Adjoint slope tomography : a velocity macro-model building tool for seismic imaging

Borhan Tavakolifaradonbeh

UNIVERSITÉ CÔTE D'AZUR

Géo AZUR
TERRE - OCÉAN - ESPACE

Ecole doctorale Sciences Fondamentales et Appliquées
Unité de recherche : Géoazur

Thèse de doctorat

Présentée en vue de l'obtention du
grade de docteur en Sciences de la Planète et de l'Univers
de
l'UNIVERSITÉ CÔTE D'AZUR

par

Borhan TAVAKOLI F.

Adjoint slope tomography: a velocity macro-model building tool for seismic imaging

---

Dirigée par **Alessandra RIBODETTI** et **Jean VIRIEUX**

Soutenue le 16 Novembre 2017 devant le Jury composé de

Thibaut ALLEMAND	Chercheur, CGG Massy	Invité
Reda BAINA	Directeur de recherche, OPERA Pau	Examinateur
Hervé CHAURIS	Professeur MINES ParisTech	Rapporteur
Gilles LAMBARÉ	Directeur de recherche, CGG Massy	Rapporteur
Mark Noble	Directeur de recherche, MINES ParisTech	Examinateur
Stéphane OPERTO	Directeur de recherche CNRS - Geoazur	Invité
René-Edouard PLESSIX	Professeur IPGP - Chercheur senior Shell	Examinateur
Alessandra RIBODETTI	Chargé de recherche IRD - Geoazur	Directeur de Thèse
Jean VIRIEUX	Professeur Université Grenoble-Alpes	Co-directeur de Thèse

Acknowledgement

First of all, I would like to thank Stéphane Operto, Alessandra Ribodetti and Jean Virieux for all of their support, help and enthusiasm for leading this research. Although Stéphane was not officially the supervisor of my thesis, but indeed he led a significant part of this research and I really appreciate all he has done for me in these years. I thank Alessandra for her endless kindness, encouragement and discussions during my PhD and thank Jean for his wise remarks, constructive critics and comments.

I thank Romain Brossier, Ludovic Metteviér and all the SEISCOPE members in Grenoble for the fruitful discussions we had during the meetings and specially Paul Wellington for sharing his experience in migration of real data with me. I thank Amir Asnaashari who introduced me this team and without his suggestion I would never have this opportunity to work with SEISCOPE. I would like to thank Alain Miniussi and specially Laure Combe for their great advice on parallel implementation of my code and their kindness and friendship.

My life in France without my friends could not be so pleasant, who made me nice memories. I would like to express my appreciation to Stephen Beller and Alexandre Belli for all the fun we had together and their support. I thank Farhood and Masomeh, for a close friendship in Sophia, and Serge for sharing office with me and all the discussions we had together. I would also like to acknowledge my other friends in GeoAzur: Laury, Vishnu, Théa, Huyen, Zoe, Mathild, Eduard, Sargis, Sara, Nicholas, Monica and .... I would like to thank my friends in Iran which despite the distance between us, they always kept in touch with me and supported me, specially Mahboobeh and Alireza for their kindness and pure friendship. I would like to thank all the Geoazur members specifically Valérie Mercier-Valero, Véronique Large, Olivier Laurain, Jenny Trévisan, Jean-Xavier Dessa, Magali Aufeuvre and Muarin Vidal for all their help and considerations which made GeoAzur a great place for me to work.

Prof. Chauris and Dr. Lambaré reviewed my thesis and Dr. Baina, Dr. Plessix, Dr. Noble and Dr. Allemand accepted to be a part of my defence jury. I would like to thank them for their time, insightful comments and suggestions which helped to improve this manuscript. I would like to thank the sponsors of SEISCOPE consortium, giving special mention to CGG for providing the real data, Dr. Lambaré and Dr. Allemand from CGG-Massy for the discussions on my research.

My greatest appreciations go for my father, mother, brother, sisters, brother and sister in law. Words can not express what they have done for me and how much I love them. Without them this journey was not possible. This thesis is dedicated to them and my nephews Aref, Ahoura and Farham.

