_4}, \frac{\partial r_4}{\partial m_4}, \frac{\partial t_3'}{\partial m_4}; \quad \frac{\partial \ell}{\partial m_4} = \left(\frac{1}{t_3}\frac{\partial t_3}{\partial m_4} + \frac{1}{r_4}\frac{\partial r_4}{\partial m_4} + \frac{1}{t_3'}\frac{\partial t_3'}{\partial m_4}\right)\ell \\
&j=5: \quad \frac{\partial r_4}{\partial m_5}; \quad \frac{\partial \ell}{\partial m_4} = \frac{1}{r_4}\frac{\partial r_4}{\partial m_5}\ell
\end{aligned} \tag{5.22}$$

where j is the layer number.

Reflection and transmission coefficients in Rüger's equations are a function of phase angle of incidence and refraction but the ray tracing is carrying out in terms of ray parameter. Using equation (2.14) and the definition of ray parameter we have,

$$\sin(\theta) = \frac{p\alpha_0}{\left(1 - p^2\alpha_0^2 H_p\right)^{\frac{1}{2}}}\ , \tag{5.23}$$

for P-waves and

$$\sin(\theta) = \frac{p\beta_0}{\left(1 - p^2\beta_0^2 H_s\right)^{\frac{1}{2}}}\ , \tag{5.24}$$

for SV-waves. Other trigonometric relations can easily be calculated.

The derivatives of Rüger's equations for reflection and transmission coefficients for different modes, $P \to P$, $P \to S_v$, $S_v \to S_v$, and $S_v \to P$ are cumbersome and are not shown here. We have stored all the derivatives in Maple format on a compact disk. Readers may refer to the compact disk attached to this thesis.

5.3 Unconstrained inverse model for traveltime inversion in a multi layered media

As in the previous section, we assume a transversely isotropic horizontal or dipping layered medium (Figures 2.1 and 2.9), where the elastic properties and geometry of each layer are laterally invariant. In the case of dipping layers we assume that the layers are not truncated.

A quadratic objective function, without imposing a specific structure given by Tarantola (2005), can be used to minimize the residual error as:

$$S(m) = \frac{1}{2}\left(d_{obs} - d_{syn}\right)^T C_D^{-1}\left(d_{obs} - d_{syn}\right) + \frac{1}{2}\left(m - m_{prior}\right)^T C_M^{-1}\left(m - m_{prior}\right)\ . \tag{5.25}$$

Furthermore, in most examples in this thesis we ignore the second term of (5.25) and simply minimize the residual in data space. This is sufficient when we have a parameter estimation problem (Oldenburg and Li, 2005) and the system is overdetermined. Using an iterative scheme we update the prior model vector m according to Tarantola (2005):

$$m_{i+1} = m_i - \alpha_i C_M \nabla_S^T C_D^{-1} e_i\ . \tag{5.26}$$

In practice, we assume that data and model parameters are independent, hence, the correlation between two different elements is zero. So, the covariance matrices of model C_M and data C_D are diagonal and contain the variances. For now, we take the inverse of the data covariance matrix out and look at the product of gradient ∇_S^T and data residual vector $e = d_{obs} - d_{syn}$ as:

$$\frac{\partial S}{\partial m_k} = \nabla_{S,k}^T e = -\sum_{i=1}^{N}\sum_{j=1}^{M-1}\frac{\partial T_{ij}}{\partial m_k}e_{ij} \ , \tag{5.27}$$

where the indices i , j, and k repeat over the number of offsets N, the number of layers $M-1$, and the model parameters respectively. In a matrix notation equation (5.27) can be expressed as:

$$\Delta m = J^T g \ , \tag{5.28}$$

where J^T , g , and Δm are given in (4.39), (4.40), and (4.41) respectively. Here, g is in fact the vector representation of matrix e and Δm refers to vector $\frac{\partial S}{\partial m_k}$. The objective function S in an indicial notation can be shown in the following form:

$$S = \frac{1}{2}\sum_{i=1}^{N_T}\sum_{j=1}^{M-1}\left(T_{ij}^{obs} - T_{ij}^{syn}\right)\left[C_D^{ii}\right]^{-1}\left(T_{ij}^{obs} - T_{ij}^{syn}\right), \tag{5.29}$$

where N_T is the number of traces, *obs* stands for measured traveltimes, and *syn* for computed traveltimes. We assume there is no correlation between the data, hence, the off-diagonal elements are zero. The diagonal elements or the variances could simply be the errors in picking the traveltimes.

5.4 Unconstrained inverse model for waveform inversion in a multi layered media

In chapter three we computed the synthetic seismogram as the product of a source wavelet w and a reflectivity series l, in the frequency domain as:

$$U(f) = w(f)\ell(f) \ . \tag{5.30}$$

The partial derivative of seismic amplitude U with respect to each model parameter m_k at each frequency f can be written as:

$$\frac{\partial U(f)}{\partial m_k} = w(f)\frac{\partial \ell(f)}{\partial m_k} + \ell(f)\frac{\partial w(f)}{\partial m_k} \ . \tag{5.31}$$

$\partial w(f)/\partial m_k = 0$, hence (5.31) can be simplified as:

$$\frac{\partial U(f)}{\partial m_k} = w(f)\frac{\partial \ell(f)}{\partial m_k} \ . \tag{5.32}$$

Here, $\partial \ell(f)/\partial m_k$ is computed using a DFT algorithm as:

$$\ell(f) = \sum_{i=1}^{M-1} \ell_i e^{j2\pi f \tau_i} \ , \tag{5.33}$$

where ℓ_i is the effective reflection coefficient (including transmission effects for the whole ray path) using Rüger's equations (Rüger, 1996) at each interface for a specific wave mode. τ_i is the two way traveltime for a specific event $(P \to P \text{ or } P \to S_v)$ reflecting from interface i. M is the number of layers. Taking derivative of (5.33) with respect to model parameter m_k gives:

$$\begin{aligned}
\frac{\partial \ell(f)}{\partial m_k} &= \frac{\partial}{\partial m_k}\left(\sum_{i=1}^{M-1} \ell_i e^{j2\pi f \tau_i}\right) \\
&= \sum_{i=1}^{M-1}\left(\frac{\partial}{\partial m_k}\left(\ell_i e^{j2\pi f \tau_i}\right)\right) \\
&= \sum_{i=1}^{M-1}\left(e^{j2\pi f \tau_i}\frac{\partial \ell_i}{\partial m_k} + \ell_i \frac{\partial}{\partial m_k}\left(e^{j2\pi f \tau_i}\right)\right) \\
&= \sum_{i=1}^{M-1}\left(e^{j2\pi f \tau_i}\left(\frac{\partial \ell_i}{\partial m_k} + j2\pi f \ell_i \frac{\partial \tau_i}{\partial m_k}\right)\right) .
\end{aligned} \tag{5.34}$$

Using Euler's formula,

$$e^{j2\pi f \tau_i} = \cos(2\pi f \tau_i) + j\sin(2\pi f \tau_i) \ , \tag{5.35}$$

equation (5.34) can be further extended to:

$$\begin{aligned}
\mathrm{Re}\left[\frac{\partial \ell(f)}{\partial m_k}\right] &= \sum_{i=1}^{M-1}\left(\frac{\partial \ell_i}{\partial m_k}\cos(2\pi f \tau_i) - 2\pi f \ell_i \frac{\partial \tau_i}{\partial m_k}\sin(2\pi f \tau_i)\right) \\
\mathrm{Im}\left[\frac{\partial \ell(f)}{\partial m_k}\right] &= \sum_{i=1}^{M-1}\left(\frac{\partial \ell_i}{\partial m_k}\sin(2\pi f \tau_i) + 2\pi f \ell_i \frac{\partial \tau_i}{\partial m_k}\cos(2\pi f \tau_i)\right) .
\end{aligned} \tag{5.36}$$

So, (5.32) can be written as:

$$\frac{\partial U(f)}{\partial m_k} = \mathrm{Re}\left[w(f)\right]\mathrm{Re}\left[\frac{\partial \ell(f)}{\partial m_k}\right] - \mathrm{Im}\left[w(f)\right]\mathrm{Im}\left[\frac{\partial \ell(f)}{\partial m_k}\right] + j\left(\mathrm{Re}\left[w(f)\right]\mathrm{Im}\left[\frac{\partial \ell(f)}{\partial m_k}\right] + \mathrm{Im}\left[w(f)\right]\mathrm{Re}\left[\frac{\partial \ell(f)}{\partial m_k}\right]\right). \tag{5.37}$$

The derivative of the objective function (5.25) with respect to model parameters can be expressed as:

$$\frac{\partial S}{\partial m_k} = \nabla_{S,k}^{T} e = -\sum_{i=1}^{N}\sum_{j=1}^{n}\frac{\partial U_{ij}}{\partial m_k} e_{ij} \ , \tag{5.38}$$

where n is the number of frequency samples and e_{ij} is the difference between the measured seismic traces and synthetic seismograms in the frequency domain. Here, the index notation of S, ignoring the model space term could be of the form,

$$S = \frac{1}{2}\sum_{i=1}^{N_T}\sum_{j=1}^{N_S}\left(U_{ij}^{obs} - U_{ij}^{syn}\right)^{\dagger}\left[C_D^{ii}\right]^{-1}\left(U_{ij}^{obs} - U_{ij}^{syn}\right) , \tag{5.39}$$

where N_T is the number of traces and N_S is number of frequency samples. The seismic amplitudes U are complex and † is the conjugate complex operator. The diagonal elements of the covariance matrix are real valued and could be the signal to noise ratio. We assume that seismic amplitudes are stationary and uncorrelated, so, off-diagonal elements are zero. The frequency spectrum can be extended up to the Nyquist frequency, but both computation of S and $\partial S/\partial m_k$ will be intensive. It makes sense to limit the frequency range to the spectrum bandwidth.

To test whether the analytic derivatives are correctly derived, we have computed (5.38) for a system of horizontal layers in a prior model. Computation of numerical derivatives is straightforward, however, much care has to be taken. We found the numerical derivatives with respect to Thomsen's anisotropy parameters are highly unstable and a proper interval at which the derivatives are to be taken is very crucial. An improper interval could easily lead to inaccurate derivatives. In a small range of the function where the assumption of linearity is valid, we evaluate the objective function in 1000 successive points. During the evaluation of the objective function, we only change one parameter and the rest of the other model parameters are unchanged. Then, using a

linear regression we calculated the derivative. This will compensate the roundoff error which may cause inaccurate computation of the objective function. As a general rule of thumb, 0.001 percent of parameter can be an optimal interval for R/T coefficients and 0.1 percent for traveltime derivatives respectively. By computing (5.38), the derivatives of traveltimes with respect to the model parameters are also computed.

Table 5.1 shows the prior model for a system of transversely isotropic layers used for computation of the derivatives of compressional wavefield. We have used 48 traces and sampled the compressional wavefield by 2048 time samples every 2 milliseconds. Nyquist frequency is 250 Hz. Table 5.2 demonstrates the analytic derivatives (5.38) while Table 5.3 illustrates the numerically computed partial derivatives with respect to model parameters. There is a good agreement between the two tables. In Table 5.4 we have only used one quarter of frequency samples which covers the most useful part of the spectrum. This reduces the computation of the objective function and its derivatives considerably. A comparison between tables 5.2 and 5.4 shows a good agreement.

Layer	$Vp\left(m/\sec\right)$	$Vs\left(m/\sec\right)$	δ	ε	$z\left(metre\right)$	$\rho\left(gr/cm^3\right)$
1	1697.5	877.5	0.02625	0.1125	875	2.28375
2	1872.5	752.5	-0.01717	0.1575	180	1.8025
3	2497.5	1001.25	-0.42916	0.0875	78.75	1.82
4	1750	875	0.0875	0.1575	35	2.295
5	2238.75	1113.75	0.04375	0.1125	112.5	2.295
6	2137.5	831.25	0.045	0.135	166.25	2.2725
7	2475	1006.25	0.0675	0.1575	236.25	2.34
8	1793.75	988.75	0.0875	0.225	148.75	2.30625
9	2318.75	1312.5	0.07875	0.1125	348.75	2.41875
10	3093.75	1338.75	0.1125	0.1575	262.5	1.89875
11	2970	1676.25	0.035	0.07	123.75	2.41875

Table 5.1: the prior model has been used to compute the analytic and numerically derived partial derivatives.

Layer	Vp	Vs	δ	ε	z	ρ
1	-6.26511	-0.09949	-6241.51	-22572.8	-24.6499	-277.906
2	-7.60709	0.646668	-1251.33	-17631.5	-28.4609	-1273.21
3	18.548	-1.33439	5194.3	30107.4	-16.9072	2633.49
4	-2.74135	0.801856	-1229.82	-1449.9	-18.1381	-1365.76
5	1.66819	-0.3268	710.519	583.14	-4.18432	627.284
6	-0.7408	0.153815	-458.369	-419.869	-4.22447	-622.637
7	2.26564	-0.34857	947.578	584.702	-0.86093	1378.19
8	-2.4241	0.265867	-602.955	-187.201	-0.30585	-1521.85
9	0.307755	-0.06551	-31.3482	-68.2764	0.909414	394.479
10	0.043983	-0.02994	123.551	104.164	0.167985	-93.0508
11	0.300365	-0.03883	84.9804	21.2973	0	311.625

Table 5.2: Analytically computed partial derivatives of the objective function using the amplitudes of compressional waves. All of the frequency samples have been used.

Layer	Vp	Vs	δ	ε	z	ρ
1	-6.26621	-0.09953	-6273.07	-21944	-26.6804	-277.906
2	-7.44387	0.646046	-1381.38	-17101.3	-28.5673	-1273.2
3	18.56206	-1.33198	5080.25	30686.53	-17.2689	2633.494
4	-2.78341	0.802407	-1262.2	-1403.07	-18.0754	-1365.75
5	1.667401	-0.32669	706.4241	580.706	-4.185	627.2889
6	-0.73696	0.153924	-453.123	-409.382	-4.23169	-622.629
7	2.265887	-0.34865	948.284	585.0287	-0.85651	1378.193
8	-2.42499	0.266044	-605.434	-190.269	-0.30679	-1521.84
9	0.308048	-0.06548	-30.6895	-67.2433	0.910864	394.4844
10	0.044387	-0.02994	123.3061	103.9389	0.167737	-93.04
11	0.300542	-0.03882	84.99086	21.2978	0	311.6278

Table 5.3: Numerically computed partial derivatives of the objective function using the amplitudes of compressional waves. All of the frequency samples have been used.

Layer	Vp	Vs	δ	ε	z	ρ
1	-6.26541	-0.09948	-6238.69	-22563	-24.6491	-277.726
2	-7.60276	0.646301	-1250.76	-17623.4	-28.4577	-1272.5
3	18.5402	-1.33373	5192.59	30094	-16.912	2632.14
4	-2.73944	0.801568	-1229.14	-1449.48	-18.1554	-1365.07
5	1.66769	-0.32664	710.203	582.881	-4.1893	626.917
6	-0.7398	0.153728	-458.043	-419.635	-4.22943	-622.299
7	2.265	-0.34836	947.131	584.478	-0.86658	1377.38
8	-2.42163	0.265705	-602.43	-187.026	-0.31518	-1520.88
9	0.307701	-0.06546	-31.2536	-68.186	0.908211	394.218
10	0.0439	-0.02992	123.466	104.095	0.168217	-93.0287
11	0.30017	-0.0388	84.9204	21.2815	0	311.426

Table 5.4: Analytically computed partial derivatives of the objective function using the amplitudes of compressional waves. Only a quarter of the frequency samples have been used.

Appendices

Appendix A: Partial derivatives of offset and traveltime in a horizontal VTI layered medium

The following equations in this appendix and appendix 5.B are only given for P-wave traveltimes. These can be extended to shear wave traveltimes simply by changing the v_p to v_s and H_p to H_s. Care has to be taken where taking the derivatives of H_s with respect to v_p or v_s.

$$\left.\frac{\partial x}{\partial v_p}\right|_{p=Const.} = \frac{\left(\frac{\partial}{\partial v_{p,j}} H_{p,j} + \frac{1}{2} p \left(\frac{\partial^2}{\partial v_{p,j} \partial p} H_{p,j}\right)\right) p v_{p,j} z_j}{\sqrt{1-\left(1+H_{p,j}\right) p^2 v_p^2}} + \frac{\left(1+H_{p,j} + \frac{1}{2} p \left(\frac{\partial}{\partial p} H_{p,j}\right)\right) p z_j}{\sqrt{1-\left(1+H_{p,j}\right) p^2 v_p^2}}$$

$$- \frac{1}{2} \frac{\left(1+H_{p,j} + \frac{1}{2} p \left(\frac{\partial}{\partial p} H_{p,j}\right)\right) p v_{p,j} z_j \left(-\left(\frac{\partial}{\partial v_{p,j}} H_{p,j}\right) p^2 v_{p,j}^2 - 2\left(1+H_{p,j}\right) p^2 v_{p,j}\right)}{\left(1-\left(1+H_{p,j}\right) p^2 v_{p,j}^2\right)^{3/2}} \quad \text{, A-1}$$

$$\left.\frac{\partial x}{\partial v_s}\right|_{p=Const.} = \frac{\left(\frac{\partial}{\partial v_{s,j}} H_{p,j} + \frac{1}{2} p \left(\frac{\partial^2}{\partial v_{s,j} \partial p} H_{p,j}\right)\right) p z_j v_{p,j}}{\sqrt{1-\left(1+H_{p,j}\right) p^2 v_{p,j}^2}}$$

$$+ \frac{1}{2} \frac{\left(1+H_{p,j} + \frac{1}{2} p \left(\frac{\partial}{\partial p} H_{p,j}\right)\right) p z_j v_{p,j}^3 \left(\frac{\partial}{\partial v_{s,j}} H_{p,j}\right) p^2}{\left(1-\left(1+H_{p,j}\right)\right) p^2 v_{p,j}^2\Big)^{(3/2)}} \quad ,$$

A-2

$$\left.\frac{\partial x}{\partial \delta_j}\right|_{p=Const.} = \frac{\left(\frac{\partial}{\partial \delta_j} H_{p,j} + \frac{1}{2} p \left(\frac{\partial^2}{\partial \delta_j \partial p} H_{p,j}\right)\right) p z_j v_{p,j}}{\sqrt{1-\left(1+H_{p,j}\right) p^2 v_{p,j}^2}}$$

$$+ \frac{1}{2} \frac{\left(1+H_{p,j} + \frac{1}{2} p \left(\frac{\partial}{\partial p} H_{p,j}\right)\right) p z_j v_{p,j}^3 \left(\frac{\partial}{\partial \delta_j} H_{p,j}\right) p^2}{\left(1-\left(1+H_{p,j}\right)\right) p^2 v_{p,j}^2\Big)^{(3/2)}} \quad ,$$

A-3

$$\left.\frac{\partial x}{\partial \varepsilon_j}\right|_{p=Const.} = \frac{\left(\frac{\partial}{\partial \varepsilon_j} H_{p,j} + \frac{1}{2} p \left(\frac{\partial^2}{\partial \varepsilon_j \partial p} H_{p,j}\right)\right) p z_j v_{p,j}}{\sqrt{1-\left(1+H_{p,j}\right) p^2 v_{p,j}^2}} + \frac{1}{2} \frac{\left(1+H_{p,j}+\frac{1}{2} p \left(\frac{\partial}{\partial p} H_{p,j}\right)\right) p z_j v_{p,j}^3 \left(\frac{\partial}{\partial \varepsilon_j} H_{p,j}\right) p^2}{\left(1-\left(1+H_{p,j}\right)\right) p^2 v_{p,j}^2)^{(3/2)}} \quad , \qquad \text{A-4}$$

$$\left.\frac{\partial x}{\partial z_j}\right|_{p=Const.} = \frac{\left(1+H_{p,j}+\frac{1}{2} p \frac{\mathrm{d}}{\mathrm{d}p} H_{p,j}\right) v_{p,j}}{\sqrt{1-\left(1+H_{p,j}\right) p^2 v_{p,j}^2}} \quad , \qquad \text{A-5}$$

$$\left.\frac{\partial T}{\partial v_{p,k}}\right|_{p=Const.} = \frac{\left(p^3 v_{p,k}\left(\frac{\partial}{\partial p} H_{p,k}\right) + \frac{1}{2} p^3 v_{p,k}^2 \left(\frac{\partial^2}{\partial v_{p,k} \partial p} H_{p,k}\right)\right) z_k}{v_{p,k}\sqrt{1-\left(1+H_{p,k}\right) p^2 v_{p,k}^2}} - \frac{\left(1+\frac{1}{2} p^3 v_{p,k}^2 \left(\frac{\partial}{\partial p} H_{p,k}\right)\right) z_k}{v_{p,k}^2 \sqrt{1-\left(1+H_{p,k}\right) p^2 v_{p,k}^2}}$$

$$- \frac{1}{2} \frac{\left(1+\frac{1}{2} p^3 v_{p,k}^2 \left(\frac{\partial}{\partial p} H_{p,k}\right)\right) z_k \left(-\left(\frac{\partial}{\partial v_{p,k}} H_{p,k}\right) p^2 v_{p,k}^2 - 2\left(1+H_{p,k}\right) p^2 v_{p,k}\right)}{v_{p,k}\left(1-\left(1+H_{p,k}\right) p^2 v_{p,k}^2\right)^{(3/2)}} \quad , \qquad \text{A-6}$$

$$\left.\frac{\partial T}{\partial v_{s,k}}\right|_{p=Const.} = \frac{1}{2} \frac{p^3 v_{p,k}\left(\frac{\partial^2}{\partial v_{s,k} \partial p} H_{p,k}\right) z_k}{\sqrt{1-\left(1+H_{p,k}\right) p^2 v_{p,k}^2}} + \frac{1}{2} \frac{\left(1+\frac{1}{2} p^3 v_{p,k}^2 \left(\frac{\partial}{\partial p} H_{p,k}\right)\right) z_k v_{p,k} \left(\frac{\partial}{\partial v_{s,k}} H_{p,k}\right) p^2}{\left(1-\left(1+H_{p,k}\right) p^2 v_{p,k}^2\right)^{(3/2)}}$$

, A-7

$$\left.\frac{\partial T}{\partial \delta_k}\right|_{p=Const.} = \frac{1}{2} \frac{p^3 v_{p,k}\left(\frac{\partial^2}{\partial \delta_k \partial p} H_{p,k}\right) z_k}{\sqrt{1-\left(1+H_{p,k}\right) p^2 v_{p,k}^2}} + \frac{1}{2} \frac{\left(1+\frac{1}{2} p^3 v_{p,k}^2 \left(\frac{\partial}{\partial p} H_{p,k}\right)\right) z_k v_{p,k} \left(\frac{\partial}{\partial \delta_k} H_{p,k}\right) p^2}{\left(1-\left(1+H_{p,k}\right) p^2 v_{p,k}^2\right)^{(3/2)}}$$

, A-8

$$\left.\frac{\partial T}{\partial \varepsilon_k}\right|_{p=Const.} = \frac{1}{2} \frac{p^3 v_{p,k}\left(\frac{\partial^2}{\partial \varepsilon_k \partial p} H_{p,k}\right) z_k}{\sqrt{1-\left(1+H_{p,k}\right) p^2 v_{p,k}^2}} + \frac{1}{2} \frac{\left(1+\frac{1}{2} p^3 v_{p,k}^2 \left(\frac{\partial}{\partial p} H_{p,k}\right)\right) z_k v_{p,k} \left(\frac{\partial}{\partial \varepsilon_k} H_{p,k}\right) p^2}{\left(1-\left(1+H_{p,k}\right) p^2 v_{p,k}^2\right)^{(3/2)}}$$