I sincerely apologize if I forgot some names, but there were many others who helped me during these three years.

This study was granted access to the HPC resources of SIGAMM (http://crimson.oca.eu) and CIMENT computer centers and CINES/IDRIS under the allocation 046091 made by GENCI.

Abstract

Velocity macro-model building is a crucial step in seismic imaging workflows as it provides the necessary background model for migration or full waveform inversion. Slope tomography, as a reliable alternative for conventional tomography, provides a tool to achieve this purpose where one picks the local coherent events rather than continuous events. These approaches are based on the slopes and traveltimes of the local coherent events which are tied to a reflecting/diffracting point (scatterer) in the subsurface. The forward modelling in these methods is based on the ray tracing which may suffer from poor sampling of the medium in presence of complex subsurface structures and long offset acquisitions. On the other hand, the inverse problem is mostly solved based on the explicit calculation of Freéchet derivative matrix which for large data sets can be computationally expensive.

In this thesis, I introduce an anisotropic slope tomographic approach which aim at macro-model building for subsurface properties in 2D tilted transversely isotropic (TTI) media. In this method, I reformulate the stereotomography, as an slope tomographic tool, such that I replace the ray-based forward engine with a TTI eikonal solver and take advantage of the adjoint state method to calculate the gradients. In result, I can efficiently calculate the traveltimes for complex media and long offset acquisition on a regular grid of the subsurface and formulate a matrix-free framework for the inversion.

The eikonal solver is based on the fast sweeping method to solve the elliptic eikonal equation while through the fixed point iterations a right hand side term updates the elliptic eikonal and accounts for the anellipticity of TTI eikonal equation. Also, in order to achieve highly accurate traveltimes, a factorization scheme is considered in the elliptic eikonal solver to remove the upwind point source singularity. Due to this definition for forward modelling, compared to the stereotomography, the model and data space is smaller and this may lead to less ill-posed inverse problem. The inverse problem is solved with the quasi-Newton L-BFGS method in which the Hessian is estimated with few gradients of the previous iterations. The gradients are achieved by the adjoint state method mainly through the resolution of two adjoint state equations.

Different synthetic examples including the isotropic Marmousi and 2D TTI BP-salt model are considered to assess the potential of the method in subsurface parameter estimations. Also, through a simple example, the footprint of parameter cross-talk is investigated for the Thomsen parametrization. As a real data application, the proposed method is applied on a 2D marine BroadSeis data set (provided by CGG) to retrieve the vertical velocity of the subsurface while during the inversion the Thomsen anisotropy parameters are kept fixed.

Résumé

La construction du macro-modèle de vitesse est une étape cruciale et complexe de la chaîne d'imagerie sismique car de sa capacité à prédire les temps de propagation des ondes sismiques dépendent la focalisation et le positionnement des réflecteurs dans les images migrées et la convergence des méthodes d'inversion non linéaires de formes d'onde complètes (FWI: Full Waveform Inversion) vers une solution proche du minimum global de la fonction coût. Parmi les approches possibles pour construire ce macro-modèle, la tomographie des pentes est fondée sur le pointé d'évènements localement cohérents caractérisés par leur temps de trajet et leurs pentes dans les collections de sismogrammes à source et réflecteur communs. Chaque évènement dans les observables est associé à un petit segment de réflecteur dans le sous-sol caractérisé par sa position et son pendage. Le caractère local du pointé est l'élément déterminant favorisant l'automatisation de cette tâche et une exploitation plus exhaustive de l'information contenue dans les données sismiques.

Les méthodes existantes de tomographie des pentes sont fondées sur des techniques de tracé de rai pour le calcul des temps et des pentes. Une limite de ces techniques réside dans la difficulté de contrôler l'échantillonnage du sous-sol par les rais particulièrement dans le cas de milieux géologiques complexes et de dispositifs d'acquisition impliquant de forts déports source-capteur. Par ailleurs, le problème inverse repose sur le calcul explicite de la matrice de sensibilité (ou de Fréchet) dont la construction et la résolution matricielle en découlant peuvent se révéler coûteuses lors du traitement de gros volumes de données.