, A-9

$$\left.\frac{\partial T}{\partial z_k}\right|_{p=Const.} = \frac{1+\frac{1}{2}p^3 v_{p,k}^2 \frac{\partial}{\partial p} H_{p,k}}{v_{p,k}\sqrt{1-\left(1+H_{p,k}\right)p^2 v_{p,k}^2}} \quad , \qquad \text{A-10}$$

$$\frac{dT}{dp} = \sum_{k=1}^{N-1}\left(\frac{\left(\frac{3}{2}p^2 v_{p,k}^2 \frac{d}{dp}H_{p,k} + \frac{1}{2}p^3 v_{p,k}^2 \frac{d^2}{dp^2}H_{p,k}\right)z_k}{v_{p,k}\sqrt{1-\left(1+H_{p,k}\right)p^2 v_{p,k}^2}}\right.$$

$$\left. -\frac{1}{2}\frac{\left(1+\frac{1}{2}p^3 v_{p,k}^2 \frac{d}{dp}H_{p,k}\right)z_k\left(-p^2 v_{p,k}^2 \frac{d}{dp}H_{p,k} - 2\left(1+H_{p,k}\right)pv_{p,k}^2\right)}{v_{p,k}\left(1-\left(1+H_{p,k}\right)p^2 v_{p,k}^2\right)^{\frac{3}{2}}}\right) . \qquad \text{A-11}$$

Appendix B: Partial derivatives of offset and traveltime in a dipping TI layered medium

As we mentioned in Chapter 2, each ray element has its own ray parameter. For simplicity we avoid subscript layer index k for ray parameter p in the following equations,

$$\left.\frac{\partial x}{\partial m_k}\right|_{p=Const.} = -\frac{\left(Z_k + a_k tan\left(\gamma_l\right) + a_k tan\left(\gamma_u\right)\right)\frac{d}{dm_k}cot\left(\alpha\right)}{\left(cot\left(\alpha\right)+tan\left(\gamma_l\right)\right)^2} ,$$

where m could be $v_p, v_s, \delta, \varepsilon$. $cot\left(\alpha\right)$ is given in Chapter 2.

$$\left.\frac{dcot\left(\alpha\right)}{dv_{p,k}}\right|_{p=Const.} =$$

$$\frac{1}{2}\frac{-2p^2 v_{p,k}\left(1+H_{p,k}\right) - p^2 v_{p,k}^2\left(\frac{\partial}{\partial v_{p,k}}H_{p,k}\right)}{\sqrt{1-p^2 v_{p,k}^2\left(1+H_{p,k}\right)}\, p\, v_{p,k}\left(1+H_{p,k}+0.5p\left(\frac{\partial}{\partial p}H_{p,k}\right)\right)} - \frac{\sqrt{1-p^2 v_{p,k}^2\left(1+H_{p,k}\right)}}{p\, v_{p,k}^2\left(1+H_{p,k}+0.5p\left(\frac{\partial}{\partial p}H_{p,k}\right)\right)}$$

$$- \frac{\sqrt{1-p^2 v_{p,k}^2\left(1+H_{p,k}\right)}\left(\frac{\partial}{\partial v_{p,k}}H_{p,k}+0.5p\left(\frac{\partial^2}{\partial v_{p,k}\partial p}H_{p,k}\right)\right)}{p\, v_{p,k}\left(1+H_{p,k}+0.5p\left(\frac{\partial}{\partial p}H_{p,k}\right)\right)^2} , \qquad \text{B-1}$$

$$\left.\frac{dcot(\alpha)}{dv_{s,k}}\right|_{p=Const.} =$$

$$-\frac{1}{2}\,\frac{p\,v_{p,k}^2\left(\frac{\partial}{\partial v_{s,k}}\,H_{p,k}\right)}{\sqrt{1-p^2\,v_{p,k}^2\left(1+H_{p,k}\right)}\;v_{p,k}\left(1+H_{p,k}+0.5p\left(\frac{\partial}{\partial p}\,H_{p,k}\right)\right)}$$

$$-\frac{\sqrt{1-p^2\,v_{p,k}^2\left(1+H_{p,k}\right)}\left(\frac{\partial}{\partial v_{s,k}}\,H_{p,k}+0.5\,p\left(\frac{\partial^2}{\partial v_{s,k}\,\partial p}\,H_{p,k}\right)\right)}{p\,v_{p,k}\left(1+H_{p,k}+0.5\,p\left(\frac{\partial}{\partial p}\,H_{p,k}\right)\right)^2}\;, \qquad \text{B-2}$$

$$\left.\frac{dcot(\alpha)}{d\delta_k}\right|_{p=Const.} =$$

$$-\frac{1}{2}\,\frac{p\,v_{p,k}^2\left(\frac{\partial}{\partial \delta_k}\,H_{p,k}\right)}{\sqrt{1-p^2\,v_{p,k}^2\left(1+H_{p,k}\right)}\;v_{p,k}\left(1+H_{p,k}+0.5p\left(\frac{\partial}{\partial p}\,H_{p,k}\right)\right)}$$

$$-\frac{\sqrt{1-p^2\,v_{p,k}^2\left(1+H_{p,k}\right)}\left(\frac{\partial}{\partial \delta_k}\,H_{p,k}+0.5\,p\left(\frac{\partial^2}{\partial \delta_k\,\partial p}\,H_{p,k}\right)\right)}{p\,v_{p,k}\left(1+H_{p,k}+0.5\,p\left(\frac{\partial}{\partial p}\,H_{p,k}\right)\right)^2}\;, \qquad \text{B-3}$$

$$\left.\frac{dcot(\alpha)}{d\varepsilon_k}\right|_{p=Const.} =$$

$$\frac{1}{2}\,\frac{p\,v_{p,k}^2\left(\frac{\partial}{\partial \varepsilon_k}\,H_{p,k}\right)}{\sqrt{1-p^2\,v_{p,k}^2\left(1+H_{p,k}\right)}\;v_{p,k}\left(1+H_{p,k}+0.5p\left(\frac{\partial}{\partial p}\,H_{p,k}\right)\right)}$$

$$-\frac{\sqrt{1-p^2\,v_{p,k}^2\left(1+H_{p,k}\right)}\left(\frac{\partial}{\partial \varepsilon_k}\,H_{p,k}+0.5\,p\left(\frac{\partial^2}{\partial \varepsilon_k\,\partial p}\,H_{p,k}\right)\right)}{p\,v_{p,k}\left(1+H_{p,k}+0.5\,p\left(\frac{\partial}{\partial p}\,H_{p,k}\right)\right)^2}\;, \qquad \text{B-4}$$

$$\left.\frac{dx}{dz_k}\right|_{p=Const.} = -\frac{1}{\left(cot(\alpha)+tan(\gamma_l)\right)^2}\;, \qquad \text{B-5}$$

$$\left.\frac{\partial T}{\partial v_{p,k}}\right|_{p=Const.} = \frac{\left(p^3\,v_{p,k}^2\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)+\frac{1}{2}\,p^3\,v_{p,k}^2\left(\frac{\partial^2}{\partial p\,\partial v_{p,k}}\,H_{p,k}\right)\right)X_k}{p\,v_{p,k}^2\left(1+H_{p,k}+\frac{1}{2}\,p\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)}$$

$$+\frac{\left(1+\frac{1}{2}\,p^3\,v_{p,k}^2\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)\left(\frac{\mathrm{d}}{\mathrm{d}v_{p0}}\,X_k\right)}{p\,v_{p,k}^2\left(1+H_{p,k}+\frac{1}{2}\,p\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)}-\frac{2\left(1+\frac{1}{2}\,p^3\,v_{p,k}^2\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)X_k}{p\,v_{p,k}^3\left(1+H_{p,k}+\frac{1}{2}\,p\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)}$$

$$-\frac{\left(1+\frac{1}{2}\,p^3\,v_{p,k}^2\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)X_k\left(\frac{\mathrm{d}}{\mathrm{d}v_{p,k}}\,H_{p,k}+\frac{1}{2}\,p\left(\frac{\partial^2}{\partial p\,\partial v_{p,k}}\,H_{p,k}\right)\right)}{p\,v_{p,k}^2\left(1+H_{p,k}+\frac{1}{2}\,p\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)^2}\quad,\qquad\text{B-6}$$

$$\left.\frac{\partial T}{\partial v_{s,k}}\right|_{p=Const.}=$$

$$\frac{1}{2}\,\frac{p^2\left(\frac{\partial^2}{\partial v_{s,k}\,\partial p}\,H_{p,k}\right)X_k}{1+H_{p,k}+\frac{1}{2}\,p\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)}+\frac{\left(1+\frac{1}{2}\,p^3\,v_{p,k}^2\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)\left(\frac{\mathrm{d}}{\mathrm{d}v_{s,k}}\,X_k\right)}{p\,v_{p,k}^2\left(1+H_{p,k}+\frac{1}{2}\,p\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)}$$

$$-\frac{\left(1+\frac{1}{2}\,p^3\,v_{p,k}^2\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)X_k\left(\frac{\partial}{\partial v_{s,k}}\,H_{p,k}+\frac{1}{2}\,p\left(\frac{\partial^2}{\partial v_{s,k}\,\partial p}\,H_{p,k}\right)\right)}{p\,v_{p,k}^2\left(1+H_{p,k}+\frac{1}{2}\,p\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)^2}\quad,\qquad\text{B-7}$$

$$\left.\frac{\partial T}{\partial \delta_k}\right|_{p=Const.}=$$

$$\frac{1}{2}\,\frac{p^2\left(\frac{\partial^2}{\partial \delta_k\,\partial p}\,H_{p,k}\right)X_k}{1+H_{p,k}+\frac{1}{2}\,p\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)}+\frac{\left(1+\frac{1}{2}\,p^3\,v_{p,k}^2\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)\left(\frac{\mathrm{d}}{\mathrm{d}\delta_k}\,X_k\right)}{p\,v_{p,k}^2\left(1+H_{p,k}+\frac{1}{2}\,p\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)}$$

$$-\frac{\left(1+\frac{1}{2}\,p^3\,v_{p,k}^2\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)X_k\left(\frac{\partial}{\partial \delta_k}\,H_{p,k}+\frac{1}{2}\,p\left(\frac{\partial^2}{\partial \delta_k\,\partial p}\,H_{p,k}\right)\right)}{p\,v_{p,k}^2\left(1+H_{p,k}+\frac{1}{2}\,p\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)^2}\quad,\qquad\text{B-8}$$

$$\left.\frac{\partial T}{\partial \varepsilon_k}\right|_{p=Const.}=\frac{1}{2}\,\frac{p^2\left(\frac{\partial^2}{\partial \varepsilon_k\,\partial p}\,H_{p,k}\right)X_k}{1+H_{p,k}+\frac{1}{2}\,p\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)}+\frac{\left(1+\frac{1}{2}\,p^3\,v_{p,k}^2\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)\left(\frac{\mathrm{d}}{\mathrm{d}\varepsilon_k}\,X_k\right)}{p\,v_{p,k}^2\left(1+H_{p,k}+\frac{1}{2}\,p\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)}$$

$$-\frac{\left(1+\frac{1}{2}\,p^3\,v_{p,k}^2\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)X_k\left(\frac{\partial}{\partial \varepsilon_k}\,H_{p,k}+\frac{1}{2}\,p\left(\frac{\partial^2}{\partial \varepsilon_k\,\partial p}\,H_{p,k}\right)\right)}{p\,v_{p,k}^2\left(1+H_{p,k}+\frac{1}{2}\,p\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)^2}\quad,\qquad\text{B-9}$$

$$\left.\frac{\partial T}{\partial z_k}\right|_{p=Const.}=\frac{\left(1+\frac{1}{2}\,p^3\,v_{p,k}^2\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)\frac{\mathrm{d}}{\mathrm{d}\,z_k}\,X}{p\,v_{p,k}^2\left(1+H_{p,k}+\frac{1}{2}\,p\left(\frac{\mathrm{d}}{\mathrm{d}p}\,H_{p,k}\right)\right)}\quad,\qquad\text{B-10}$$

where dX/dz_k is defined in (B-5).

$$\frac{dT}{dp} = \sum_{k=1}^{N-1} \frac{dT}{dp_k}, \tag{B-11}$$

$$\begin{aligned}\frac{dT}{dp_k} = & \frac{\left(\frac{3}{2} p^2 v_{p,k}^2 \left(\frac{\mathrm{d}}{\mathrm{d}p} H_{p,k}\right) + \frac{1}{2} p^3 v_{p,k}^2 \left(\frac{\mathrm{d}^2}{\mathrm{d}p^2} H_{p,k}\right)\right) X_k}{p v_{p,k}^2 \left(1 + H_{p,k} + \frac{1}{2} p \left(\frac{\mathrm{d}}{\mathrm{d}p} H_{p,k}\right)\right)} \\ & + \frac{\left(1 + \frac{1}{2} p^3 v_{p,k}^2 \left(\frac{\mathrm{d}}{\mathrm{d}p} H_{p,k}\right)\right) \left(\frac{\mathrm{d}}{\mathrm{d}p} X_k\right)}{p v_{p,k}^2 \left(1 + H_{p,k} + \frac{1}{2} p \left(\frac{\mathrm{d}}{\mathrm{d}p} H_{p,k}\right)\right)} - \frac{\left(1 + \frac{1}{2} p^3 v_{p,k}^2 \left(\frac{\mathrm{d}}{\mathrm{d}p} H_{p,k}\right)\right) X_k}{p^2 v_{p,k}^2 \left(1 + H_{p,k} + \frac{1}{2} p \left(\frac{\mathrm{d}}{\mathrm{d}p} H_{p,k}\right)\right)} \\ & - \frac{\left(1 + \frac{1}{2} p^3 v_{p,k}^2 \left(\frac{\mathrm{d}}{\mathrm{d}p} H_{p,k}\right)\right) X_k \left(\frac{3}{2} \frac{\mathrm{d}}{\mathrm{d}p} H_{p,k} + \frac{1}{2} p \left(\frac{\mathrm{d}^2}{\mathrm{d}p^2} H_{p,k}\right)\right)}{p v_{p,k}^2 \left(1 + H_{p,k} + \frac{1}{2} p \left(\frac{\mathrm{d}}{\mathrm{d}p} H_{p,k}\right)\right)^2}, \end{aligned} \tag{B-12}$$

Appendix C: Partial derivatives of H_P and H_S with respect to model parameters and ray parameter

We can write down equation (2.15) in each layer with a good accuracy using the first few terms only. To avoid extra indices we omit the layer indices. The subscripts p and s refer to P-wave and to SV-waves respectively. We express H_s as a function H_p to reduce the computations. a terms are given in Chapter 2.

$$H_p = 2\delta + a_1 \left(pv_p\right)^2 + a_2 \left(pv_p\right)^4 + a_3 \left(pv_p\right)^6, \tag{C-1}$$

$$\frac{dH_p}{dp} = 2a_1 pv_p^2 + 4a_2 p^3 v_p^4 + 6a_3 p^5 v_p^6 , \tag{C-2}$$

$$\frac{dH_p}{dv_p} = \left(pv_p\right)^2 \frac{da_1}{dv_p} + 2p\left(pv_p\right)a_1 + \left(pv_p\right)^4 \frac{da_2}{dv_p} + 4p\left(pv_p\right)^3 a_2 + \left(pv_p\right)^6 \frac{da_3}{dv_p} + 6p\left(pv_p\right)^5 a_3 , \tag{C-3}$$

$$\frac{dH_p}{dv_s} = \left(pv_p\right)^2 \frac{da_1}{dv_s} + \left(pv_p\right)^4 \frac{da_2}{dv_s} + \left(pv_p\right)^6 \frac{da_3}{dv_s} , \tag{C-4}$$

$$\frac{dH_p}{d\delta} = 2 + \left(pv_p\right)^2 \frac{da_1}{d\delta} + \left(pv_p\right)^4 \frac{da_2}{d\delta} + \left(pv_p\right)^6 \frac{da_3}{d\delta} \ , \qquad \text{C-5}$$

$$\frac{dH_p}{d\varepsilon} = \left(pv_p\right)^2 \frac{da_1}{d\varepsilon} + \left(pv_p\right)^4 \frac{da_2}{d\varepsilon} + \left(pv_p\right)^6 \frac{da_3}{d\varepsilon} \ , \qquad \text{C-6}$$

$$\frac{\partial^2 H_p}{\partial v_p \partial p} = 4\left(pv_p\right)a_1 + 2pv_p^2 \frac{da_1}{dv_p} + 16\left(pv_p\right)^3 a_2 + 4p^3 v_p^4 \frac{da_2}{dv_p} + 36\left(pv_p\right)^5 a_3 + 6p^5 v_p^6 \frac{da_3}{dv_p} \ , \qquad \text{C-7}$$

$$\frac{\partial^2 H_p}{\partial v_s \partial p} = 2pv_p^2 \frac{da_1}{dv_s} + 4p^3 v_p^4 \frac{da_2}{dv_s} + 6p^5 v_p^6 \frac{da_3}{dv_s} \ , \qquad \text{C-8}$$

$$\frac{\partial^2 H_p}{\partial \delta \partial p} = 2a_1 pv_p^2 \frac{da_1}{d\delta} + 4a_2 p^3 v_p^4 \frac{da_2}{d\delta} + 6a_3 p^5 v_p^6 \frac{da_3}{d\delta} \ , \qquad \text{C-9}$$

$$\frac{\partial^2 H_p}{\partial \varepsilon \partial p} = 2a_1 pv_p^2 \frac{da_1}{d\varepsilon} + 4a_2 p^3 v_p^4 \frac{da_2}{d\varepsilon} + 6a_3 p^5 v_p^6 \frac{da_3}{d\varepsilon} \ , \qquad \text{C-10}$$

$$\frac{da_1}{dv_p} = -\frac{8\left(\varepsilon - \delta\right)\delta\gamma_0}{\left(\gamma_0^2 - 1\right)^2 v_s} \ , \qquad \text{C-11}$$

$$\frac{da_1}{dv_s} = -\gamma_0 \frac{da_1}{dv_p} \ , \qquad \text{C-12}$$

$$\frac{da_2}{dv_p} = \frac{da_1}{dv_p} \frac{\gamma_0^2\left(-6\delta^2 + 4\varepsilon\delta + \varepsilon - 2\delta\right) + \left(-2\delta^2 - \varepsilon + 2\delta\right)}{\delta\left(\gamma_0^2 - 1\right)} \ , \qquad \text{C-13}$$

$$\frac{da_2}{dv_s} = -\gamma_0 \frac{da_2}{dv_p} \ , \qquad \text{C-14}$$

$$\frac{da_3}{dv_p} = \left(\frac{a_3}{a_2} + \frac{a_2}{a_1}\right)\frac{da_2}{dv_p} + \left(\frac{a_2}{2\left(\delta - \sigma\right)} - \left(\frac{a_2}{a_1}\right)^2\right)\frac{da_1}{dv_p} + \frac{a_1 a_2 \sigma}{v_p\left(\delta - \sigma\right)^2} \ , \qquad \text{C-15}$$

$$\frac{da_3}{dv_s} = -\gamma_0 \frac{da_3}{dv_p} \ , \qquad \text{C-16}$$

$$\frac{da_1}{d\delta} = \frac{4\gamma_0^2}{\gamma_0^2 - 1}\left(\varepsilon - 2\delta\right) - 2 \ , \qquad \text{C-17}$$

$$\frac{da_2}{d\delta} = -2\frac{da_1}{d\delta}\left(\frac{\delta - \sigma}{\gamma_0^2 - 1}\right) - 2a_1 \frac{\gamma_0^2 + 1}{\gamma_0^2 - 1} \ , \qquad \text{C-18}$$

$$\frac{da_3}{d\delta} = \frac{da_1}{d\delta}\left(\frac{a_2}{2\left(\delta - \sigma\right)} - \left(\frac{a_2}{a_1}\right)^2\right) + \frac{da_2}{d\delta}\left(\frac{a_1}{2\left(\delta - \sigma\right)} + \frac{2a_2}{a_1}\right) - \frac{a_1 a_2\left(1 + \gamma_0^2\right)}{2\left(\delta - \sigma\right)^2} \ , \qquad \text{C-19}$$

$$\frac{da_1}{d\varepsilon} = 2\left(1 + \frac{2\gamma_0^2 \delta}{\gamma_0^2 - 1}\right) \ , \qquad \text{C-20}$$

$$\frac{da_2}{d\varepsilon} = -2\frac{(\delta-\sigma)}{\gamma_0^2-1}\frac{da_1}{d\varepsilon} + 2\frac{a_1\gamma_0^2}{\gamma_0^2-1} \ , \qquad \text{C-21}$$

$$\frac{da_3}{d\varepsilon} = \left(2\frac{a_2}{a_1} + \frac{a_1}{2(\delta-\sigma)}\right)\frac{da_2}{d\varepsilon} + \left(\frac{a_2}{2(\delta-\sigma)} - \left(\frac{a_2}{a_1}\right)^2\right)\frac{da_1}{d\varepsilon} + \frac{a_2 a_1\gamma_0^2}{2(\delta-\sigma)^2} \ , \qquad \text{C-22}$$

Using the equation (2.18) we can express H_s derivative as:

$$\frac{dH_s}{dv_p} = 4\frac{\sigma}{v_p} - \frac{dH_p}{dv_p}, \qquad \text{C-23}$$

$$\frac{dH_s}{dv_s} = -4\frac{\sigma}{v_s} - \frac{dH_p}{dv_s} \ , \qquad \text{C-24}$$

$$\frac{dH_s}{d\delta} = 2 - 2\gamma_0^2 - \frac{dH_p}{d\delta} \ , \qquad \text{C-25}$$

$$\frac{dH_s}{d\varepsilon} = 2\gamma_0^2 - \frac{dH_p}{d\varepsilon} \ , \qquad \text{C-26}$$

$$\frac{\partial^2 H_s}{\partial v_p \partial p} = -\frac{\partial^2 H_p}{\partial v_p \partial p} \ , \qquad \text{C-27}$$

$$\frac{\partial^2 H_s}{\partial v_s \partial p} = -\frac{\partial^2 H_p}{\partial v_s \partial p} \ , \qquad \text{C-28}$$

$$\frac{\partial^2 H_s}{\partial \delta \partial p} = -\frac{\partial^2 H_p}{\partial \delta \partial p} \ , \qquad \text{C-29}$$

$$\frac{\partial^2 H_s}{\partial \varepsilon \partial p} = -\frac{\partial^2 H_p}{\partial \varepsilon \partial p} \ , \qquad \text{C-30}$$

References

Dahl, T., and B. Ursin, 1991, Parameter estimation in one-dimensional anelastic medium: Journal of Geophysical Research, **96**, 20217-20233.

Oldenburg, D. W., and Y. Li, 2005, Inversion for applied geophysics: A tutorial, *in* D. K. Butler, ed., Near-Surface Geophysics: Society of Exploration Geophysicists, 89-150.

Rüger, A., 1996, Reflection coefficients and azimuthal AVO analysis in anisotropic media: PhD, Colorado School of Mines.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: SIAM.

Chapter 6

Numerical results of unconstrained inversion of traveltime and amplitude

Introduction

This chapter implements the methods and algorithms discussed in previous chapters. We choose a stack of transversely isotropic layers and assign them a wide range of elastic parameters. We simulate a prestack offset gather of the traveltimes and synthetic seismograms for a wide range of offsets over a 2-dimensional line. To demonstrate the efficiency of proposed unconstrained minimization algorithms, we perturb the true model in different ways and try to recover the true model via different strategies. Over a horizontally layered media we use both traveltime and waveform inversion, however, only traveltime inversion will be used over a dipping layered medium. We limit the waveform inversion solely to estimation of Thomsen's anisotropy parameters to utilize its capability in recovering the high frequency part of the model. Due to intensive computations involved in dipping media, we use a preconditioning conjugate gradient algorithm to converge to the solution in fewer iterations. This limits the parameter estimation problem to the parameters with a more or less similar scale. To accomplish the inversion we have developed a package in C++ in object oriented environment. The class diagram is given in the appendix at the end of this chapter.

6.1 Traveltime inversion in a horizontally layered VTI media

In the first minimization example we investigate the effect of the offset and uncertainty in the prior models on estimation of model parameters. We restrict the parameter estimation to one layer only and estimate layer properties such as P-wave velocity, thickness, and Thomsen's anisotropy parameters simultaneously from the inversion of near and far offset P-wave traveltimes. We compute 96 P-wave traveltimes corresponding offsets from zero to 4750 metres in 50 metre intervals in a layer with following properties: Vp = 1940 metres/sec, Vs = 780 metres/sec, Z = 1000 metres, δ =

0.03, and ε = 0.1. we consider far and near offsets corresponding all traveltimes from 0 to 4750 metres and zero to 3000 metres, respectively. In each case, the prior model for the P-wave velocity, shear wave velocity, and thickness are kept the same (Vp=2750 metres/sec, Vs=1375 metres/sec, Z=500 metres) but the prior models for δ and ε are drawn randomly from a uniform distribution with upper and lower boundaries in the context of week seismic anisotropy ($-0.05 \le \delta \le 0.2$ and $0 \le \varepsilon \le 0.25$). The condition $\varepsilon \ge \delta$ is applied during the random selection. Further, to investigate the effect of uncertainty in the Thomsen's anisotropy parameters on parameter estimation we minimize the traveltimes with 20 different prior models of δ and ε.

Since the estimation of Thomsen's anisotropy parameter δ depends on both compressional and shear wave velocities and because P-wave traveltimes in a transverse isotropic medium are depending weakly on S-wave velocity, we are not expecting a good estimate of δ while shear wave velocity itself is under estimation. To moderate this effect we constrained shear wave velocity by $Vp/Vs = 2$. This may, however, bias the estimation of δ.

Figures 6.1 and 6.3 illustrate the convergence pattern of P-wave velocity and thickness for 20 realizations from the inversion of far offset traveltimes using a nonlinear conjugate gradient algorithm. After a few iterations both P-wave velocity and thickness have converged to a solution. It can be also seen that the initial Thomsen's anisotropy parameters have a crucial effect on the estimation of P-wave velocity and thickness, where the uncertainty in P-wave velocity and thickness is large. The mean of the P-wave velocity is 1955 metres/sec and standard deviation is 87 metres/sec, while the mean of thickness is 1008 metres and standard deviation is 47 meters. This indicates that the relative error in estimation of P-wave velocity and thickness is almost the same. In figure 6.2 and figure 6.4 we have illustrated the convergence pattern of P-wave velocity and thickness from the inversion of near offset traveltimes for 20 realizations. We have used the same prior models as above for compressional and shear wave velocities and thickness but different random prior models for δ and ε for minimization. It can be seen that the uncertainty in estimation of P-wave velocity and thickness is larger compared to the uncertainty in estimation from the far offsets. The mean of the P-wave velocity is 1925 metres/sec and standard deviation is 117 metres/sec, while the mean of

thickness is 993 metres and standard deviation is 58 meters. As expected, the effect of seismic anisotropy is more significant in far offset, hence, it results in better estimation of P-wave velocity and thickness. Figures 6.5 and 6.6 illustrate the convergence of the objective function for far and near offsets, respectively. In each figure a different curve belongs to a minimization from a prior model with different δ and ε. The Frobenius norm depicts the amount of minimized residuals and does not necessarily mean a better solution. For example, if one parameter, such as thickness or P-wave velocity that has a major effect on the objective function resolves well, the objective function dramatically drops, yet the other parameters may have not been converged to true solutions.

Tables 6.1 and 6.2 show the prior model and estimated model parameters from the previous minimization. They show there is no major update in Thomsen's anisotropy parameters. One of the reasons is the large uncertainty in the prior model of parameters such as P-wave velocity and thickness that have a much larger impact on the objective function. These parameters have larger directional derivatives compared to δ and ε whenever they are far from the solution, hence, they do not comply well with the search direction vectors along the directional derivatives of δ and ε. Another reason is due the large difference in the scale of Thomsen's anisotropy parameters with velocity and thickness. A thorough revising of the traveltime and offset equations by linear scaling of velocity and thickness to the same magnitude of Thomsen's anisotropy parameters may compensate the scaling effect (Ravindran et al., 2006). As an alternative, we may think of inverting the elastic stiffness coefficients instead, which have almost the same scale. We scaled also the P-wave and S-wave velocities and thicknesses to their mean values. We implemented both approaches, but no improvement was achieved. In general, we take the approach of Gunning and Glinsky (2004) to estimate the model parameters that strongly dominate the objective function and then the model parameters with smaller effect. In general, where the uncertainty in the velocity and thickness is large, it may lead up to 10 percent error in estimation of layer thickness and P-wave velocity, respectively.

In the following examples we reduce the uncertainty in the model parameters within 25 percent which is more realistic. This let us to estimate the model parameters in a layered media simultaneously. To compute the traveltimes in a stack of horizontal transversely isotropic layers we use a model (Table 6.3) given by Ursin and Stovas

(2006) with a few modifications. This model provides a wide range of compressional and shear wave velocities, thicknesses, and Thomsen's anisotropic parameters in the context of weak anisotropy. We compute the traveltimes for $P \to P$ and $P \to S_v$ waves from 9 reflectors for 96 traces placed inline at 50 metre intervals. The nearest trace is placed at 200 metre offset from shot point and the farthest at 4950 metre. Figures 2.6 and 2.7 illustrate the $P \to P$ and $P \to S_v$ traveltimes from the above model with the given geometry. These traveltimes will be considered as the observed traveltimes in the following minimization examples. To start the minimization we set a lower and upper boundary for each model parameter in Table 6.3 by ± 25 percent deviation from the true model (Table 6.3). We draw a random prior model from a uniform distribution constrained by the upper and lower boundaries (Table 6.4).

In the second minimization example, we implement a nonlinear conjugate gradient algorithm using the Polak-Ribiére method (Nocedal and Wright, 1999; Bonnans et al., 2003) to simultaneously estimate all the 45 model parameters by inverting the noise free $P \to P$ traveltimes from 96 traces (Nadri and Hartley, 2007a; , 2007b). We continued the minimization up to 500 iterations. Table 6.5 shows the estimated model parameters for the last iteration. P-wave velocities, layer thicknesses, and to some extents ε are fairly well estimated, however, estimates of δ and especially S-wave velocities, as was expected, are not very accurate. Such a good estimation of thickness and P-wave velocity is mainly due to the fact that prior information for Thomsen's anisotropy parameters is close to the true solution and initial model for velocities and thickness are within the 25 percent of the true model.

In the third minimization example, we implement the same minimization algorithm as above to simultaneously estimate all the 45 model parameters by inverting the joint $P \to P$ and $P \to S_v$ traveltimes from 96 traces. We again continued the minimization up to 500 iterations. Table 6.6 demonstrates the estimated model parameters for the last iteration. P-wave and S-wave velocities, and layer thicknesses are fairly well estimated, however, ε and δ are poorly optimized.

To investigate the effect of offset on both $P \to P$ and joint $P \to P$ and $P \to S_v$ traveltime minimization, where the uncertainty in P-wave, S-wave, and thickness is not

large, we limited the number of offsets to the first 48 traces. Hence, the far offset is located at 2550 metre. We implemented the same minimization algorithm and continued up to 500 iterations. Tables 6.7 and 6.8 show the estimated model parameters for the last iteration. In general, reducing the number of offsets did not cause a major effect on the parameter estimation. This offcourse should not bring us to the conclusion that larger offsets are not necessary in parameter estimation. Such a good results may be due to noise free traveltimes or being close to the solution.

To demonstrate the effectiveness of each of the above examples we show, in Figures 6.7-6.12, the convergence pattern of some of the estimated parameters. Figure 6.7 illustrates the convergence pattern of thickness for the top layer from the inversion of $P \to P$ and joint $P \to P$ and $P \to S_v$ traveltimes for the two sets of 48 and 96 traces. The blue curve depicts the $P \to P$ inversion of traveltimes from 96 traces, red the same but for 48 traces, green and cyan curves show the convergence pattern of joint $P \to P$ and $P \to S_v$ traveltime inversion for 96 and 48 traces respectively. All the four minimizations converged to true solutions. To see the numerical values of last iteration refer to Tables 6.5-6.8. Figure 6.8 illustrates the thickness for layer 7 using the same inversion and colour coding as Figure 6.7.

The estimated P-wave velocities for the top layer and layer 7 from the minimization of $P \to P$ and joint $P \to P$ and $P \to S_v$ traveltimes for the two sets of 48 and 96 traces are shown in Figures 6.9 and 6.10, respectively. P-wave velocity in the top layer has converged faster than for layer 7. In general, due to propagation of more rays in top layers compared to lower ones, model parameters in top layers gain more information and this result in faster and more stable convergence.

We have shown the estimated shear wave velocities for the top layer and layer 7 in Figure 6.11 and Figure 6.12 respectively with the same conjugate gradient algorithm as in the above examples. We only plotted the result of joint inversion of $P \to P$ and $P \to S_v$ traveltimes as not much update has occurred from the inversion of $P \to P$ traveltimes. The S-wave velocity from the inversion of longer offsets seems to converge faster than when using only near range offsets. This is a because of the large moveout of shear wave traveltimes at longer offsets.

Figure 6.13 shows the convergence of the objective function of four different minimization examples discussed above. The vertical axis is the Frobenius norm which is simply the objective function in equation (5.29). The objective functions are not normalized as they are generally a representation of the data residuals (errors) left from minimization. The objective function for joint compressional and shear waves obviously uses more measured data, hence their mean square error is generally bigger than that using only compressional waves.

In multidimensional nonlinear minimization problems, it is not easy to interpret the behavior of the objective function or convergence pattern of a specific model parameter. In a linear conjugate gradient algorithm, as mentioned in chapter 4, the model parameters should theoretically converge to the solution in n iteration, where n is the number of model parameters. In nonlinear problems, even in the absence of finite precision problems, which cause loss of orthogonality of the conjugate directions, there is no guarantee to converge even to a local minimum within n iterations. We may however, attribute the rapid jump in objective function to the convergence property of the conjugate gradient algorithm. The convergence pattern of each model parameter is an interaction of all other model parameters with respect to the shape of the objective function in *n- dimension.* Hence, it is not easy to interpret their evolution patterns independently.

We have so far been concerned about simultaneous inversion of all different parameters such as P-wave and S-wave velocity, layer thickness, and Thomsen's anisotropy parameters. Hereafter, we assume that we, by some means, for examples check shot data have a fairly good estimation of vertical velocity and structure model of the layered medium and solely looking at estimation of Thomsen's anisotropy parameters. This will gives us more freedom on the prior knowledge of ε and δ. In the following, we implement a preconditioning conjugate gradient algorithm (Saad, 2000) to estimate Thomsen's anisotropy parameters in the context of weak anisotropy. This means, with the assumption of weak anisotropy (equation 3.30), we do not need to have an explicit prior model for Thomsen's anisotropy parameters. The preconditioning utilizes a Numerically computed Hessian matrix based on the method of BFGS (Nocedal and Wright, 1999; Press et al., 2002). Because, Thomsen's anisotropy parameters have

almost the same scale and the Hessian matrix is homogenous and well conditioned, it could be used as a preconditioner.

Figure 6.14 shows the convergence of the objective function from minimization of $P \to P$ traveltimes over 96 traces using a preconditioned conjugate gradient algorithm. Minimization has been carried out over five different prior models randomly drawn from a uniform distribution constrained by the weak anisotropy assumption. We assume $-0.05 \le \delta \le 0.2$ and $0 \le \varepsilon \le 0.2$. We continued minimization for up to 250 iterations, but as it can be seen, all the objective functions are very small after 25 iterations.

In the upper part of the Table 6.9 we have shown 5 different prior models for ε and δ and in the lower part, the estimated model parameters from the last iteration of the minimization for the related prior model. It can be seen that all of the model parameters have converged to their true solution. The first two columns correspond to the red curve in Figure 6.14, the second two columns to blue, the third two columns to green, the fourth two columns to brown, and the last two columns to the magenta curve. The brown curve has reached the least value of the objective function. This is due to the fact that its corresponding prior model has been drawn by coincidence closer to the true model. We can see the closeness of the prior model for the brown curve in Figures 6.15-6.20.

Figure 6.15 and Figure 6.16 illustrate the convergence pattern of δ and ε from five realizations from the top layer respectively. All realizations have converged to the solution in 20 iterations. The preconditioning has reduced the computational burden significantly. In Figure 6.17 and Figure 6.18 the convergence pattern of δ and ε from five realizations for layer 7 are shown respectively. It can be seen that all realizations have converged in 30 iterations. In Figure 6.19 and Figure 6.20 we have demonstrated the convergence pattern of δ and ε from five realizations for layer 9 respectively. It can be seen that all realizations have converged in 25 iterations.

To compare the effect of preconditioning in conjugate gradient algorithm we have estimated the Thomsen's anisotropy parameters from a system of 9 layers by inverting the $P \to P$ traveltime from 96 traces. The convergence pattern of the objective function as a criterion for comparison has been shown in Figure 6.21. The blue curve illustrates the Polak-Ribiére conjugate gradient algorithm without preconditioning, while a preconditioned algorithm has been shown by the red curve. Both algorithms use the same

prior model. The preconditioned conjugate gradient algorithm has a significant superiority over that without preconditioning. The Polak-Ribiére conjugate gradient algorithm without preconditioning is robust enough and has finally converged to the solution but at the cost of 10 times more iterations.

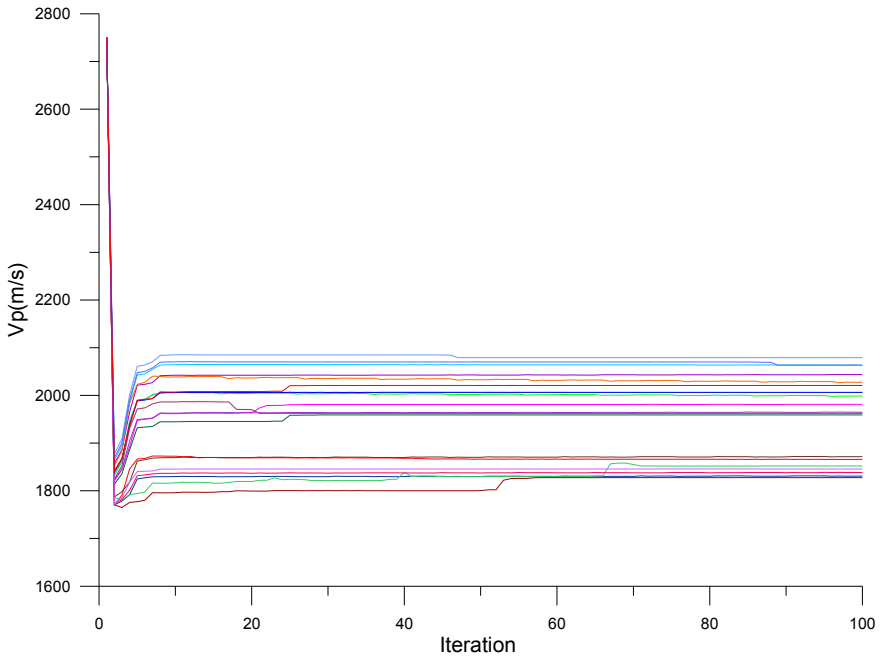


Figure 6.1: Convergence pattern of the 20 realizations of P-wave velocity of layer 1 from the inversion of far offset traveltimes for different δ and ε. The true value is 1940 m/sec. Mean is 1955 m/s and standard deviation is 87 m/sec.

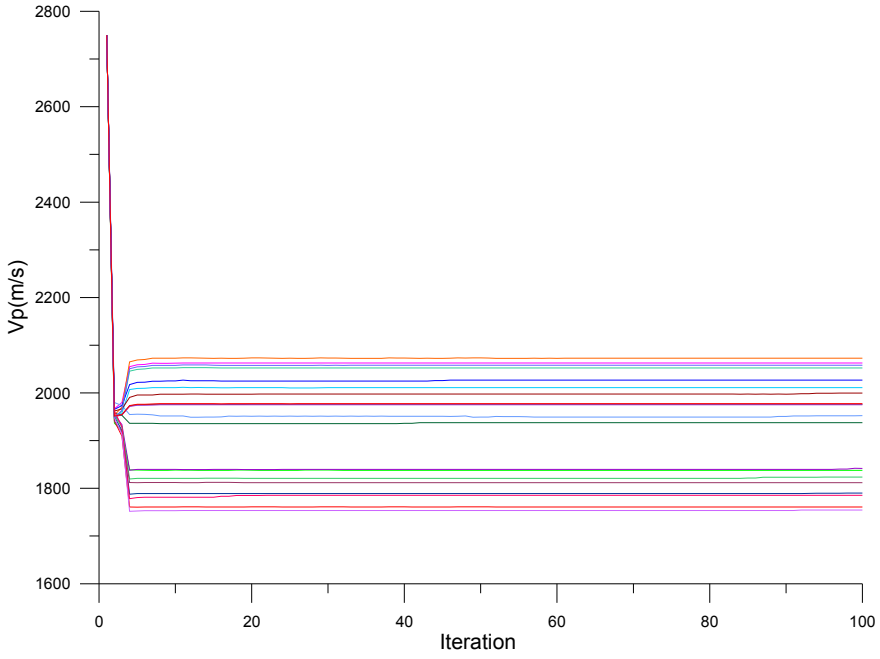


Figure 6.2: Convergence pattern of the 20 realizations of P-wave velocity of layer 1 from the inversion of near offset traveltimes for different δ and ε. The true value is 1940 m/sec. Mean is 1925 m/s and standard deviation is 111 m/sec.

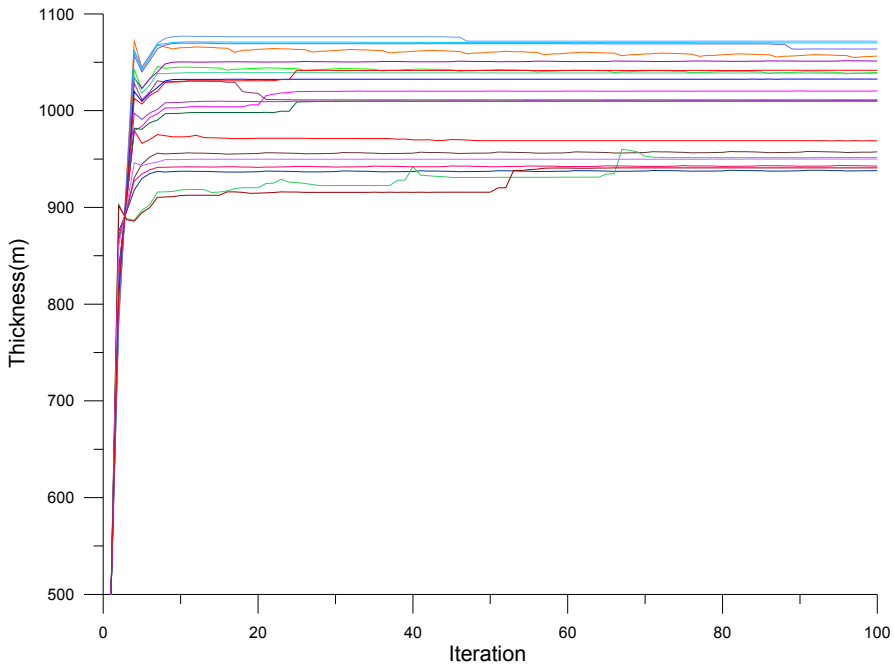


Figure 6.3: Convergence pattern of the 20 realizations of thickness of layer 1 from the inversion of far offset traveltimes for different δ and ε. The true value is 1000 metres. Mean is 1008 metres and standard deviation is 47 metres.

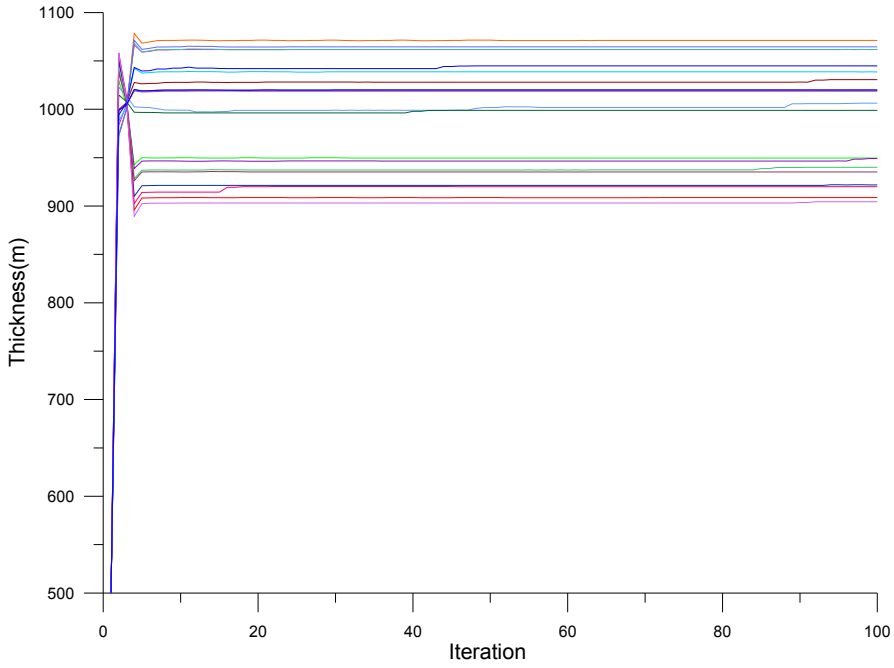


Figure 6.4: Convergence pattern of the 20 realizations of thickness of layer 1 from the inversion of near offset traveltimes for different δ and ε. The true value is 1000 metres. Mean is 993 metres and standard deviation is 58 metres.

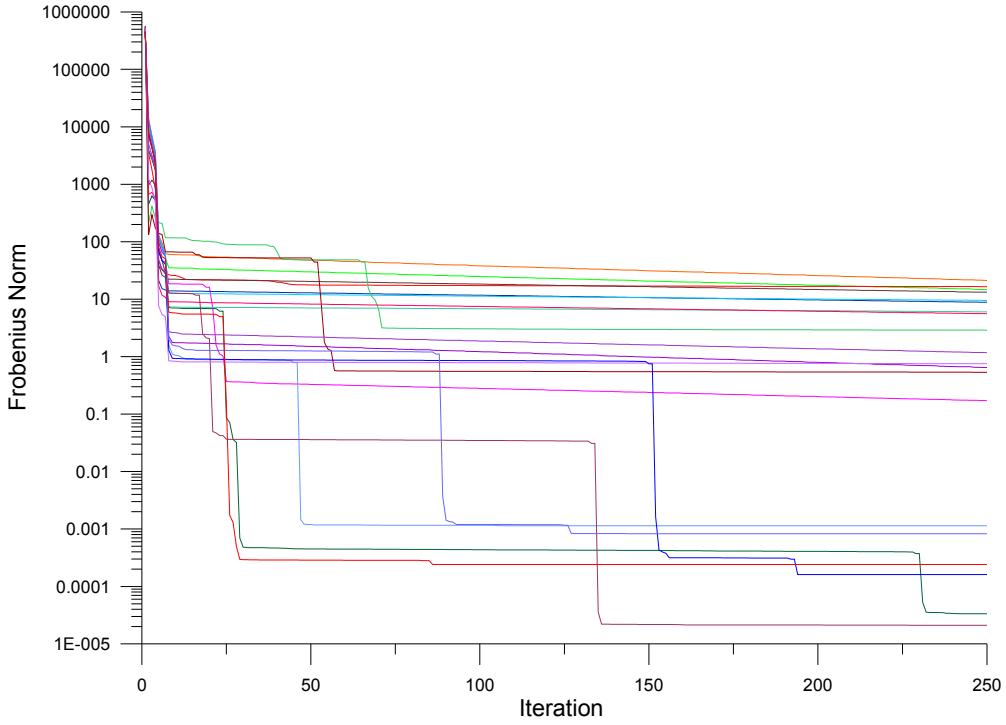


Figure 6.5: Convergence pattern of the 20 realizations of objective function from the inversion of far offset traveltimes of layer 1 for different δ and ε.