Cette thèse propose une reformulation de la tomographie des pentes pour palier aux deux limitations sus-mentionnées. Les temps de trajet sont calculés à l'aide d'un solveur eikonal et la méthode de l'état adjoint est utilisée pour calculer efficacement le gradient de la fonction coût dans le contexte de méthodes d'optimisation locale implémentées avec l'algorithme de quasi-Newton l-BFGS. Le solveur eikonal est fondé sur une méthode aux différences finies pour la discrétisation des opérateurs différentiels, la méthode du "fast sweeping" pour propager les fronts d'onde et d'une approche factorisée pour s'affranchir de la singularité à la source. La méthode est implémentée pour des milieux 2D transverses isotropes avec un axe de symétrie dont l'orientation varie spatialement (milieux TTI). Dans ce contexte, l'équation eikonale TTI est résolue avec la méthode du point fixe après l'avoir ré-écrite sous forme d'une équation eikonale "elliptique" dont le terme de droite concentre les effets d'anellipticité. Les pentes à une source et à un récepteur donnés sont calculées au sens des différences finies en utilisant des simulations calculées pour les sources et les récepteurs voisins et en exploitant la réciprocité de la propagation entre un point d'émission et de réception.

La méthode est évaluée avec différents examples synthétiques de complexité croissante.

Le rôle des pentes pour localiser les segments de réflecteurs est tout d'abord illustré avec un modèle isotrope simple à une couche avant de d'aborder la cas du modèle complexe Marmousi pour lequel un pointé réaliste est utilisé. Cet example permet de démontrer que la tomographie des pentes fournit un modèle de vitesse initial suffisamment précis pour la FWI. Dans le cas de milieux TTI, les couplages pouvant exister entre les paramètres de différente nature (vitesse verticale et paramètres de Thomsen ou toute autre combinatoire et position des segments) sont analysés avec un cas synthétique canonique avant de présenter une application sur un modèle synthétique plus réaliste (le modèle 2D TTI BP-salt) où la vitesse verticale et le paramètre de Thomsen ϵ sont reconstruits de manière conjointe.

Cette thèse se conclut par une application à des données de sismique réflexion multitrace large bande fournie par la Compagnie Générale de Géophysique (CGG). Le modèle de vitesse vertical reconstruit avec l'approche développée dans cette thèse est comparable avec celui obtenu par la CGG. La pertinence de ce modèle est validé par ailleurs en l'utilisant comme modèle de référence pour effectuer une migration par renversement temporel (RTM: Reverse Time Migration) et vérifier l'horizontalité des réflecteurs dans le volume migré avant sommation.

A titre de conclusion, quelques perspectives de ce travail sont évoquées parmi lesquelles l'inversion conjointe des temps de trajet des primières arrivées et des ondes réfléchies ainsi que l'implémentation de régularisations ou d'une paramtrisation parsimonieuse de l'espace des modèles permettant de réduire les ambiguités potentielles entre les vitesses sismiques et la position des segments de réflecteurs.

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

Chapter overview: In this chapter, I first review the basic principles of seismic methods in exploration seismology, the geometry of the main seismic acquisitions and the anatomy of the seismic measurements. Then, I introduce the motivation beyond the concept of scale separation between the velocity macro-model and the reflectivity, which describe the long and short wavelength components of the subsurface, respectively. Most of the seismic imaging workflows based on reflected waves rely on this concept through the alternate application of tomography (or migration-based velocity analysis) for the velocity macro-model building and migration for the reflectivity imaging. Then I focus more specifically on the velocity-model building methods to which this thesis is more closely related. I first introduce the basics of reflection traveltime tomography before extending the discussion to slope tomography methods, the tomographic methods that exploit the slope of locally-coherent events in addition to their traveltimes to update the subsurface. I first review the pioneering ray-based stereotomography method before extending the discussion to various recent variants of non linear slope tomography. I also present a short review of anisotropy in the Earth as almost all of the tomographic approaches in exploration geophysics need to account for anisotropic effects. I conclude this introduction by the main objectives of this thesis and the outline of the chapters.