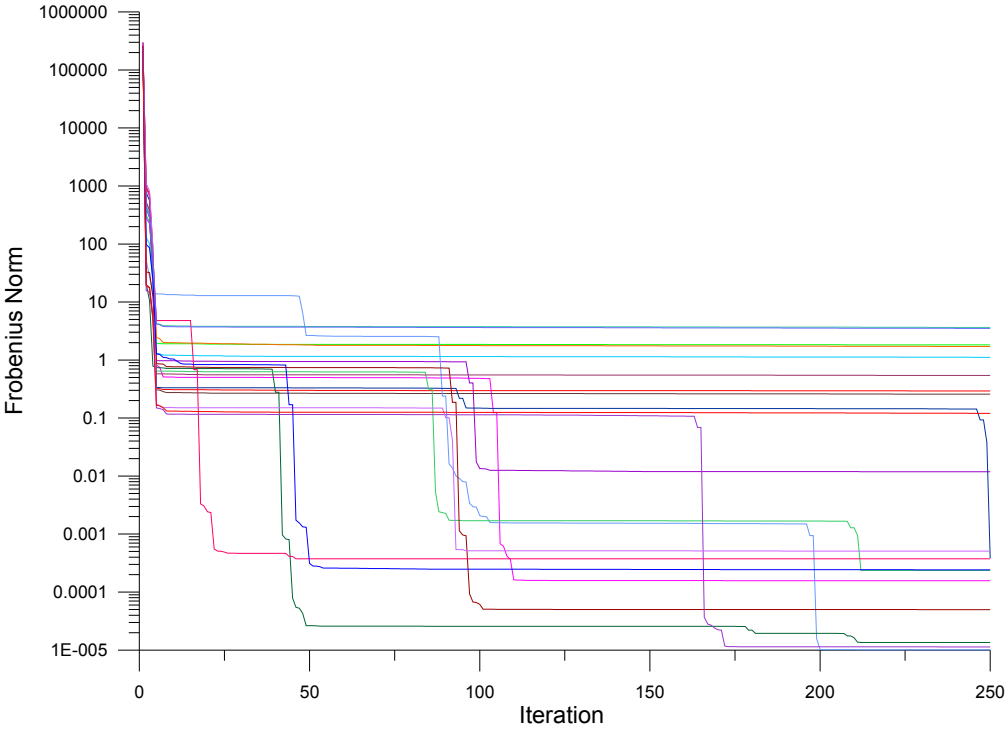


Figure 6.6: Convergence pattern of the 20 realizations of objective function from the inversion of near offset traveltimes of layer 1 for different δ and ε.

Vp(m/sec) (est.)	Z(m) (est.)	δ (prior)	δ (est.)	ε (prior)	ε (est.)	Obj. Func. (Init.)	Obj. Func. (250 Iter)
2063	1064	-0.0205	-0.03	0.028	0.0303	416258	0.000826688
2021	1042	-0.026	-0.011	0.072	0.0527	451401	0.00024071
1991	1033	0.0595	0.043	0.05	0.0583	443940	14.7082
1865	968	0.132	0.122	0.135	0.1353	515782	16.4982
1982	1021	-0.024	0.0039	0.105	0.0758	476887	0.170106
2045	1053	-0.0315	-0.023	0.050	0.0419	433141	0.646181
2016	1048	0.0615	0.0375	0.026	0.0420	423577	21.2554
1874	960	0.018	0.0255	0.169	0.1537	526922	13.2523
2079	1072	-0.029	-0.0376	0.021	0.0222	409329	0.00114151
1833	940	0.0515	0.0565	0.195	0.1811	548048	8.8612
2062	1069	0.00425	-0.0005	0.025	0.0227	416691	9.52178
1959	1010	0.00125	0.0197	0.111	0.0879	484230	3.3435E-005
1852	952	-0.02375	0.0603	0.232	0.1631	560844	2.8879
2005	1038	0.02475	0.0214	0.059	0.0545	447002	6.02033
1828	941	0.018	0.0868	0.237	0.1777	570216	0.535326
1963	1012	0.04425	0.0174	0.068	0.0854	456407	2.10973E-005
1840	944	0.05575	0.0596	0.187	0.1739	543262	5.57953
1966	1011	0.003	0.0048	0.097	0.0868	473556	1.16838
1846	950	0.07525	0.0739	0.173	0.1649	536494	0.753365
2011	1037	-0.0115	-0.0066	0.07	0.0578	450957	0.000161741

Table 6.1: Prior and estimated model of the 20 realizations from traveltime inversion of far offset data for layer 1. Vp(prior)=2750 m/s, Vs(prior)=1375 m/s, and Z(prior) =500 m.

Vp(m/sec) (est.)	Z(m) (est.)	δ (prior)	δ (est.)	ε (prior)	ε (est.)	Obj. Func. (Init.)	Obj. Func. (250 Iter)
2063	1064	-0.0205	-0.03	0.028	0.0303	416258	0.000826688
2021	1042	-0.026	-0.011	0.072	0.0527	451401	0.00024071
1991	1033	0.0595	0.043	0.05	0.0583	443940	14.7082
1865	968	0.132	0.122	0.135	0.1353	515782	16.4982
1982	1021	-0.024	0.0039	0.105	0.0758	476887	0.170106
2045	1053	-0.0315	-0.023	0.050	0.0419	433141	0.646181
2016	1048	0.0615	0.0375	0.026	0.0420	423577	21.2554
1874	960	0.018	0.0255	0.169	0.1537	526922	13.2523
2079	1072	-0.029	-0.0376	0.021	0.0222	409329	0.00114151
1833	940	0.0515	0.0565	0.195	0.1811	548048	8.8612
2062	1069	0.00425	-0.0005	0.025	0.0227	416691	9.52178
1959	1010	0.00125	0.0197	0.111	0.0879	484230	3.3435E-005
1852	952	-0.0237	0.0603	0.232	0.1631	560844	2.8879
2005	1038	0.0247	0.0214	0.059	0.0545	447002	6.02033
1828	941	0.018	0.0868	0.237	0.1777	570216	0.535326
1963	1012	0.0442	0.0174	0.068	0.0854	456407	2.10973E-005
1840	944	0.0557	0.0596	0.187	0.1739	543262	5.57953
1966	1011	0.003	0.0048	0.097	0.0868	473556	1.16838
1846	950	0.0752	0.0739	0.173	0.1649	536494	0.753365
2011	1037	-0.0115	-0.0066	0.07	0.0578	450957	0.000161741

Table 6.2: Prior and estimated model of the 20 realizations from traveltime inversion of near offset data for layer 1. The prior model for Vp, Vs, and Thickness is the same as above table 6.1.

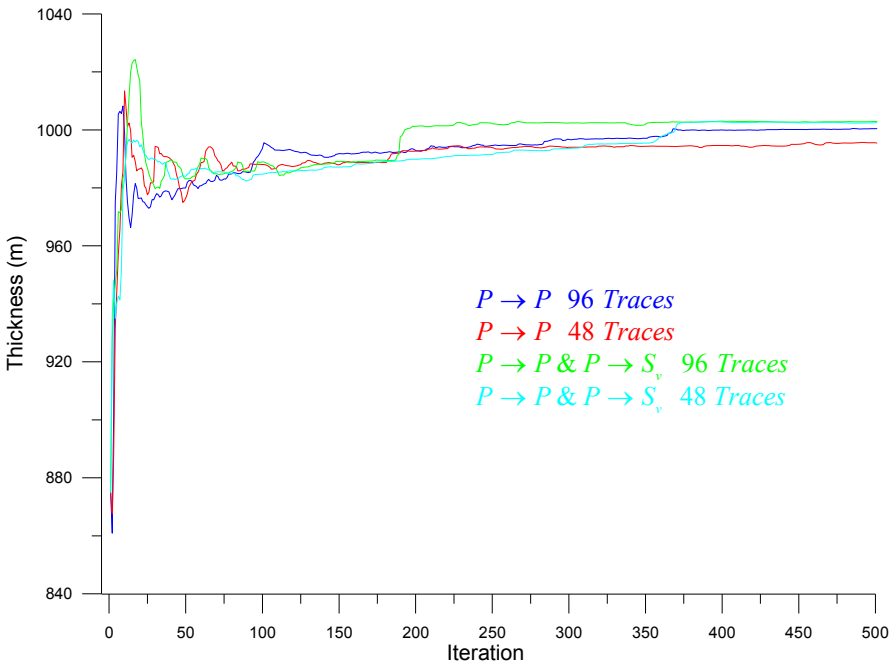


Figure 6.7: Convergence pattern of the thickness of layer 1 from inversion of the traveltime data. The true value is 1000 metres.

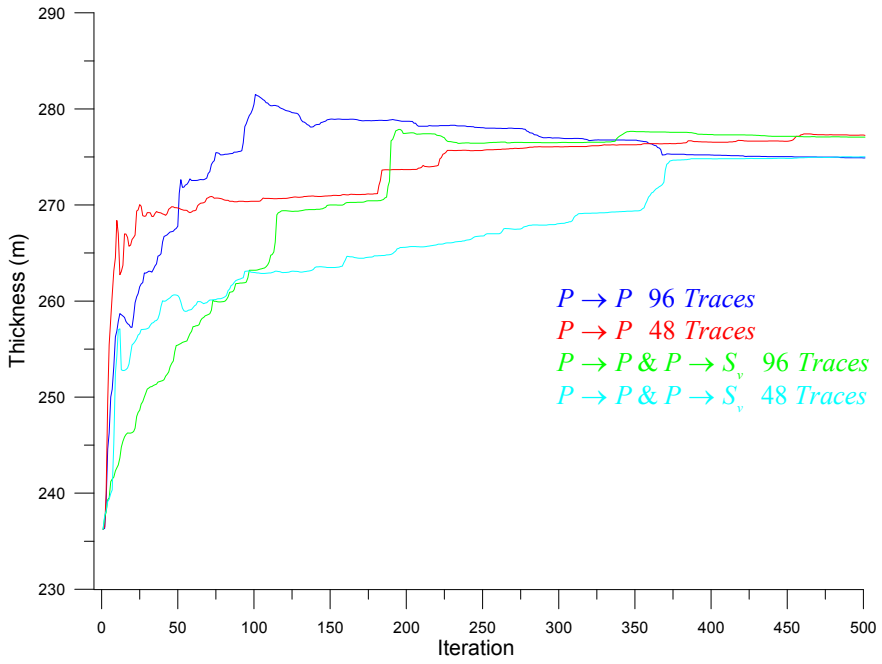


Figure 6.8: Convergence pattern of the thickness of the layer 7 from inversion of the traveltime data. The true value is 270 metres.

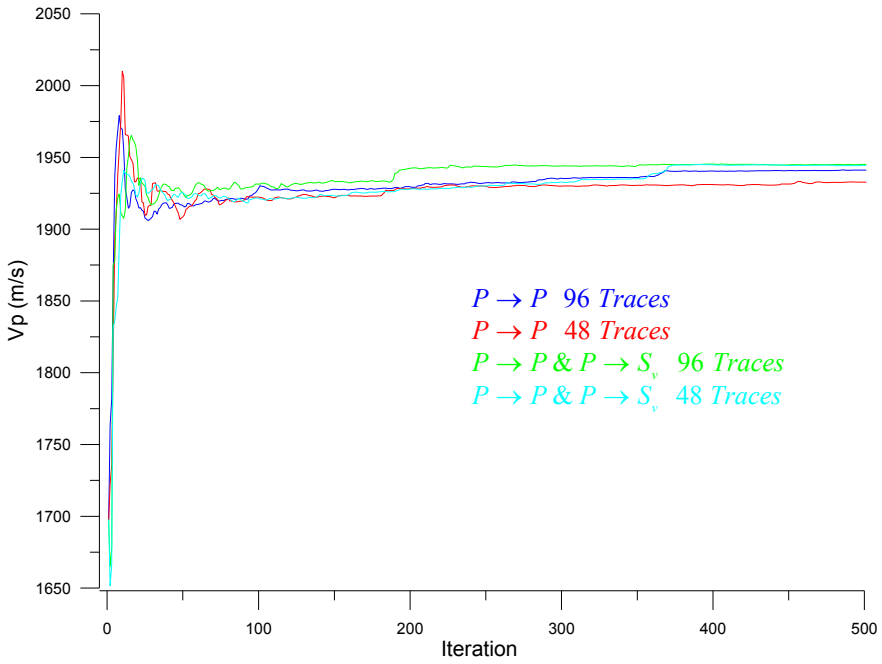


Figure 6.9: Convergence pattern of the P-wave velocity of layer 1 from inversion of the traveltime data. The true value is 1940 m/sec.

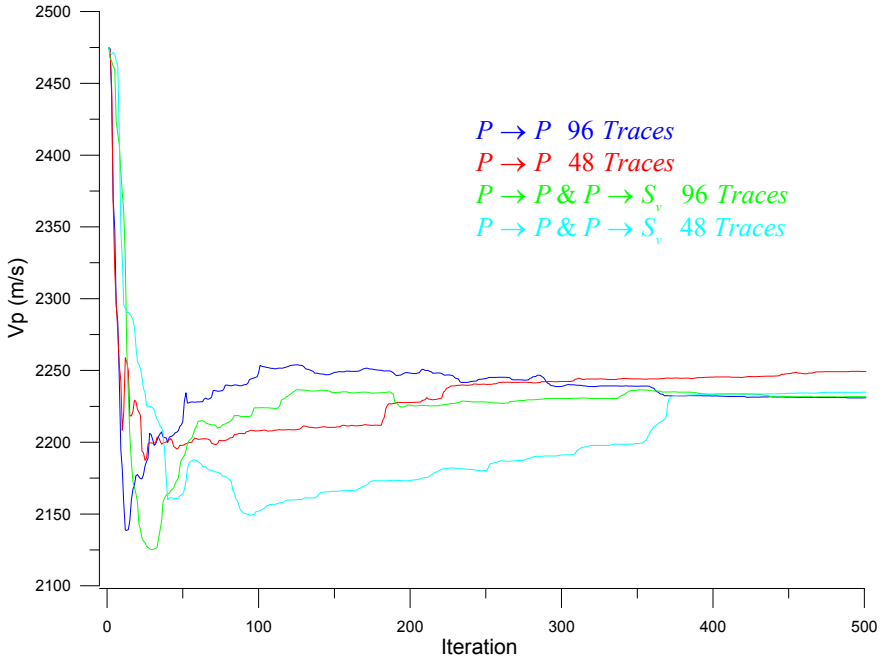


Figure 6.10: Convergence pattern of the P-wave velocity of layer 7 from inversion of the traveltime data. The true value is 2200 m/sec.

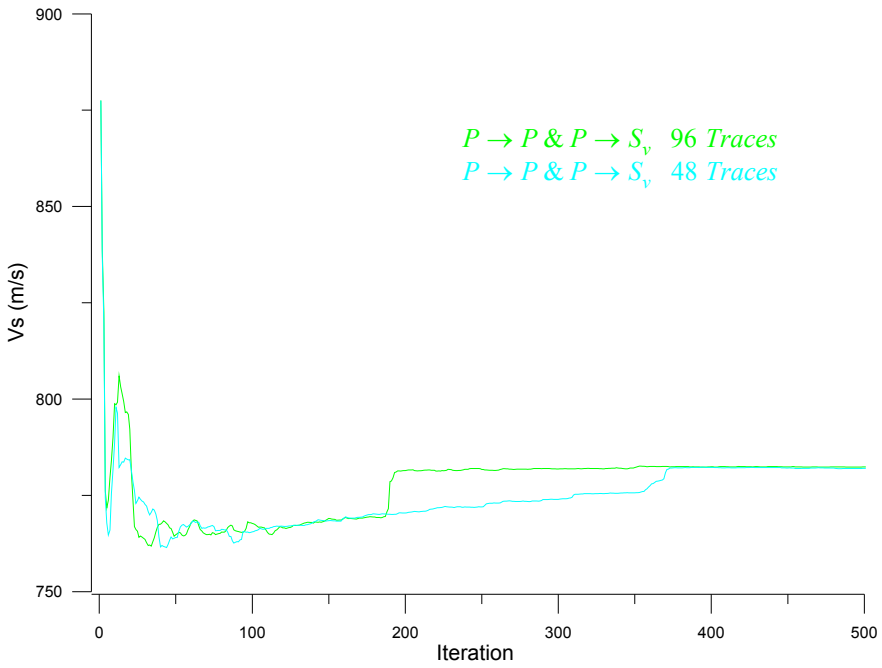


Figure 6.11: Convergence pattern of the S-wave velocity of layer 1 from inversion of the traveltime data. The true value is 780 m/sec.

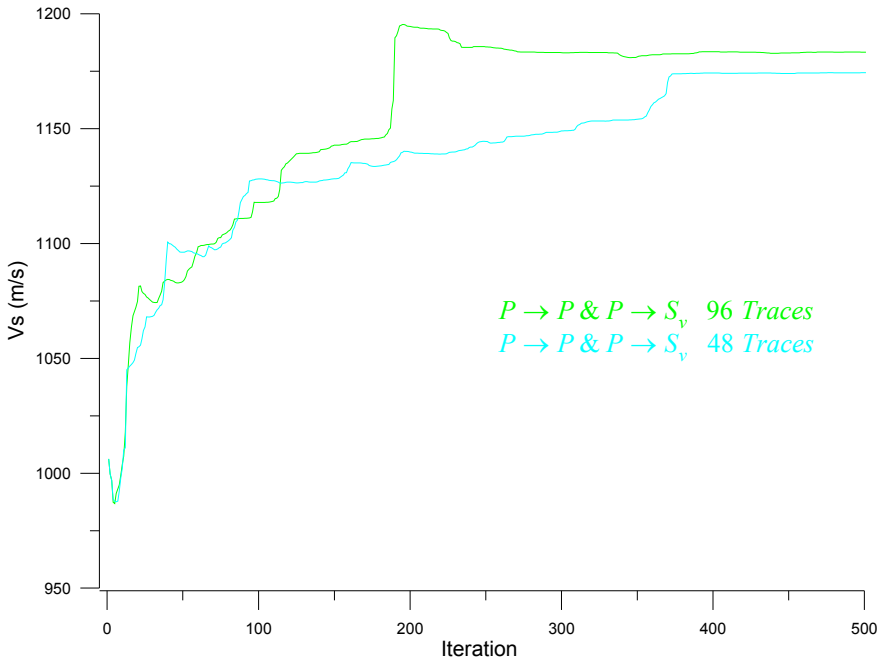


Figure 6.12: Convergence pattern of the S-wave velocity of layer 7 from inversion of the traveltime data. The true value is 1150 m/sec.

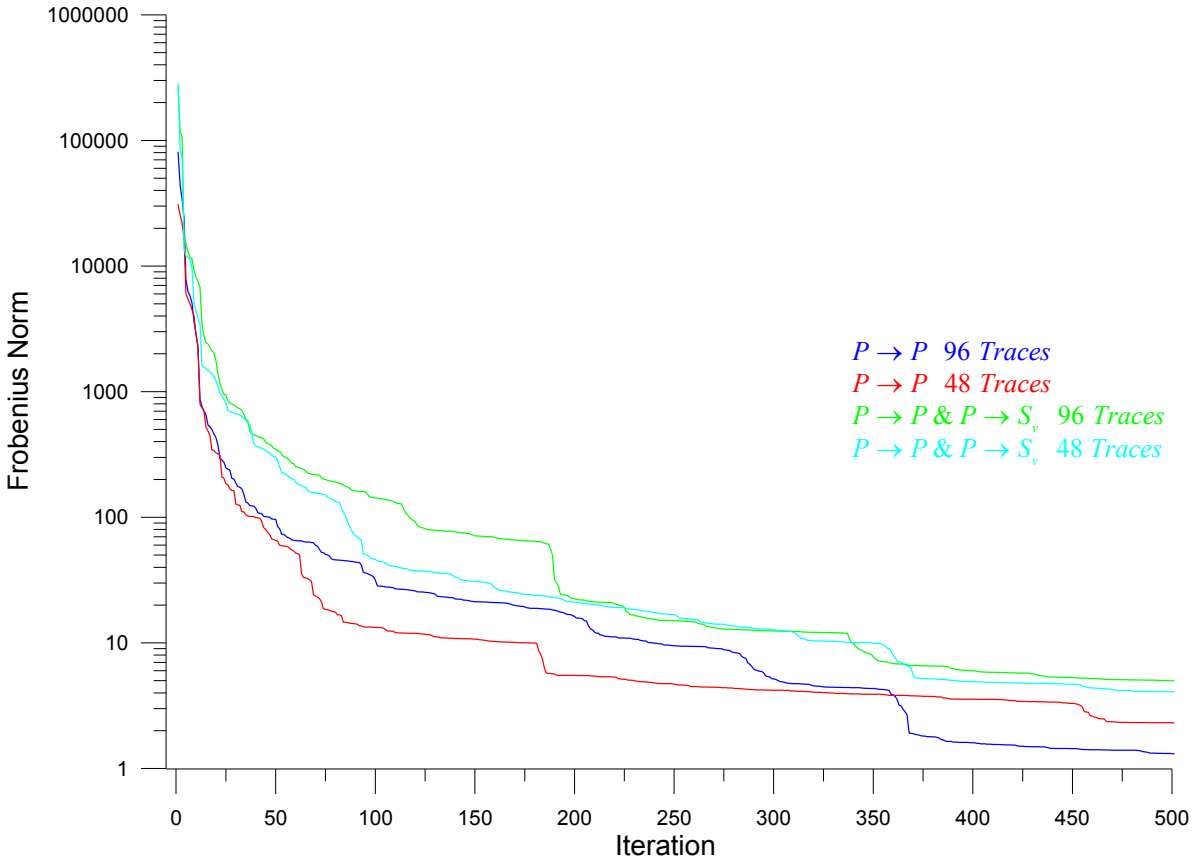


Figure 6.13: The Frobenius Norm (objective function) shows the convergence rate of the simultaneous traveltime inversion of vertical P-wave velocity, vertical S-wave velocity, thickness, and Thomsen's anisotropy parameters of a VTI media consist of 10 horizontal layers. The Frobenius norm is not scaled. Polak-Ribiére conjugate gradient algorithm without preconditioning has been used.

Layer	$Vp\left(m/\sec\right)$	$Vs\left(m/\sec\right)$	δ	ε	$\Delta z\left(metre\right)$
1	1940	780	0.03	0.1	1000
2	2140	860	-0.02	0.14	160
3	2220	890	-0.05	0.1	90
4	2000	1000	0.1	0.14	40
5	1990	990	0.05	0.1	100
6	1900	950	0.04	0.12	190
7	2200	1150	0.06	0.18	270
8	2050	1130	0.1	0.2	170
9	2650	1500	0.07	0.1	310

Table 6.3: True layer properties which have been used to create observed $P \to P$ and $P \to S_v$ traveltimes.

Layer	$Vp\left(m/\sec\right)$	$Vs\left(m/\sec\right)$	δ	ε	$\Delta z\left(metre\right)$
1	1698	878	0.0263	0.113	875
2	1873	753	-0.0172	0.158	180
3	2498	1001	-0.429	0.088	79
4	1750	875	0.0875	0.158	35
5	2239	1114	0.0438	0.113	113
6	2138	831	0.045	0.135	166
7	2475	1006	0.0675	0.158	236
8	1794	989	0.0875	0.225	149
9	2319	1313	0.0788	0.113	349

Table 6.4: The prior model which has been used in minimization of $P \to P$ and $P \to S_v$ traveltimes.

Layer	$Vp(m/\sec)$	$Vs(m/\sec)$	δ	ε	$\Delta z(metre)$
1	1941	877	0.0281	0.099	1000
2	2128	751	-0.017	0.149	159.
3	2249	1001	-0.0428	0.087	92
4	1872	875	0.0881	0.1587	35
5	2030	1113	0.0437	0.1082	104
6	1874	831	0.0450	0.1333	187
7	2231	1006	0.0677	0.1560	275
8	2021	991	0.0893	0.2332	167
9	2628	1312	0.0789	0.1113	308

Table 6.5: Estimated model parameters from minimization of the $P \to P$ for the last iteration in Figure 6.13. 96 traces have been used in the minimization.