1.1 Exploration seismology

1.1.1 History

Seismic imaging triggered by observing the first seismogram near the end of 19th century. The first application of seismology was dedicated to the localization of earthquake epicentres and it was in 1884 that Mallet reported the first controlled-source experiment (by opposition of natural sources provided by earthquakes) to characterize the subsurface. Knott (1899) developed the theory of wave partitioning at interfaces (estimation of reflection and transmission coefficients as a function of the incidence angle) and this work was later extended by Wiechert

and Zoeppritz (1907); Zoeppritz and Geiger (1907). During World War I, the efforts devoted to locate the heavy artilleries from seismic waves generated by recoil laid the foundations of seismic exploration. Discovery of a salt dome in Texas by refraction seismology made this method popular, and it was 1927 that commercial utilization of reflection seismology began in Oklahoma. The first subsurface maps was generated by recognizing one arrival on a single seismic record, and it was in 1929 that calculation of dip of events on several seismic records permitted the successful application of reflection seismology to map laterally-coherent geological features. This prompted the pioneers of exploration geophysics to design surveys with more receivers, leading to the emergence of reflection seismology for subsurface imaging. Digitizing the seismic records and mathematical developments in early 1960s revolutionised the exploration seismology by defining the "data processing" concept. Meanwhile, there were some achievements in seismic sources (e.g. in 1953 Vibroseis method developed) and seismic survey design (e.g. common midpoint method patented in 1956) (Telford et al., 1990). These were the first steps in developing the modern exploration seismology.

1.1.2 Principles

Seismic acquisition (Fig. 1.1) includes an array of receivers which records the seismic energy propagated by controlled seismic sources (i.e. dynamite, vibroseis, airgun, etc.). In land explorations, the receivers are on the surface or at depth along a vertical well in vertical seismic profiling (VSP). In marine (off-shore) surveys, the receiver array is near the sea surface and is tied to the source position (towed-streamer acquisition) or stationary on the sea bed in ocean bottom seismic (OBS), which can be carried out either with multicomponent ocean bottom cables (OBC) or autonomous ocean bottom nodes or seismometers (OBN) (Fig. 1.1).

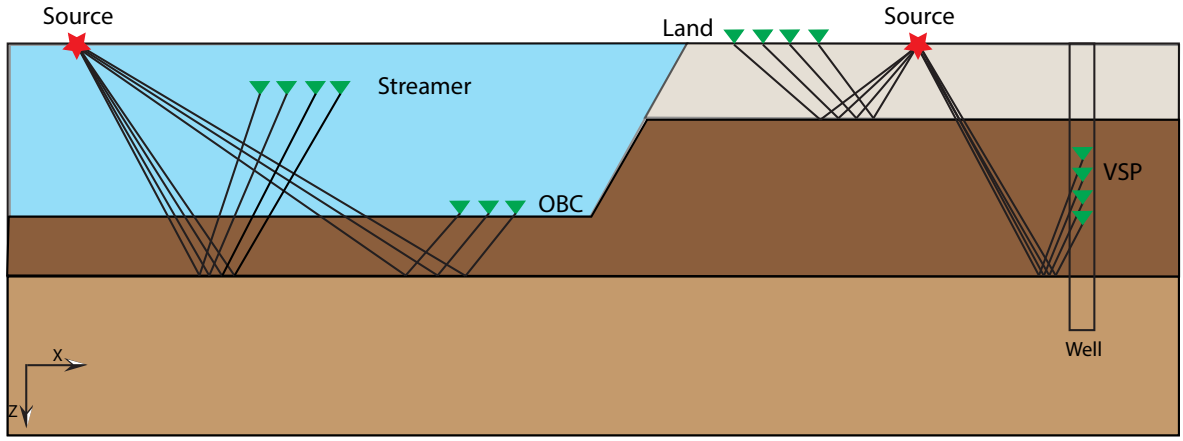

Figure 1.1 – Schematic presentation of different seismic acquisitions. Green triangles are the receivers. In marine acquisition (the blue region is sea), hydrophones are deployed in streamers and are towed by the shooting vessel, while the stationary receivers are on the sea bottom in ocean bottom seismic. For land acquisition, the geophones are either on the surface and/or along a vertical well in VSP.