Layer	$Vp(m/\sec)$	$Vs(m/\sec)$	δ	ε	$\Delta z(metre)$
1	1945	782	0.0284	0.097	1003
2	2118	856	-0.017	0.153	159
3	2257	893	-0.043	0.084	90
4	1840	873	0.088	0.157	34
5	2053	1059	0.044	0.107	108
6	1852	910	0.045	0.134	182
7	2232	1183	0.067	0.167	277
8	1969	1061	0.087	0.236	160
9	2635	1497	0.079	0.112	310

Table 6.6: Estimated model parameters from minimization of the joint $P \to P$ and $P \to S_v$ traveltimes for the last iteration in Figure 6.13. 96 traces have been used in minimization.

Layer	$Vp\left(m/\sec\right)$	$Vs\left(m/\sec\right)$	δ	ε	$\Delta z\left(metre\right)$
1	1945	782	0.0284	0.097	1003
2	2118	856	-0.017	0.153	159
3	2257	893	-0.043	0.084	90
4	1840	873	0.088	0.157	34
5	2053	1059	0.044	0.107	108
6	1852	910	0.045	0.134	182
7	2232	1183	0.067	0.167	277
8	1969	1060	0.087	0.236	160
9	2635	1497	0.079	0.112	310

Table 6.7: Estimated model parameters from minimization of the $P \to P$ for the last iteration in Figure 6.13. 48 traces have been used in minimization.

Layer	$Vp\left(m/\sec\right)$	$Vs\left(m/\sec\right)$	δ	ε	$\Delta z\left(metre\right)$
1	1944	782	0.028	0.097	1002
2	2115	846	-0.017	0.154	158
3	2241	902	-0.043	0.084	91
4	1777	875	0.088	0.157	34
5	2158	1086	0.045	0.110	111
6	1858	926	0.045	0.134	185
7	2235	1174	0.067	0.164	275
8	1871	1020	0.088	0.229	153
9	2669	1513	0.079	0.1137	314

Table 6.8: Estimated model parameters from minimization of the joint $P \to P$ and $P \to S_v$ traveltimes for the last iteration in Figure 6.13. 48 traces have been used in minimization.

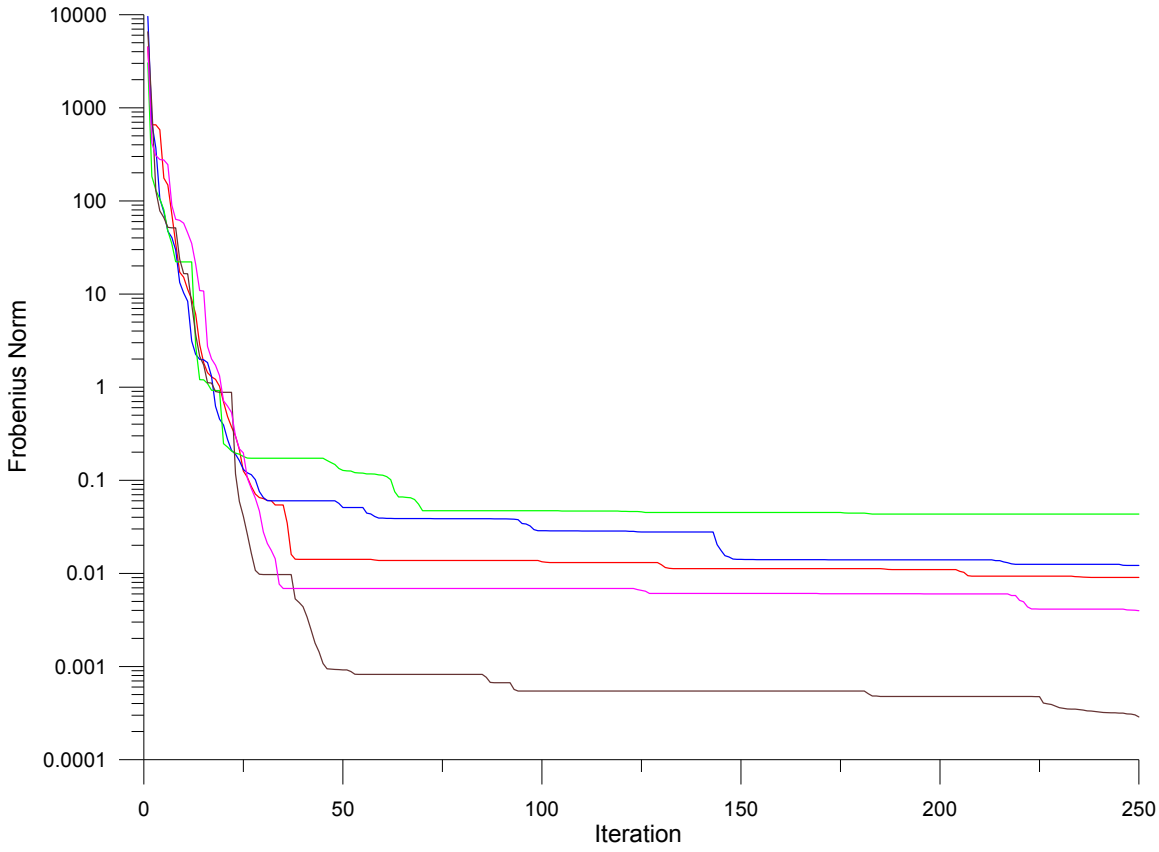


Figure 6.14: The Frobenius Norm (objective function) shows the convergence rate of the five realizations with different random prior model of simultaneous $P \to P$ traveltime inversion of Thomsen's anisotropy parameters of a VTI medium consist of 10 horizontal layers. 96 traces have been used in each realization. Preconditioning Polak-Ribiére conjugate gradient algorithm has been used. The color code is the same in the Figures 6.15-6.20. See Table 6.9 to relate each curve to a specific prior model.

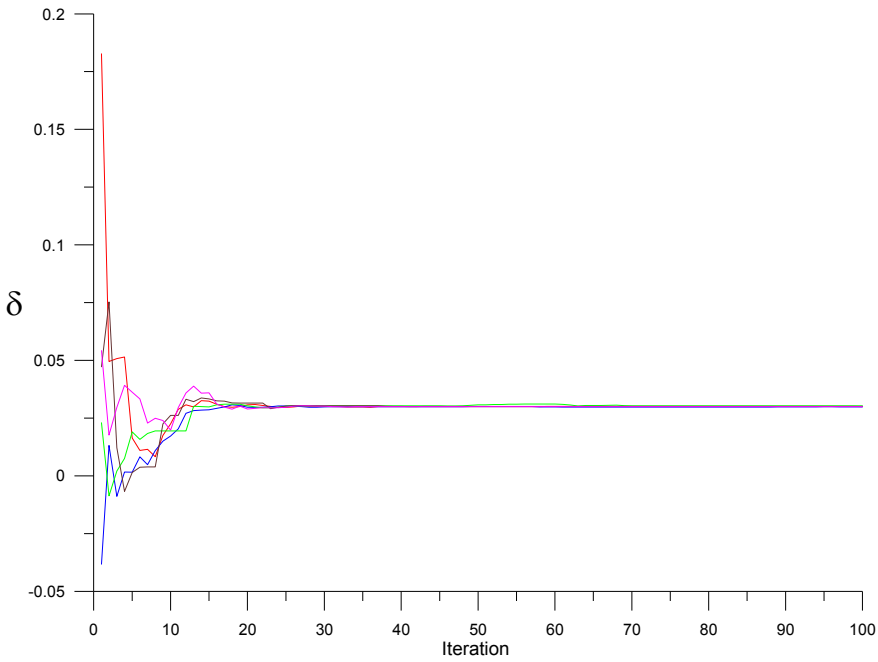


Figure 6.15: Convergence pattern of the δ for layer 1 from five realizations. A preconditioning conjugate gradient algorithm has been used for simultaneous inversion of Thomsen's anisotropy parameters from a system of 10 layers. The true value is 0.03.

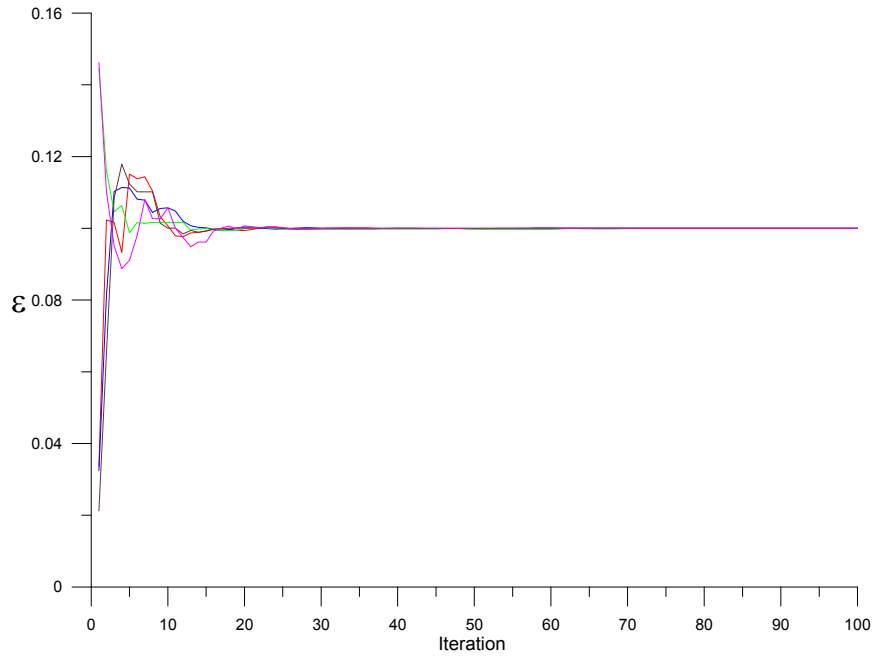


Figure 6.16: Convergence pattern of the ε for layer 1 from five realizations. A preconditioning conjugate gradient algorithm has been used for simultaneous inversion of Thomsen's anisotropy parameters from a system of 10 layers. The true value is 0.1.

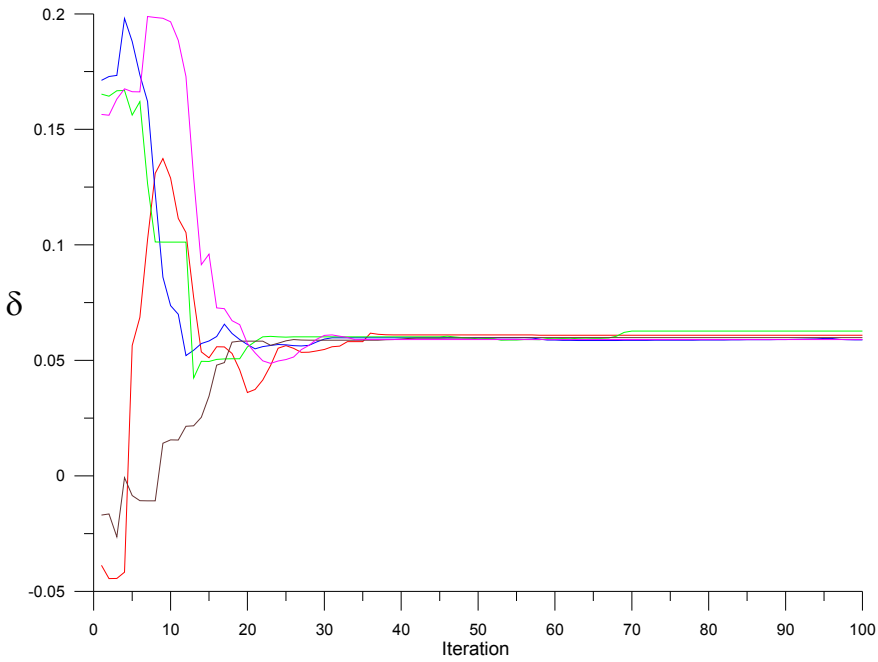


Figure 6.17: Convergence pattern of the δ for layer 7 from five realizations. A preconditioning conjugate gradient algorithm has been used for simultaneous inversion of Thomsen's anisotropy parameters from a system of 10 layers. The true value is 0.06.

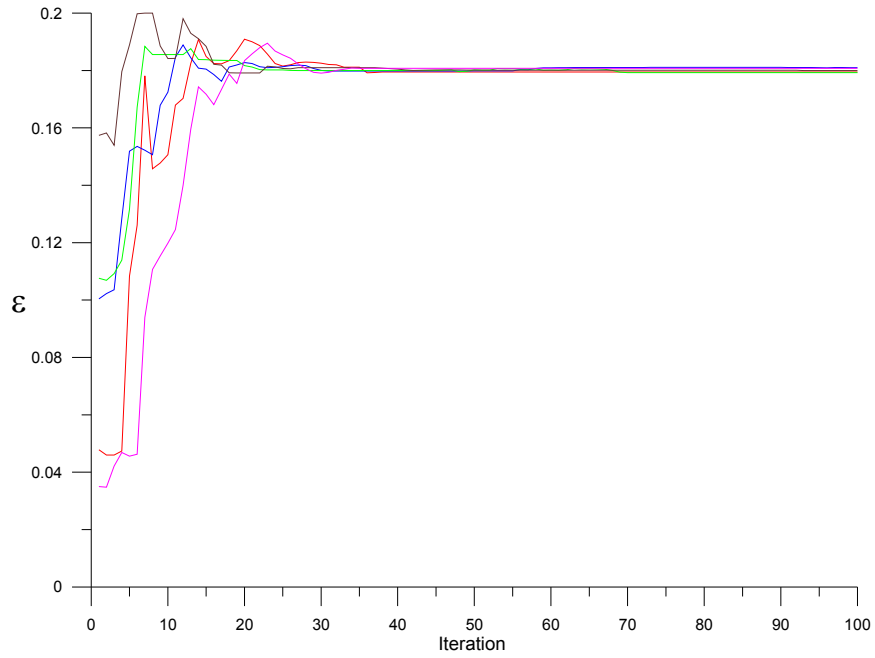


Figure 6.18: Convergence pattern of the ε for layer 7 from five realizations. A preconditioning conjugate gradient algorithm has been used for simultaneous inversion of Thomsen's anisotropy parameters from a system of 10 layers. The true value is 0.18.

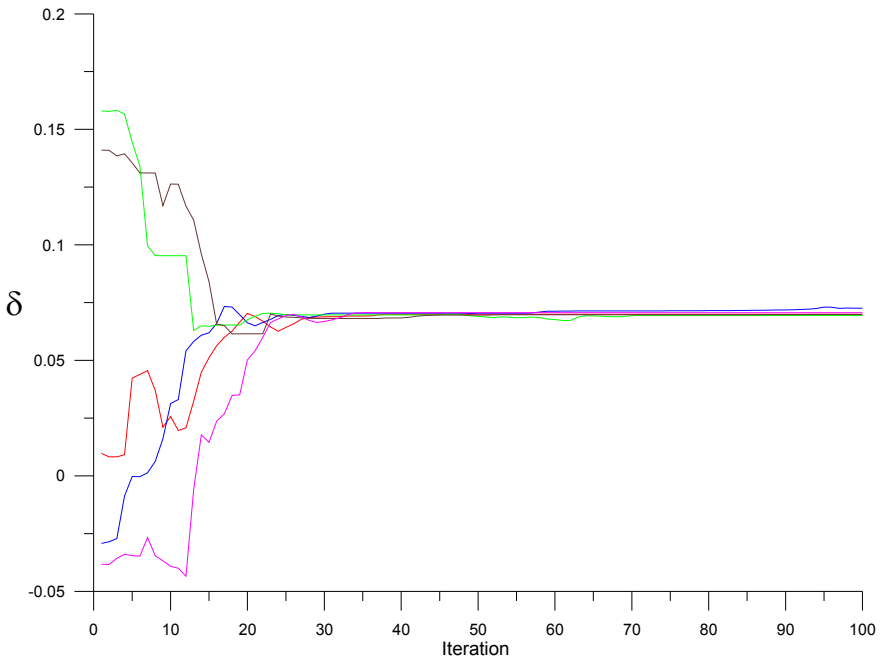


Figure 6.19: Convergence pattern of the δ for layer 9 from five realizations. A preconditioning conjugate gradient algorithm has been used for simultaneous inversion of Thomsen's anisotropy parameters from a system of 10 layers. The true value is 0.07.

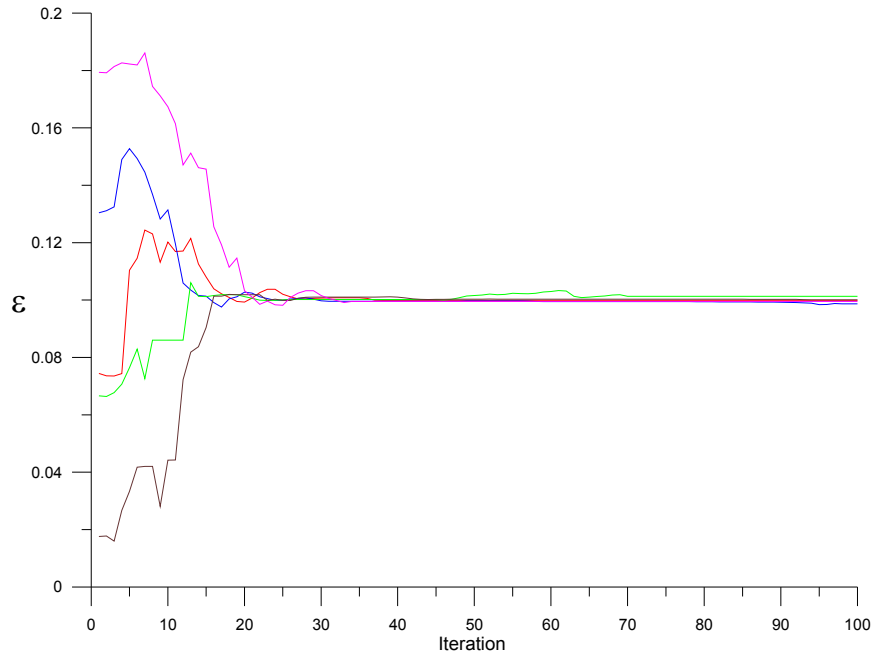


Figure 6.20: Convergence pattern of the ε for layer 9 from five realizations. A preconditioning conjugate gradient algorithm has been used for simultaneous inversion of Thomsen's anisotropy parameters from a system of 10 layers. The true value is 0.1.

Layer	δ	ε	δ	ε	δ	ε	δ	ε	δ	ε
1	0.18275	0.0336	-0.03825	0.0324	0.023	0.1446	0.04725	0.0212	0.05425	0.1462
2	-0.0255	0.1322	0.166	0.1764	-0.01575	0.1784	0.11525	0.1316	-0.02975	0.1326
3	0.16875	0.069	0.099	0.1738	0.05725	0.0872	0.13375	0.0144	0.017	0.1612
4	0.07725	0.1914	-0.04225	0.1086	-0.005	0.192	0.13775	0.119	0.07725	0.1566
5	0.055	0.1792	-0.0245	0.07	0.15075	0.1014	0.1845	0.0758	0.11575	0.1368
6	0.143	0.07	0.07975	0.0688	0.0785	0.1682	0.04075	0.03	-0.02275	0.0414
7	-0.03875	0.0478	0.17125	0.1004	0.16525	0.1076	-0.017	0.1574	0.1565	0.035
8	0.0525	0.0384	-0.03125	0.0844	-0.037	0.101	0.19625	0.1862	0.0205	0.0702
9	0.00975	0.0744	-0.02925	0.1304	0.158	0.0666	0.141	0.0176	-0.03825	0.1794
1	0.029943	0.100019	0.029772	0.100057	0.030297	0.099893	0.030031	0.099992	0.029885	0.100021
2	-0.01967	0.139997	-0.01785	0.139809	-0.02132	0.140057	-0.02037	0.140042	-0.01927	0.139963
3	-0.04598	0.09885	-0.04985	0.10014	-0.04258	0.097168	-0.04955	0.099889	-0.04975	0.099917
4	0.073966	0.151453	0.077749	0.146523	0.0633	0.156201	0.100753	0.139782	0.098205	0.140593
5	0.056008	0.096447	0.057374	0.096331	0.063294	0.093521	0.049244	0.100498	0.049909	0.100191
6	0.03973	0.11987	0.040356	0.120017	0.033693	0.124511	0.040531	0.11947	0.040615	0.119353
7	0.060445	0.179536	0.059608	0.180018	0.062754	0.179131	0.059819	0.180047	0.059901	0.179959
8	0.100056	0.2	0.099862	0.2	0.096489	0.2	0.100259	0.199839	0.099753	0.2
9	0.069676	0.100227	0.070379	0.100023	0.069195	0.101462	0.0699	0.100024	0.069996	0.100057

Table 6.9: The first two columns correspond to red color in figure 6.8, the second two columns to blue color, the third two columns to green color, the fourth two columns to brown, and the last two columns to pink color.The first set of rows correspond to prior model and the second set to the estimated model in the last iteration

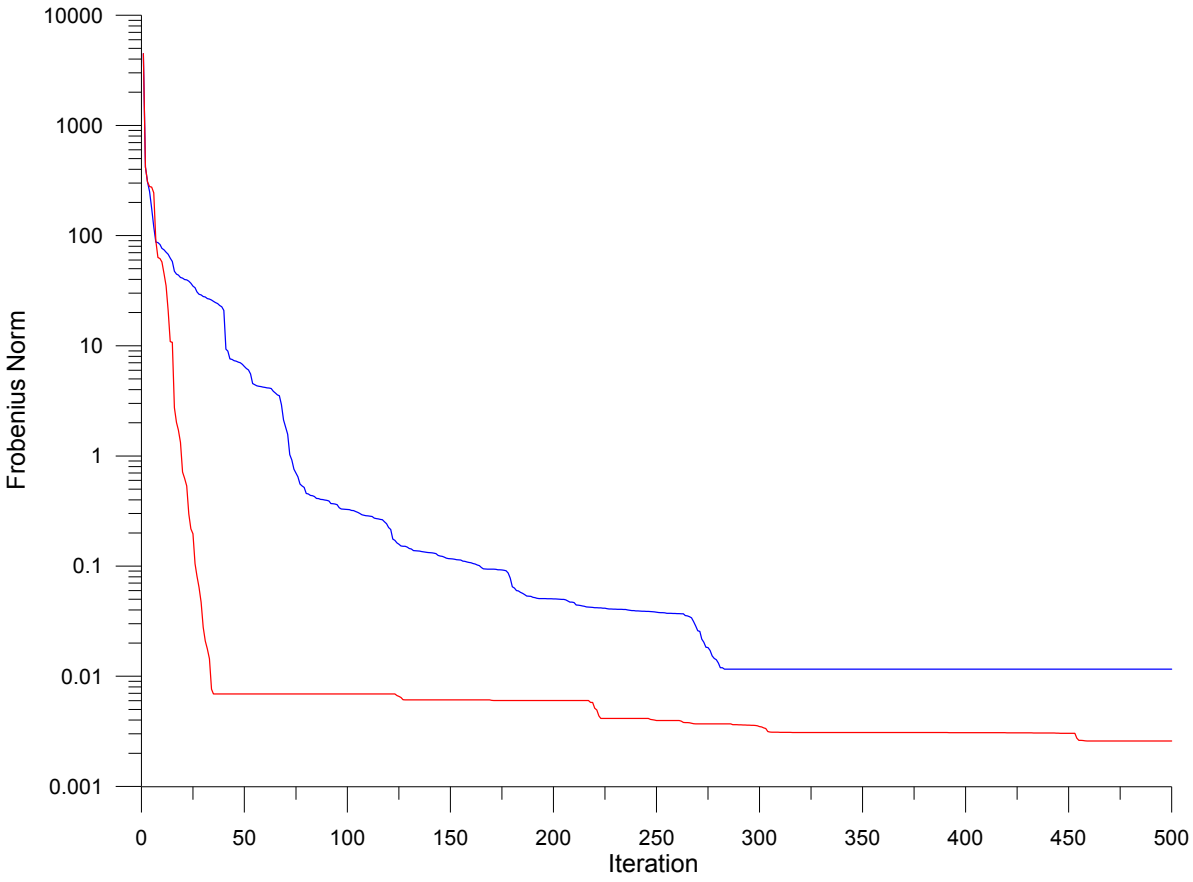


Figure 6.21: This picture illustrates the effect of the preconditioning on the convergence rate in the conjugate gradient algorithm. Thomsen's anisotropy parameters from a system of 9 horizontal layers have been optimized simultaneously by minimizing $P \to P$ traveltime from 96 traces. Both inversions use the same prior model. The blue curve illustrates the Polak-Ribiére conjugate gradient algorithm while, red curve shows the results from the preconditioned Polak-Ribiére conjugate gradient algorithm.

6.2 Waveform inversion in horizontally layered VTI media

Prestack amplitude based inversion methods suffer from problems such as loss of true amplitude during the processing, out of plane interference, noise, an incorrect wavelet, and intensive computational costs, however, they have been promising in estimation of the high frequency part of the structures where layer thickness is small. That is why traveltime data are usually used in estimation of the low frequency part of the structure. The other benefit of waveform inversion is the fact that it takes the bottom layer of the model into account, whereas traveltime inversion does not.

In this section we still take the geometry of previous sections, however, only the first 48 traces will be used in parameter estimation. This will keep the phase incidence angles below the critical angle, where Rüger's equations (Rüger, 1996) are valid. We will not consider geometrical spreading effects as the seismic data are usually corrected for this. Implementing the geometrical spreading is straightforward and readers may refer to Ursin and Hokstad (2003). We also assume that data have been corrected for multiples, but we will consider the shear converted waves and use them in parameter estimation.

The objective function of a waveform inversion problem (equation 5.39) in terms of velocity and thickness is highly nonlinear and less nonlinear in terms of Thomsen's anisotropy parameters and density. We can still use prestack waveform inversion to estimate the velocity and thickness but the prior information needs to be close enough to the solution. Apart from nonlinearity, the parameter model space contains many local minima and it is likely to converge to a local minimum rather than the global minimum. A quick evaluation of the objective function (equation 5.39) for two high impact parameters such as P-wave velocity for the first and second layers simultaneously, will show the nonlinearity and frequency of local minima. For the sake of utilizing the high frequency estimating power of prestack waveform inversion, we assume a fairly well velocity model using for example traveltime inversion is available and estimate merely Thomsen's anisotropy parameters (Nadri and Hartley, 2007b).

In the first example, we draw a random prior model for δ and ε from a uniform distribution constrained within ± 50 percent of the true model (the second two columns of Table 6.10). This will be used for both $P \to P$ and joint $P \to P \,\&\, P \to S_v$ inversion. We

also draw a random prior model from a wider uniform distribution constrained by ± 75 percent of the true model (the last two columns of the Table 6.10). We only use this prior model for inversion of $P \to P$ data. Due to high nonlinearity of the objective function corresponding to joint $P \to P \,\&\, P \to S_v$, it was not possible to converge to the solution with the latter prior model. We used the Polak-Ribiére conjugate gradient algorithm to minimize the objective function. Figure 6.22 illustrates the convergence pattern of the objective function for $P \to P$ (blue curve), joint $P \to P \,\&\, P \to S_v$ (red curve), and $P \to P$ data with a broader prior model (green curve). The blue curve has a better convergence because it used a closer prior model compared to the green curve. Minimization continued for up to 500 iterations.

The convergence patterns of Thomsen's anisotropy parameters δ and ε for the top layer are illustrated in Figure 6.23 and Figure 6.24 respectively. All the inversion schemes have resolved the model parameters well. We only plotted the first 250 iterations for a better view. It is clear that the green curves are farther from the solutions. The estimated Thomsen's anisotropy parameters for the last iteration are given in table 6.11 for different inversion schemes.

In Figures 6.25 and 6.26 we have illustrated the convergence pattern of δ and ε in layer 4. The thickness of this layer is 40 metres and the waveform from the bottom of this layer overlaps the tail of the waveform from the upper reflector (Figure 3.9). Practically, picking the traveltimes for this event is not very reliable, however, waveform inversion was able to resolve the layer properties well.

Figure 6.27 illustrates the convergence pattern of δ from layer 9. The overall convergence rate is slow and even slower for waveform inversion of $P \to P$ waves, however, they have all converged to the true solution. In Figure 6.28 the convergence rate of ε from layer 8 is illustrated. The prior model for $P \to P$ (green curve) is very far from the solution, however, it has been able to converge rapidly to true solution.

To demonstrate the impact of anisotropy on seismic amplitudes, we have computed the $P \to P$ wavefield in a prestack gather from the true model and a prior model given in the first two columns and second two columns of Table 6.10 respectively. The $P \to P$ wavefield from the true model is shown in the left panel of Figure 6.29. We

also computed the $P \to P \,\&\, P \to S_v$ wavefield in a prestack gather from the true model and a different prior model given in last two columns of Table 6.10. The $P \to P \,\&\, P \to S_v$ wavefield from the true model is also shown in the left panel of Figure 6.30. We have subtracted the $P \to P$ synthetic seismograms from the true model and prior model and have shown the resultant seismogram in the middle panel of Figure 6.29. We call the subtracted seismogram the seismic residual gather. It can be seen that the amount of the residuals is quite significant. The seismic residual gather clearly shows the effect of anisotropy on the seismic wavefield. The same procedure has been applied to $P \to P \,\&\, P \to S_v$ seismograms and the residual gather is shown in the middle panel of Figure 6.30. It is clear that the amount of residual is quite large especially at the larger offsets. To demonstrate the efficiency of the minimization algorithm in suppressing the seismic residuals we have computed the synthetic seismogram from the estimated model parameters in the last iteration of $P \to P$ waveform inversion given in the last two columns of table 6.11. We subtracted the above synthetic seismogram from the true corresponding $P \to P$ seismogram and illustrated the residual seismogram in the right panel of Figure 6.29. All panels have been shown with the same gain. It is clear that the minimization has been quite effective and almost no residual is left. The same procedure has been used for $P \to P \,\&\, P \to S_v$ wavefield. We have computed the synthetic seismogram from the estimated model parameters in the last iteration given in the second two columns of Table 6.11. The seismic residual for the $P \to P \,\&\, P \to S_v$ wavefield is also illustrated in the right panel of Figure 6.30. Same as the Figure 6.29 all gathers have been plotted with the same amplitude scaling factor. As it can be seen, minimization is quite effective and almost no residual is left.

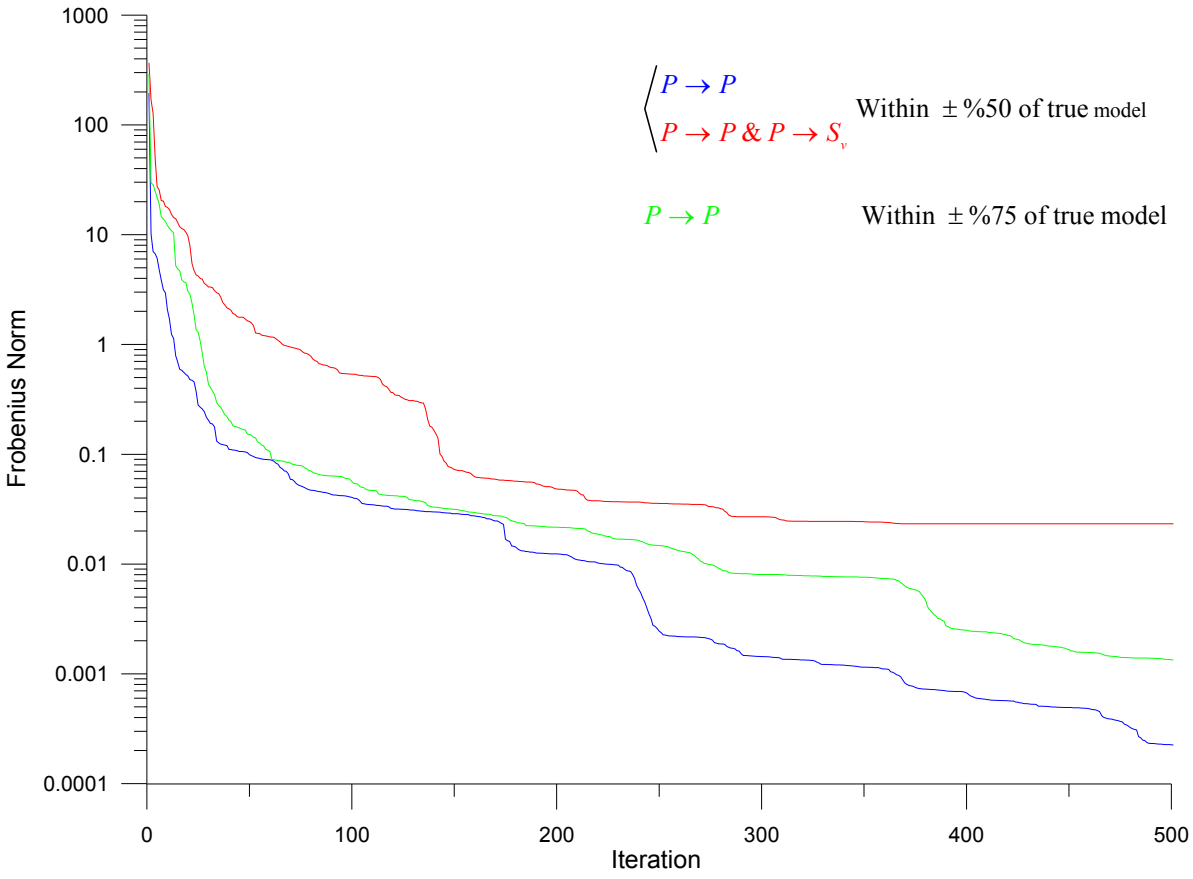


Figure 6.22: Convergence rate of the simultaneous $P \rightarrow P$ (blue and green curves) and joint $P \rightarrow P \,\&\, P \rightarrow S_v$ (red curve) waveform inversion of Thomsen's anisotropy parameters. Both red and blue curves used the same prior model drawn from a prior model within the $\pm\%50$ of the true model. The green curve has been drawn from a prior model within the $\pm\%75$ of the true model. A Polak-Ribiére conjugate gradient algorithm has been used to simultaneously minimize Thomsen's anisotropy parameters from a system of 10 horizontal layers. All prior models have been randomly selected from a uniform distribution constrained to an upper and lower limit. Only the first 48 traces are used in minimization.

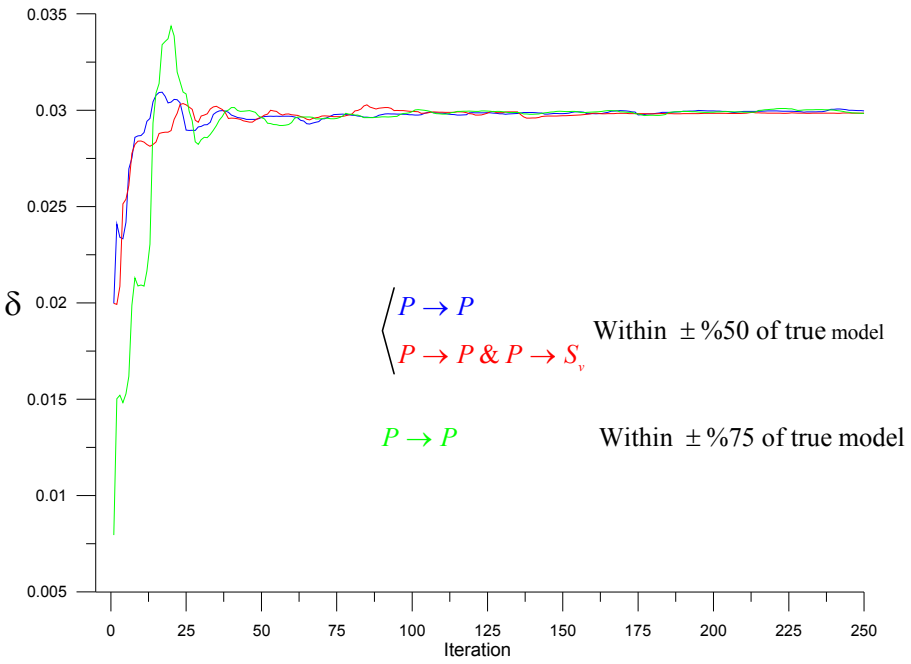


Figure 6.23: Convergence pattern of the δ for layer 1 from waveform inversion. The true value is 0.03.

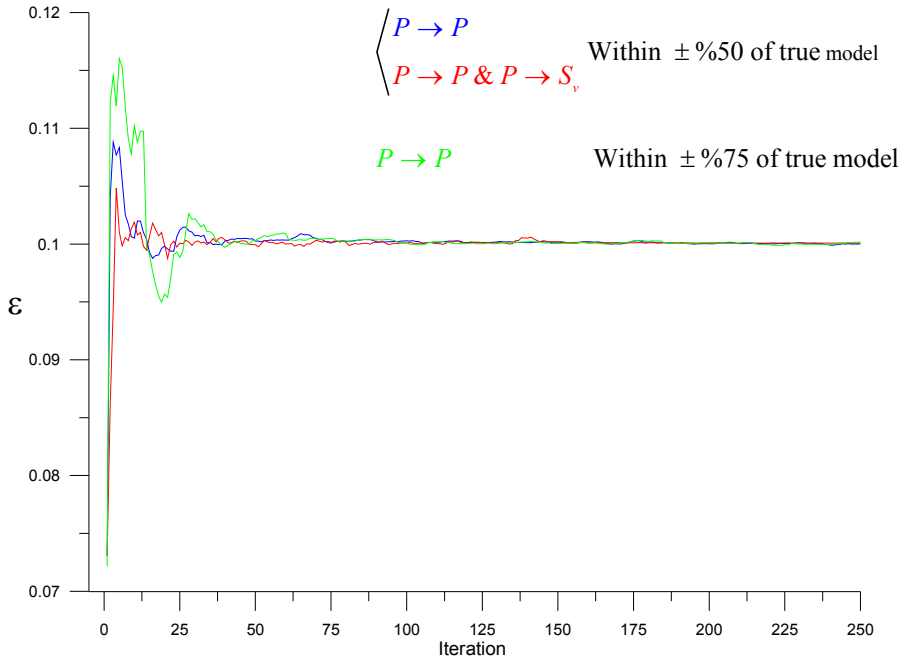


Figure 6.24: Convergence pattern of the ε for layer 1 from waveform inversion. The true value is 0.1.

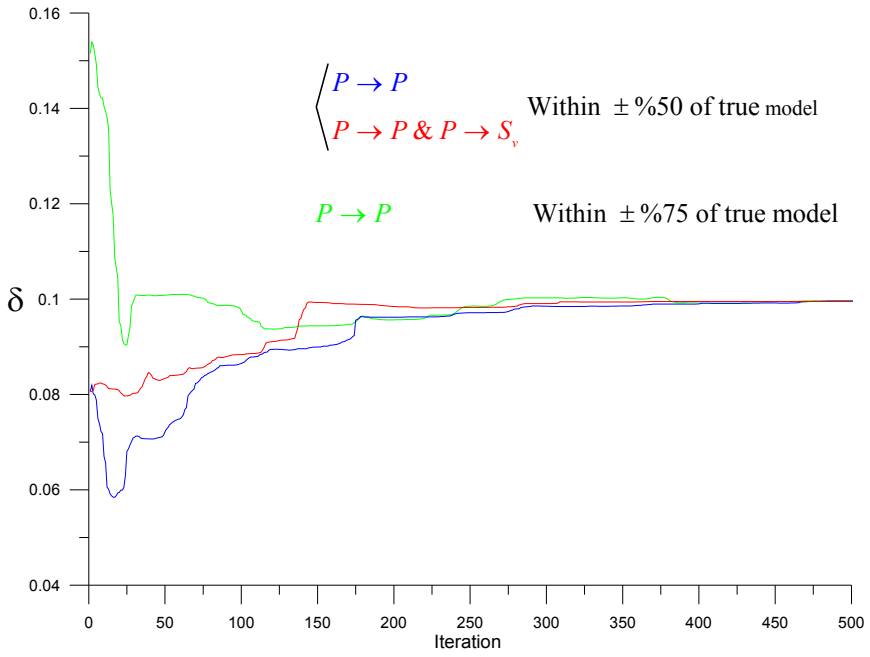


Figure 6.25: Convergence pattern of the δ for layer 4 from waveform inversion. The true value is 0.1.

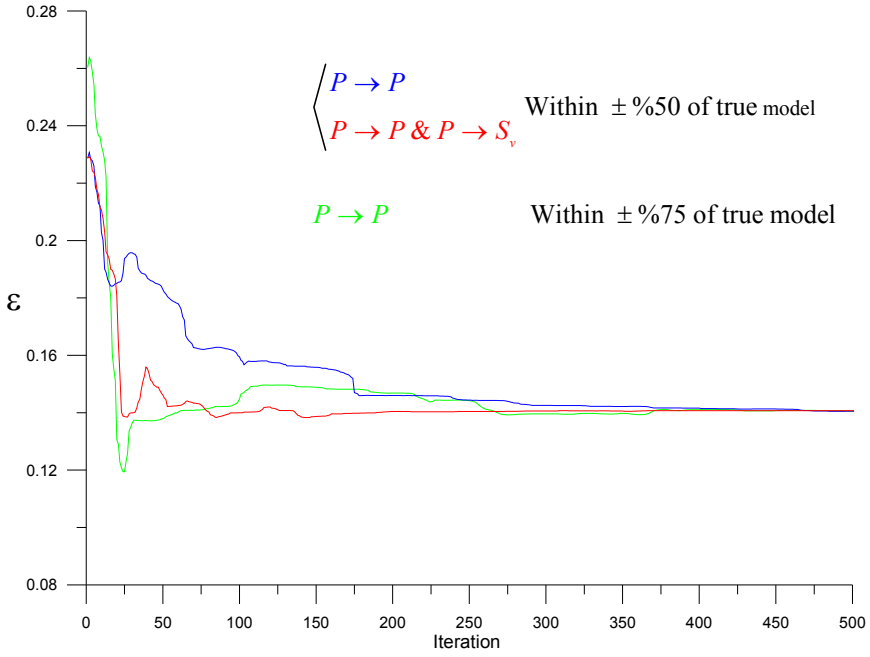


Figure 6.26: Convergence pattern of the ε for layer 4 from waveform inversion. The true value is 0.14.

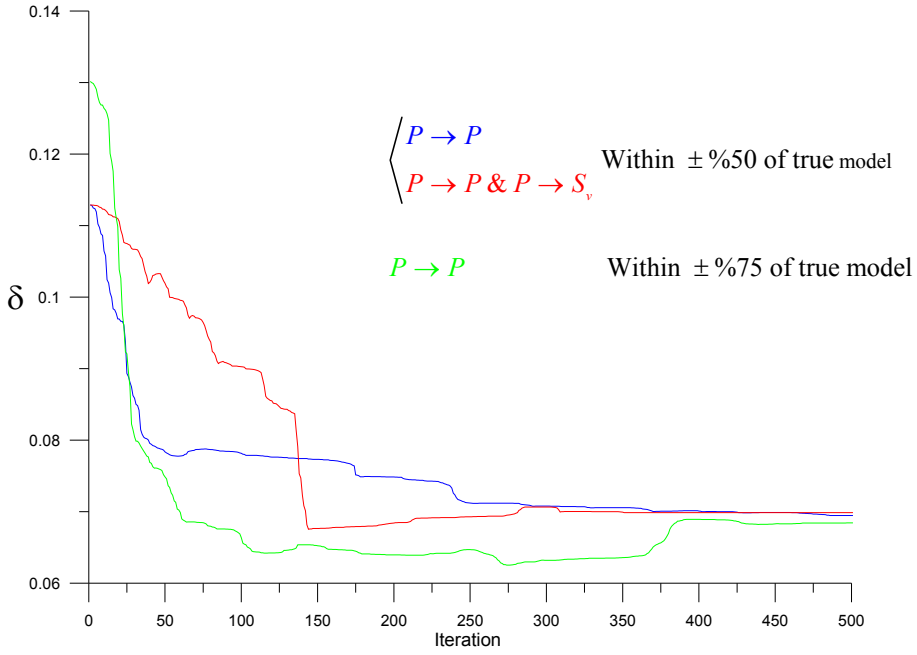


Figure 6.27: Convergence pattern of the δ for layer 9 from waveform inversion. The true value is 0.07.

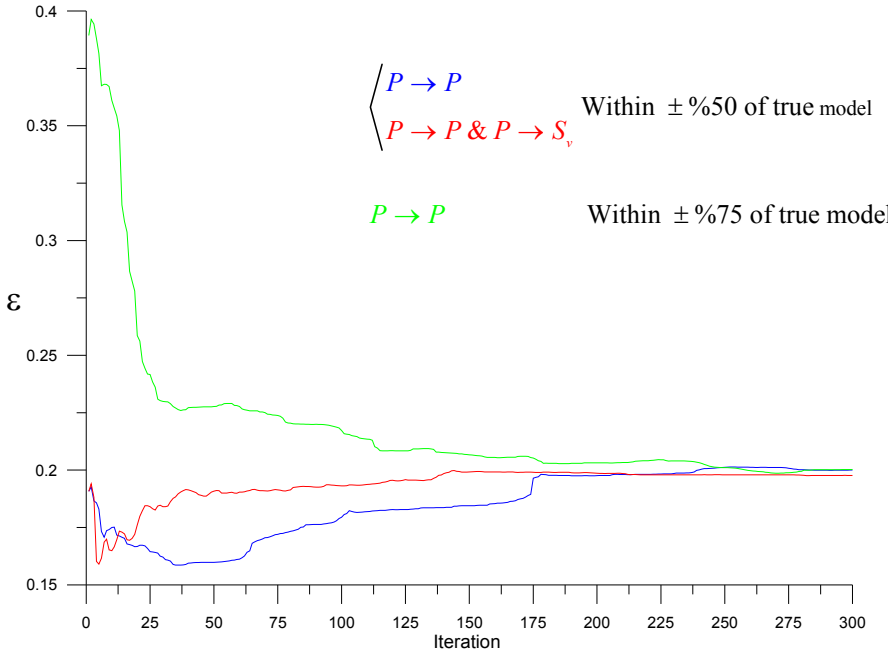


Figure 6.28: Convergence pattern of the ε for layer 8 from waveform inversion. The true value is 0.2.