The released energy by a source generates both compressional (P) or/and shear (S) waves in the subsurface. In exploration seismology, the subsurface is often assumed to be a fluid (acoustic medium), within which only P waves propagate. During their propagation in the subsurface, the waves interact with the subsurface heterogeneities through refraction, (multi-) reflection and diffraction before being recorded by receivers near the surface (Fig. 1.2). Waves, which are sent downward from the surface, are recorded near the surface during their rising path after reflection or refraction from first-order lithological discontinuities or beyond the grazing incidence for diving waves.

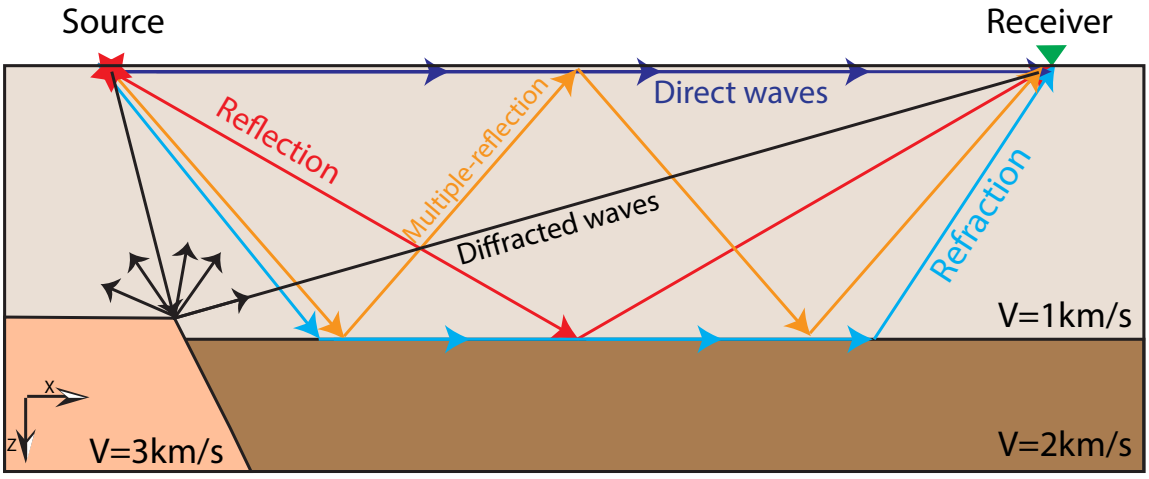

Figure 1.2 – Anatomy of wave propagations in a synthetic acoustic model. The main arrivals are: reflections, namely waves reflected from the first-order velocity contrasts or reflectors; Direct wave which travels directly from the source to the receivers without any interaction with subsurface heterogeneities; multiple reflections between the free surface and the internal reflectors; Refracted waves which travel along the velocity discontinuities. Diffracted waves which result from the interaction with an obstacle, the size of which is of the order of the propagated wavelength. Here, the arrows represent the rays, the trajectories perpendicular to the wavefront.

Each active source (shot) is recorded by several receivers and each of them records a seismogram. Collection of these seismograms builds a common-shot gather. Tracking the lateral coherency of the recorded energy in common-shot gathers, namely, in the time-offset domain where the offset is the horizontal distance between the source and the receivers, allows for a first identification of the main seismic phases and a quality control of the data. Fig. 1.3(a-c) shows a realistic synthetic subsurface acoustic model with a seismic source at the top left corner, and Fig. 1.3(e) is the recorded common shot gather with a 1D receiver array on the surface. Different phases are labelled on the gather and the associated schematic wave trajectories are plotted on the subsurface model.