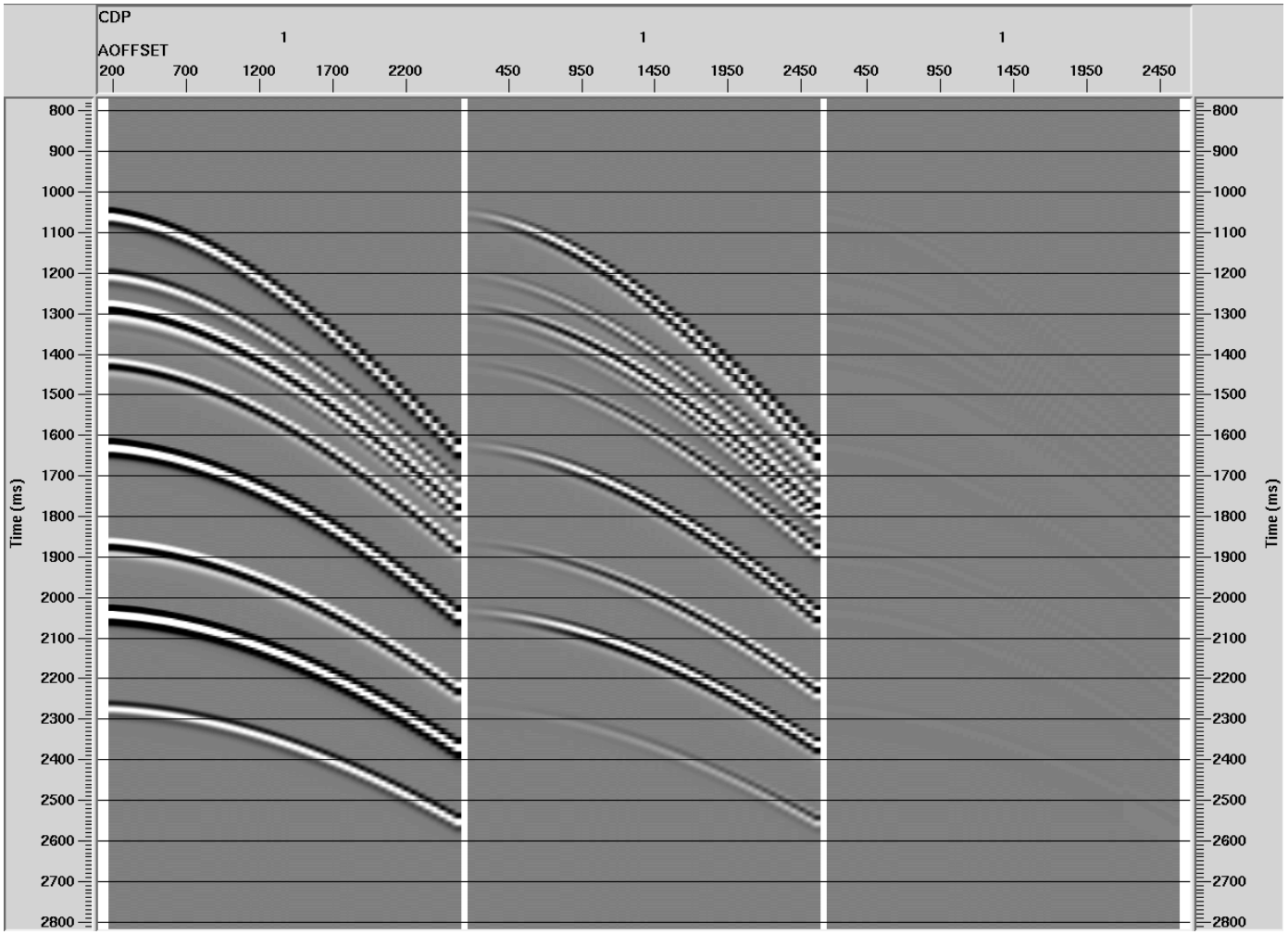


Figure 6:29: Left panel: compressional wavefield of true model (Table 2.3). Middle panel: difference between true model and a random prior model (Table 6.10). This is the prior residual has to be minimized. Right panel: the difference between true model and final model. All three panels have been shown with the same scale. The inversion scheme is explained in Figure 6.22.

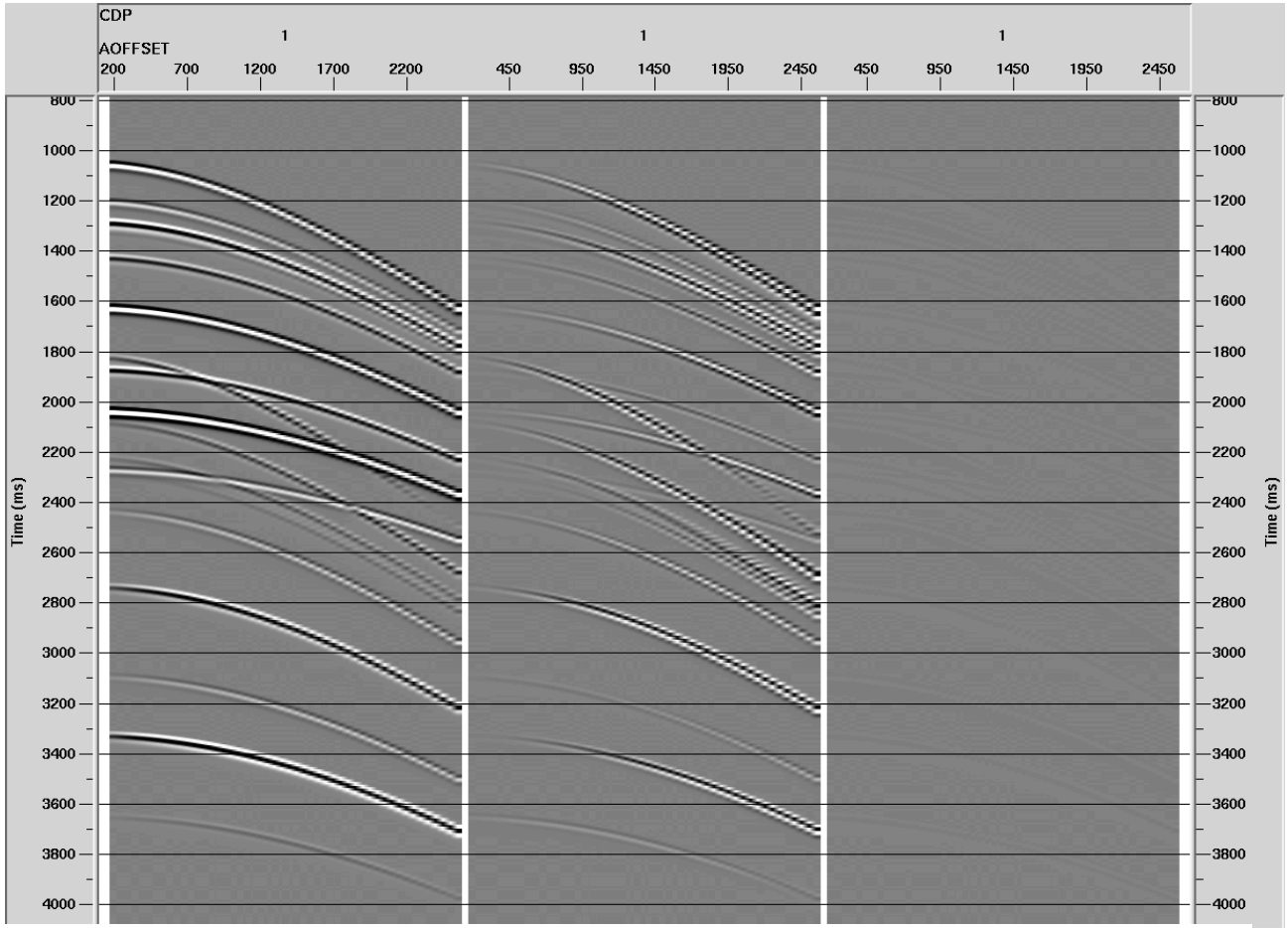


Figure 6:30: Left panel: compressional and converted shear wavefield of true model (Table 2.3). Middle panel: difference between true model and a random prior model (Table 6.10). This is the prior residual which has to be minimized. Right panel: the difference between true model and final model. All three panels have been shown with the same scale. The inversion scheme is explained in Figure 6.22.

Layer	δ	ε	δ	ε	δ	ε
1	0.03	0.1	0.02	0.073	0.0079	0.072
2	-0.02	0.14	-0.014	0.130	-0.0133	0.187
3	-0.05	0.1	-0.055	0.053	-0.053	0.128
4	0.1	0.14	0.0806	0.229	0.1516	0.260
5	0.05	0.1	0.04	0.171	0.0459	0.201
6	0.04	0.12	0.063	0.12	0.0385	0.053
7	0.06	0.18	0.046	0.226	0.0328	0.191
8	0.1	0.2	0.122	0.191	0.1135	0.389
9	0.07	0.1	0.113	0.099	0.130	0.160
10	0.1	0.14	0.072	0.169	0.063	0.331

Table 6.10: The first two columns correspond to the true model, the next two columns to the prior model which has been used in waveform inversion of $P \to P$ and joint $P \to P \,\&\, P \to S_v$ from 48 seismic traces as in Figures 6.22-6.26. The last two columns correspond to prior model which has been used in waveform inversion of $P \to P$ from 48 seismic traces as in Figures 6.22-6.26. This prior model is further from the true model parameters than the one in first two columns.

Layer	δ	ε	δ	ε	δ	ε
1	0.03	0.100	0.03	0.100	0.030	0.1
2	-0.0197	0.14	-0.02	0.14	-0.02	0.14
3	-0.05	0.10	-0.05	0.100	-0.05	0.1
4	0.1	0.14	0.099	0.140	0.1	0.14
5	0.049	0.101	0.05	0.101	0.048	0.103
6	0.040	0.119	0.04	0.12	0.04	0.118
7	0.060	0.18	0.06	0.18	0.06	0.179
8	0.1	0.20	0.1	0.2	0.1	0.201
9	0.07	0.101	0.07	0.11	0.068	0.104
10	0.095	0.157	0.10	0.15	0.087	0.183

Table 6.11: The first two columns correspond to estimated model parameters from the last iteration (Figure 6.22) of the waveform inversion of $P \to P$ and the second two columns to joint $P \to P \,\&\, P \to S_v$ data. The last two columns correspond to waveform inversion of $P \to P$ data as in Figures 6.22-6.26 where the prior model is further from the true model parameters than the one in the first two columns of Table 6.10.

6.3 Traveltime inversion in a dipping layered TI media

In Chapter 2 we gave a ray tracing algorithm for dipping layers in a TI medium and in Chapter 5 we discussed unconstrained minimization algorithms in dipping media. The minimization algorithm is the same as for horizontal layers but more computationally intensive. In Chapter 2, non-hyperbolic behavior of $P \to P$ and $P \to S_v$ traveltimes and violation of the law of conservation of horizontal slowness were discussed. These issues make the computation of traveltimes and partial derivatives very intensive. In contrast to inversion of traveltimes in horizontally layered media in which we could go for a higher number of iterations, the computational cost in dipping media makes it difficult to do large number of iterations. A preconditioning conjugate gradient can substantially decrease the number of iterations. In the dipping media, we are mainly concerned with the estimation of Thomsen's anisotropy parameters, but instead we extend the prior information knowledge to the weak anisotropy assumption. We have shown, by the assumption of weak anisotropy, that Thomsen's anisotropy parameters can be effectively estimated without the knowledge of a prior model (Nadri and Hartley, 2008).

We use Table 2.2 to compute the traveltimes for $P \to P$ and $P \to S_v$ waves. These traveltimes will be considered as the measured traveltimes during the minimization. The first two columns of Table 6.12 show the prior model parameters for both $P \to P$ and joint $P \to P \,\&\, P \to S_v$ minimization. This prior model has been drawn randomly from a uniform distribution constrained by the weak anisotropy assumption (equation 3.30). The uniform distribution parameters are $-0.05 \le \delta \le 0.2$ and $0 \le \varepsilon \le 0.2$.

A preconditioning conjugate gradient algorithm has been used to invert $P \to P$ and joint $P \to P \,\&\, P \to S_v$ traveltimes respectively. We continued $P \to P$ traveltime inversion up to 250 iterations and stopped the joint $P \to P \,\&\, P \to S_v$ traveltimes inversion after 48 iterations where it reached the true solutions. Figure 6.31 illustrates the convergence rate of the objective function (5.29) for both $P \to P$ (red curve) and joint $P \to P \,\&\, P \to S_v$ (blue curve). As can be seen, the $P \to P$ traveltimes has been quite well inverted after 50 iterations and not updated much after that. The estimated model parameters from the last iteration of $P \to P$ traveltime inversion are shown in the first

two columns of Table 6.11 while the last two columns show the estimated model parameters from the last iteration of joint inversion $P \to P \,\&\, P \to S_v$ traveltimes.

Figures 6.32 and 6.33 demonstrate the convergence pattern of δ and ε from the top layer for both $P \to P$ (red curve) and joint $P \to P \,\&\, P \to S_v$ (blue curve) respectively. Both parameters have converged to the true solution after 40 iterations. In the Figures 6.34 and 6.35 the convergence pattern of δ and ε is shown, respectively, for layer 3. It is obvious that prior models are quite far from the solutions. In Figures 6.36 and 6.37 we have illustrated the convergence rate of δ and ε from layer 7 and in Figures 6.38 and 6.39 they are plotted, respectively, for layer 9. The entire examples in this section show that a preconditioned conjugate gradient is robust in estimating the true model parameters in fewer iterations.

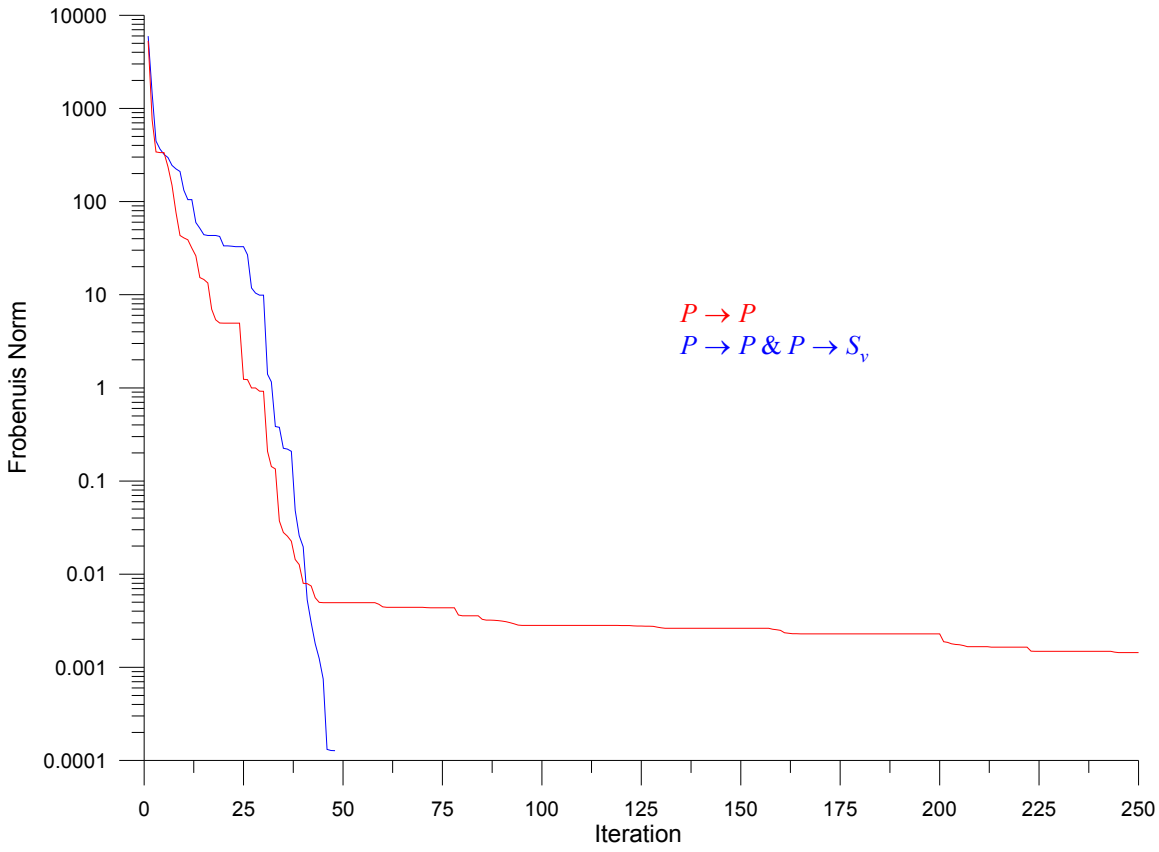


Figure 6.31: Convergence rate of the simultaneous inversion of $P \rightarrow P$ (red curve) and joint $P \rightarrow P \,\&\, P \rightarrow S_v$ (blue curve) traveltimes. Polak-Ribiére preconditioning conjugate gradient algorithm has been used to simultaneously invert Thomsen's anisotropy parameters in dipping TI media consisting of 9 layers. Both inversions use the same prior model. Joint inversion of $P \rightarrow P \,\&\, P \rightarrow S_v$ traveltimes has reached to the true solution after 48 iterations. All prior models have been drawn randomly from in a uniform distribution constrained by the weak anisotropy assumption.

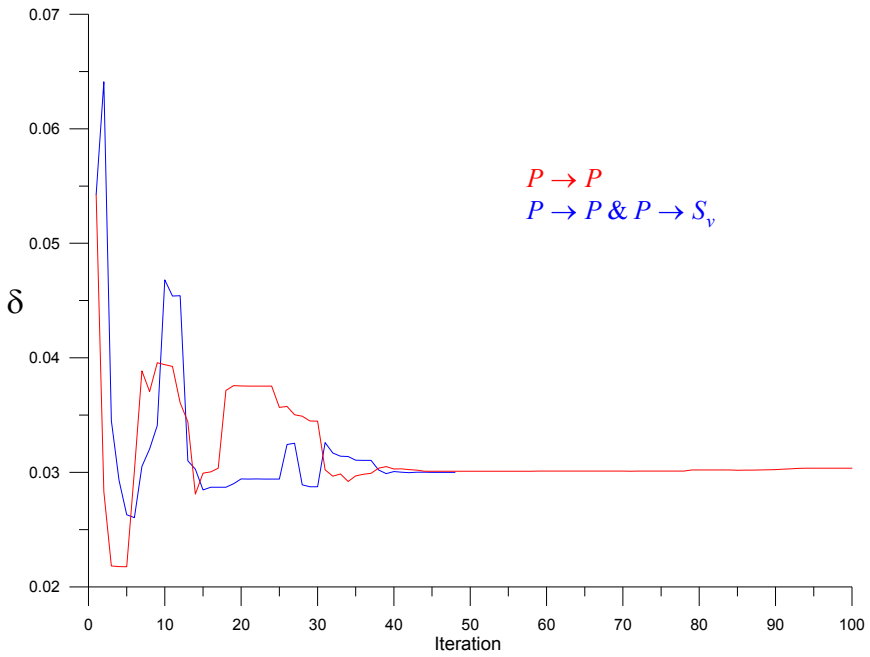


Figure 6:32 Convergence pattern of the δ for layer 1 from traveltime inversion of $P \rightarrow P$ (red curve) and joint $P \rightarrow P \,\&\, P \rightarrow S_v$ (blue curve) in dipping TI media. The true value is 0.03.

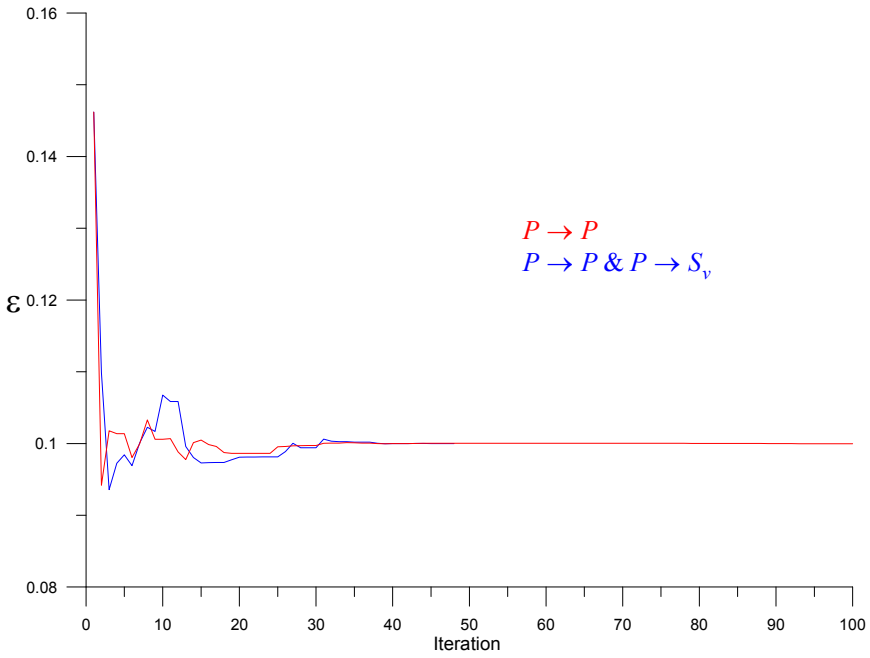


Figure 6:33 Convergence pattern of the ε for layer 1 from traveltime inversion of $P \rightarrow P$ (red curve) and joint $P \rightarrow P \,\&\, P \rightarrow S_v$ (blue curve) in dipping TI media. The true value is 0.1.

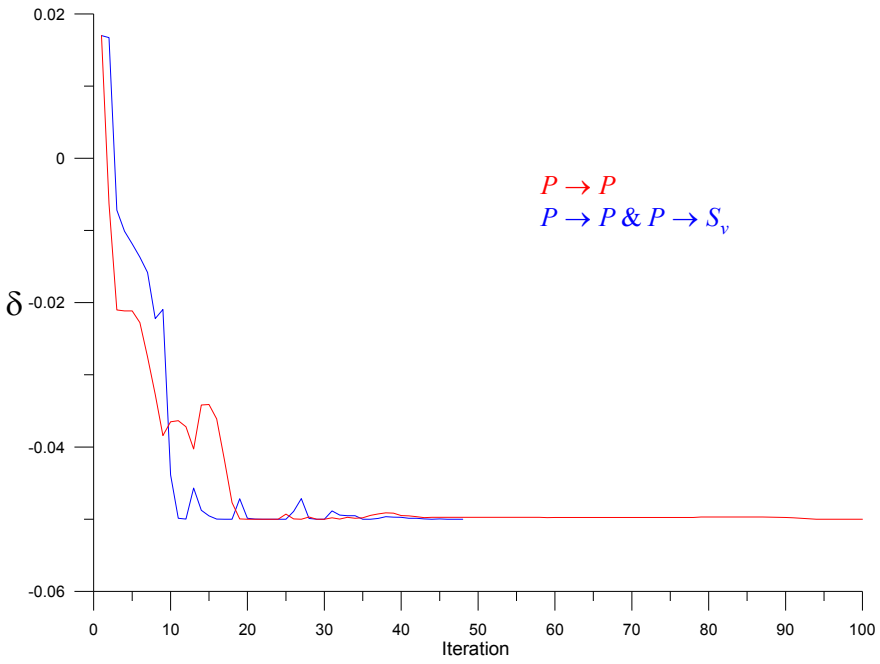


Figure 6:34 Convergence pattern of the δ for layer 3 from traveltime inversion of $P \rightarrow P$ (red curve) and joint $P \rightarrow P$ & $P \rightarrow S_v$ (blue curve) in dipping TI media. The true value is -0.05.

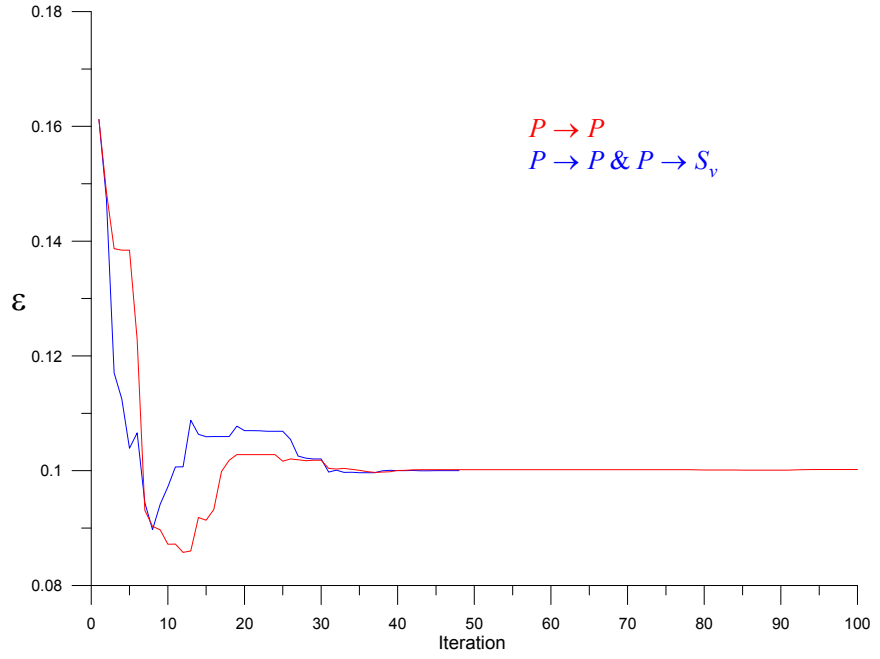


Figure 6:35 Convergence pattern of the ε for layer 3 from traveltime inversion of $P \rightarrow P$ (red curve) and joint $P \rightarrow P$ & $P \rightarrow S_v$ (blue curve) in dipping TI media. The true value is 0.1.

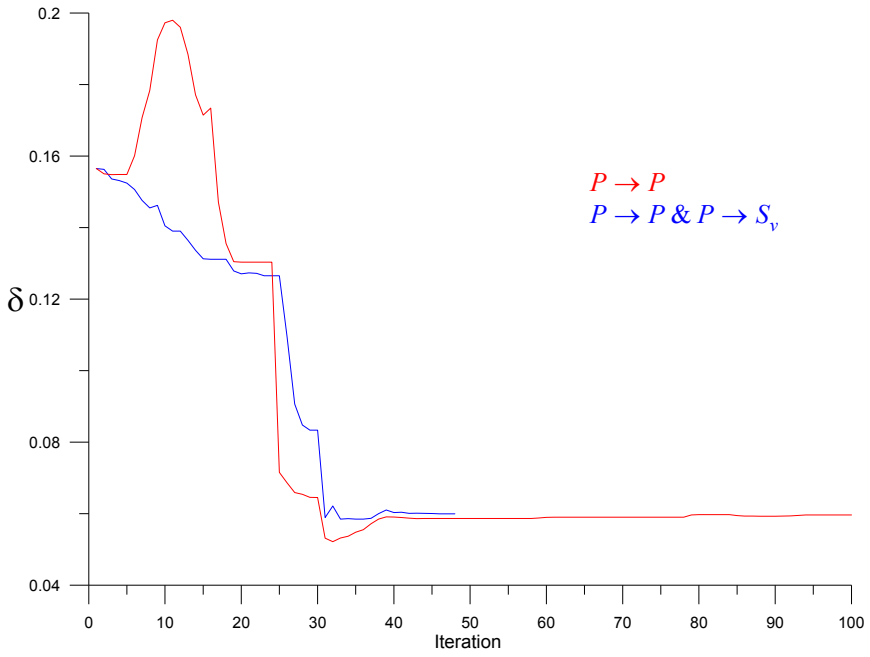


Figure 6:36 Convergence pattern of the δ for layer 7 from traveltime inversion of $P \rightarrow P$ (red curve) and joint $P \rightarrow P$ & $P \rightarrow S_v$ (blue curve) in dipping TI media. The true value is 0.06.

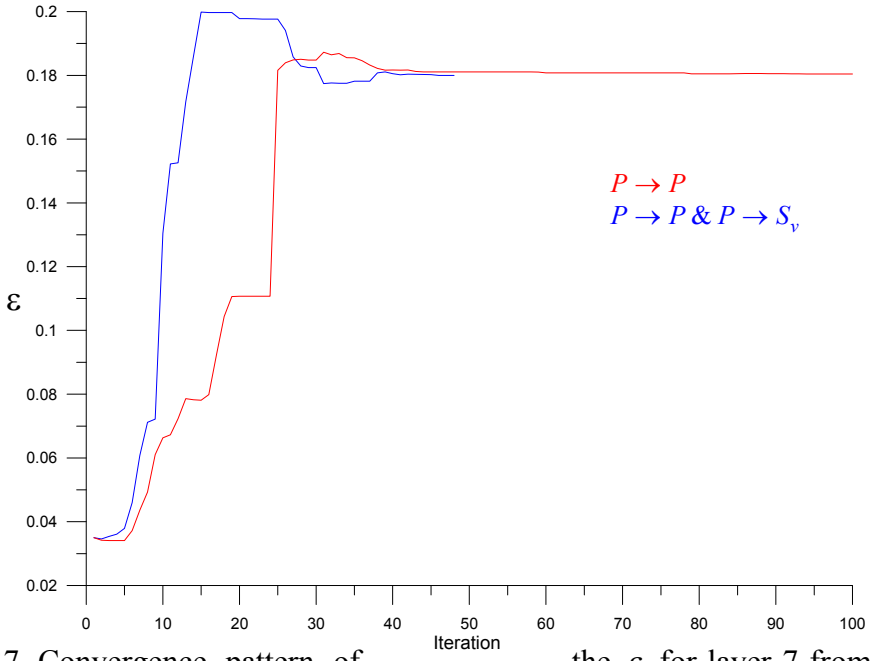


Figure 6:37 Convergence pattern of the ε for layer 7 from traveltime inversion of $P \rightarrow P$ (red curve) and joint $P \rightarrow P$ & $P \rightarrow S_v$ (blue curve) in dipping TI media. The true value is 0.18.

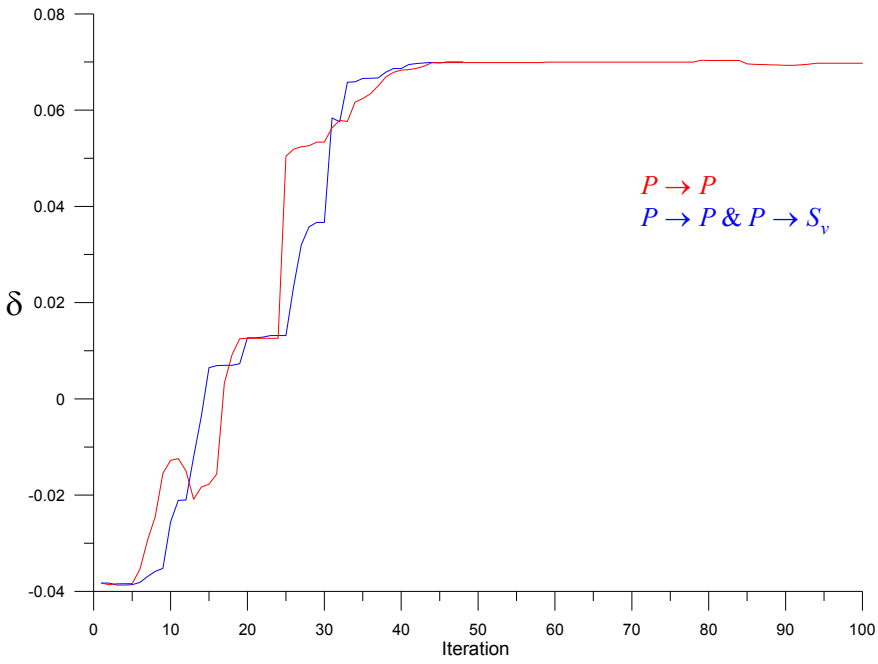


Figure 6:38 Convergence pattern of the δ for layer 9 from traveltime inversion of $P \rightarrow P$ (red curve) and joint $P \rightarrow P \,\&\, P \rightarrow S_v$ (blue curve) in dipping TI media. The true value is 0.07.

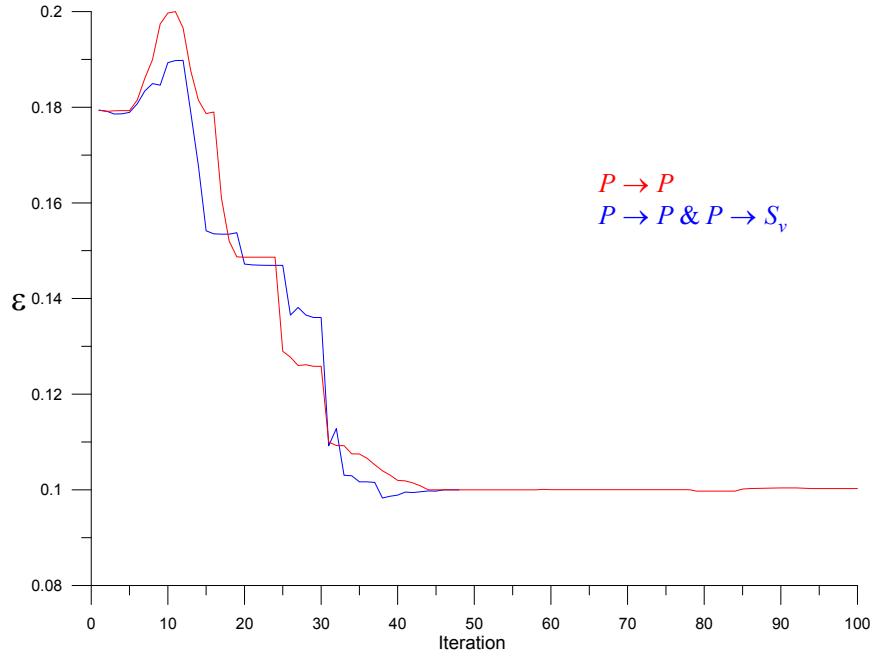


Figure 6:39 Convergence pattern of the ε for layer 9 from traveltime inversion of $P \rightarrow P$ (red curve) and joint $P \rightarrow P \,\&\, P \rightarrow S_v$ (blue curve) in a dipping TI media. The true value is 0.1.

Layer	δ	ε	δ	ε	δ	ε
1	0.054	0.146	0.03	0.1	0.030	0.1
2	-0.0297	0.133	-0.02	0.14	-0.021	0.14
3	0.017	0.161	-0.05	0.10	-0.05	0.10
4	0.077	0.157	0.1	0.14	0.10	0.14
5	0.116	0.137	0.05	0.1	0.05	0.10
6	-0.023	0.041	0.04	0.12	0.04	0.12
7	0.156	0.035	0.06	0.18	0.06	0.18
8	0.020	0.070	0.1	0.2	0.1	0.2
9	-0.038	0.179	0.07	0.1	0.07	0.10

Table 6.12: The first two columns correspond to prior model which has been used in traveltime inversion of $P \to P$ and joint $P \to P \,\&\, P \to S_v$ from 96 traces in dipping TI media. The second two columns correspond to estimated model parameters from the inversion of $P \to P$ traveltimes for the last iteration (Figure 6.31). The last two columns correspond to estimated model parameters from joint inversion of the $P \to P \,\&\, P \to S_v$ traveltimes from the last iteration (Figure 6.31). Figures 6.32-6.39 have used the current prior model.

Comments

Traveltimes are particularly useful to estimate the low frequency part of the model space. Using the far offset $P \to P$ traveltimes one can simultaneously estimate model structure, vertical P-wave velocities (α_0), however, due to the huge difference in scale between Thomsen's anisotropy parameters (δ, ε) and other parameters (α_0, β_0, z), ε and especially δ are poorly resolved. Joint traveltime inversion of $P \to P \,\&\, P \to S_v$ can efficiently estimate vertical S-wave velocity (β_0) as well as α_0 and layer thickness (z). Having a good velocity model either from check shot data or traveltime inversion, one can estimate δ and ε efficiently. This can be accomplished without a prior knowledge by using the weak anisotropy assumption. Where, the layer thickness is small compared to Rayleigh vertical resolution and picking the traveltime of corresponding reflector is difficult, waveform inversion is an alternative to estimate the high frequency part of the model in terms of Thomsen's anisotropy parameters.

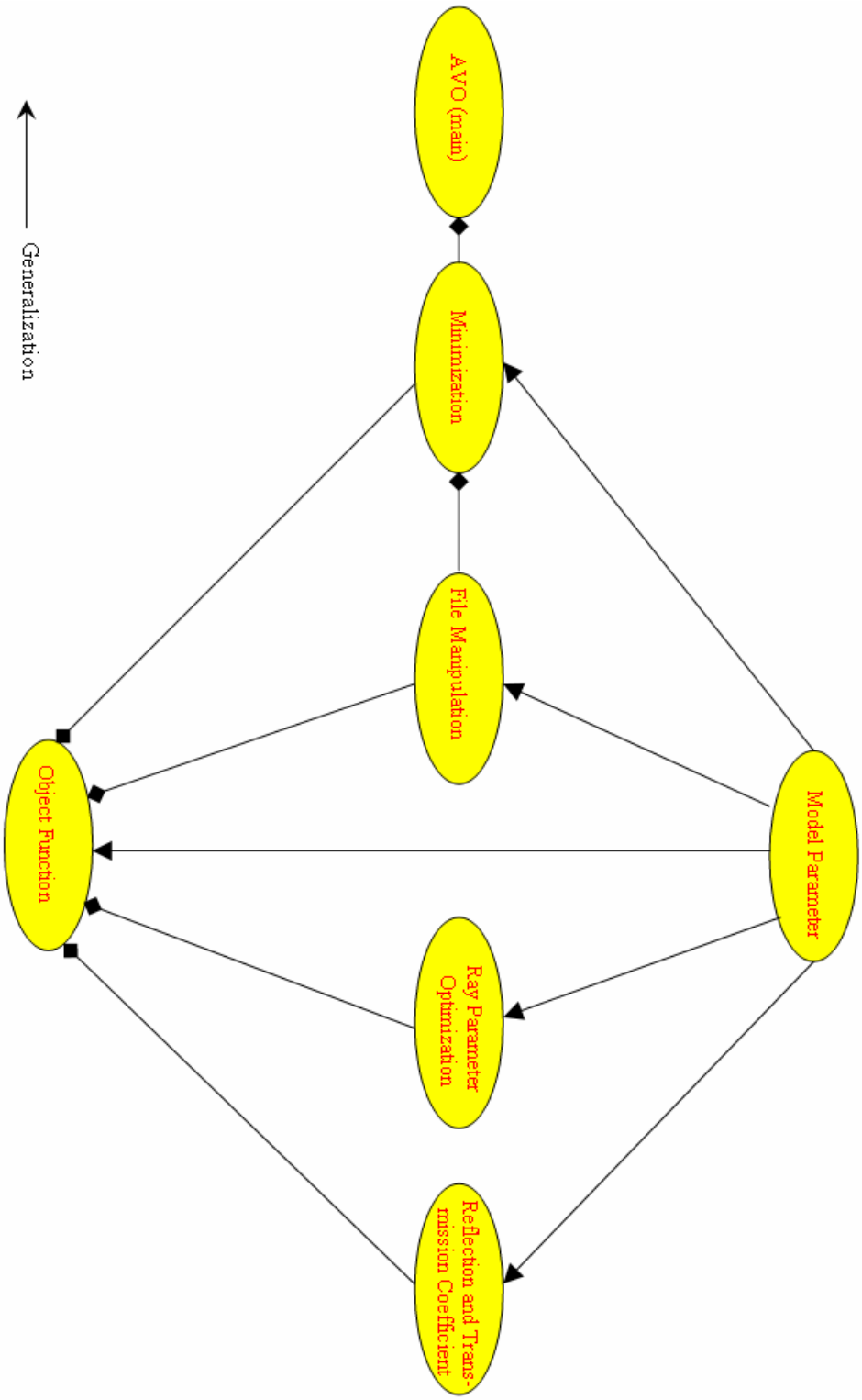


Figure 6.40: A schematic class diagram illustrates different classes of the package. Class Attributes and operations are not shown here.

References

Bonnans, J. F., J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, 2003, Numerical optimization: Theoretical and practical aspects: Springer-Verlag.

Gunning, J., and M. E. Glinsky, 2004, Delivery: an open-source model-based Bayesian seismic inversion program: Computers & Geosciences, **30**, 619-636.

Nadri, D., and B. Hartley, 2007a, Non-linear AVO inversion in a VTI medium: Presented at the EAGE 69th Conference & Exhibition, European Association of Geoscientists and Engineers, P356.

Nadri, D., and B. Hartley, 2007b, Non-linear joint AVO inversion of PP and PS waves in a VTI medium: Presented at the 19th International Geophysical Conference & Exhibition, Australian Society of Exploration Geophysicists.

Nadri, D., and B. Hartley, 2008, Non-linear estimation of Thomsen's anisotropic parameters in a dipping TI media: Accepted at the 70th EAGE Conference & Exhibition, European Association of Geoscientists and Engineers, p292.

Nocedal, J., and S. J. Wright, 1999, Numerical optimization: Springer-Verlag.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 2002, Numerical recipes in C++: the art of scientific computing: Cambridge University Press.

Ravindran, R., K. M. Ragsdell, and G. V. Reklaitis, 2006, Engineering optimization: Methods and applications: John Wiley & Sons.

Rüger, A., 1996, Reflection coefficients and azimuthal AVO analysis in anisotropic media: PhD, Colorado School of Mines.

Saad, Y., 2000, Iterative Methods for Sparse Linear Systems.

Ursin, B., and K. Hokstad, 2003, Geometrical spreading in a layered transversely isotropic medium with vertical symmetry axis: Geophysics, **68**, 2082-2091.

Ursin, B., and A. Stovas, 2006, Traveltime approximations for a layered transversely isotropic medium: Geophysics, **71**, D23-D33.

Conclusions, Discussions, and Future prospects

Conclusion and discussion

One of the achievements of this thesis is providing a framework to integrate all the relevant information to create a robust algorithm to effectively minimize the objective function. This facilitates the simultaneous estimation of different model parameters from different layers. Using the sophisticated unconstrained minimization algorithms such as nonlinear preconditioning conjugate gradient, traveltimes from a horizontally, layered vertical transversely isotropic medium have been inverted to estimate the Thomsen's anisotropy parameters. With the assumption of weak anisotropy, estimation of Thomsen's anisotropy parameters is independent of specific prior information.

We have shown that vertical P-wave velocity and thickness of the layers can be effectively estimated using traveltime inversion of compressional waves, however, joint inversion of $P \rightarrow P \,\&\, P \rightarrow S_V$ has been used to recover the vertical shear wave velocity as well as vertical compressional wave velocity and thickness, in fewer iterations.

The effect of source-receiver offset on the estimation of Thomsen's anisotropy parameters has been studied. From the implementation of both pre-critical and post-critical phase incidence angles, we have show that only pre-critical reflections are sufficient to estimate the velocity structure and Thomsen's anisotropy parameters. However, long offset generally results in faster convergence.

Another achievement of this thesis is formulating new equations for offset and exact traveltime in dipping layered transversely isotropic media. These equations have been expressed as a function of ray parameter which is suitable for ray tracing algorithms. Due to nonhyperbolic behaviour of traveltimes, especially for converted shear waves, ray tracing in dipping TI media is much more complicated than in horizontal layers. We use the Newton method to estimate the ray parameter for each ray path by inverting the offset equations. A random ray parameter drawn from a uniform distribution has been used as a prior model.

We have also demonstrated the ability of the unconstrained minimization algorithms to estimate the Thomsen's anisotropy parameters in dipping transversely

isotropic media. With the assumption of weak anisotropy both $P \to P$ and joint $P \to P \,\&\, P \to S_V$ traveltime inversions have successfully estimated the Thomsen's anisotropy parameters from the random prior model drawn from a uniform distribution. This also enables the estimation process to be independent of prior knowledge.

Horizon-based inversion schemes such as traveltime inversion are unable to recover the low-frequency part of a model structure. Reflectivity based methods at the interface of two layers where the thickness of the layer is below the tuning limit suffer from the interference effect, so estimated parameter may be biased. Amplitude inversion based schemes are more effective in estimation of the high frequency part of the model. Prestack waveform inversion is, however, highly nonlinear so there is no guarantee to converge to a solution from one realization of a prior model. To be able to estimate Thomsen's anisotropy parameters, a good prior velocity model is essential. This may be achieved by using well log data or the traveltime inversion as discussed above. Prestack waveform inversion is also computationally intensive. I have shown that restricting the frequency spectrum to the signal bandwidth, does not affect the partial derivative considerably but it significantly reduces the computations.

Gradient based inversion algorithms need partial derivatives of the objective function. It is not always possible to derive the partial derivatives of exact reflection and transmission coefficient analytically. So the common approach is to use finite difference calculations instead. Finite differences are computationally intensive and require an optimum interval. I have shown that the numerical derivatives of the objective function of the effective reflection coefficient with respect to δ and ε are highly unstable. We preferred to use Rüger's equations (1996) for different modes. They approximate the exact plane wave reflection and transmission coefficients for pre-critical phase angles of incidence, where the difference in Thomsen's anisotropy parameters is small. It is also possible to derive their analytic partial derivatives, however, it is cumbersome to extract the derivative of effective reflection coefficients, where the transmission coefficients in layered system needs to be included. I have shown that neglecting the transmission coefficients may cause error in effective reflection coefficient of up to 10 percent. I have demonstrated that prestack waveform inversion is able to effectively estimate the Thomsen's anisotropy parameters where the layer thickness is within the tuning limit. I

found Joint inversion of $P \to P \,\&\, P \to S_V$ seismic amplitudes converges to the solution faster than just $P \to P$ inversion, however, this increases the nonlinearity of the model space and we don't expect joint $P \to P \,\&\, P \to S_V$ to minimize the same prior model as $P \to P$ does.

I have used C++ as a programming platform because it is fast and its object oriented programming environment provides a framework to easily reuse the classes we have developed here. New classes can be easily integrated within the package.

Future prospects

The inversion algorithms which have been developed in this thesis, for a horizontally VTI model could be extended to an HTI model. This could be done by transformations given by Tsvankin (1997) which convert an equivalent VTI model to an HTI model. The inverse models could be further extended to estimate fracture orientation using the azimuthal variation of traveltimes. This could be done either by rewriting the Christoffel equations for an HTI model directly, in the same way as Ursin and Stovas (2006) did for a VTI model, or to use the Tsvankin (1997) equivalent VTI model (Nadri et al., 2008b; Nadri et al., 2008c). It is also possible to use the azimuthally varying $P \to P$ effective reflection coefficients from the Rüger's equations (Rüger, 1998) or joint azimuthally varying $P \to P$ (Rüger, 1998) and $P \to S_V$ (Jílek, 2002) waves.

The computation of ray parameter in dipping TI media is intensive. This is mainly due to nonhyperbolic moveout behaviour of traveltimes. So, the ray parameter should be very precise and this causes extra burden on Newton minimization. During the minimization, where a large number of function evaluations are necessary, CPU time will be large, however, a preconditioning conjugate gradient algorithm ensures faster convergence.

Parallel processing is an effective solution to distribute the computations to different CPUs in a cluster. Computation of each ray parameter for each ray path is independent of other ray paths. So, using parallel programming for example, OpenMP (Chapman et al., 2008), one could easily port the computation of ray parameters, at least for each reflector, to a specific CPU.

References

Chapman, B., G. Jost, and R. Van Der Pass, 2008, Using OpenMP: Portable shared memory parallel programming: MIT Press.

Jílek, P., 2002, Converted PS-wave reflection coefficients in weakly anisotropic media, *in* I. Pšenčik and V. Červený, eds., Seismic waves in laterally inhomogeneous media: Birkhäuser Verlag, 1527-1562.

Nadri, D., A. Bóna, and B. Hartley, 2008b, Fracture orientation estimation in TI media using nonlinear inversion of traveltimes: Submitted to The 13th International Workshop on Seismic Anisotropy, Center for Wave Phenomena at Colorado School of Mines.

Nadri, D., B. Hartley, and A. Bóna, 2008c, Nonlinear anisotropic parameter estimation in TI media: Submitted to the 78th Annual International Meeting, Society of Exploration Geophysicists.

Rüger, A., 1996, Reflection coefficients and azimuthal AVO analysis in anisotropic media: PhD, Colorado School of Mines.

Rüger, A., 1998, Variation of P-wave reflectivity with offset and azimuth in anisotropic media: Geophysics, **63**, 935-947.

Tsvankin, I., 1997, Reflection moveout and parameter estimation for horizontal transverse isotropy: Geophysics, **62**, 614-629.

Ursin, B., and A. Stovas, 2006, Traveltime approximations for a layered transversely isotropic medium: Geophysics, **71**, D23-D33.