The recorded seismic data can be sorted in different ways by considering another common attribute of the acquisition. For instance, gathering all the seismograms which are related to

the same receiver forms a common receiver gather (Sheriff and Geldart, 1995). Among other widely used gathers are common mid point (CMP) gather which includes several shot-receiver pairs with common mid point. Fig. 1.4 shows a schematic representation source and receiver layouts for common shot, CMP, common receiver and common offset gathers.

The seismic arrivals carry two types of information: 1) kinematic information which are related to the traveltime of the arrivals from the source to the receiver, 2) dynamic information which are related to the amplitude of the recorded arrivals. The kinematic information are a function of the trajectory and propagation velocity of the waves, and dynamic information are influenced by different factors such as the source wavelet, reflection/transmission coefficients of the interfaces, geometrical spreading and attenuation.

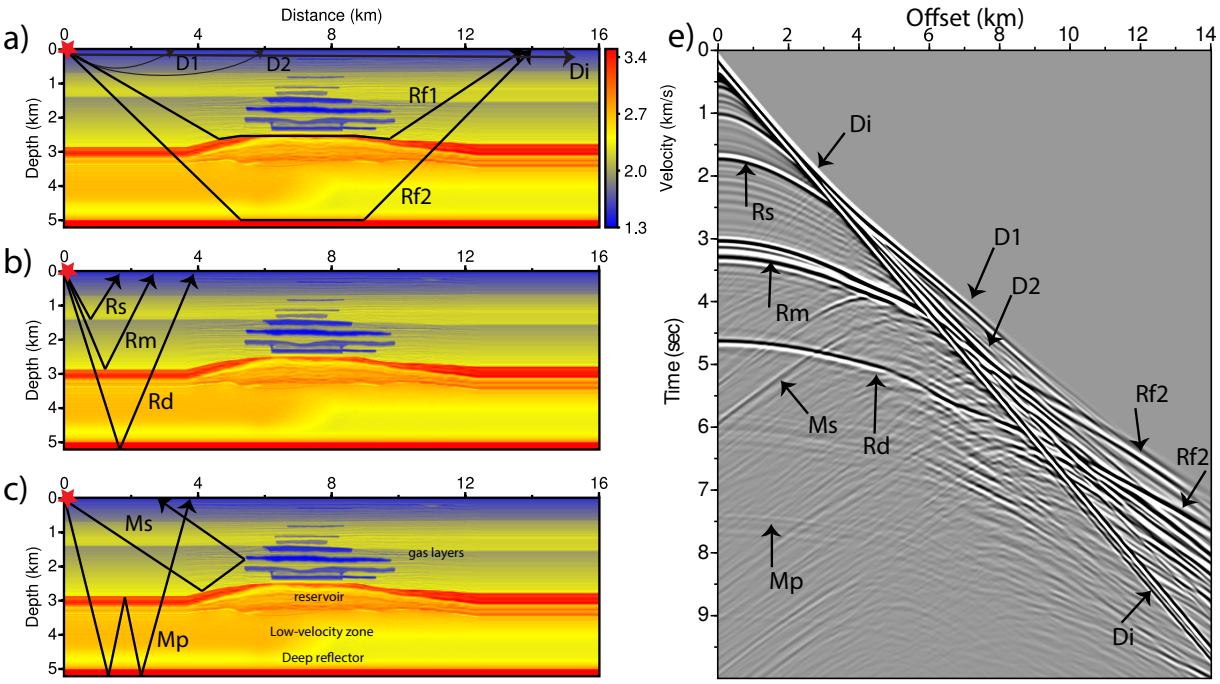

Figure 1.3 – Wave propagations and recorded shot gather in a synthetic shallow-water model representative of the North sea. a-c) The subsurface main structures are: low-velocity gas layers between $1400 - 2400$ m in depth, the reservoir at 2500 m in depth, and a flat reflector at 5000 m in depth. e) The modelled synthetic acoustic seismograms for the subsurface model shown in the left panels with a source on the top left corner (red ★) and the receivers on the surface. The labels designate different phases on the seismograms and the associated schematic ray-paths on the model: direct waves (Di) with straight path, diving waves (D1, D2) in the upper structure, refraction (Rf1, Rf2) from the velocity discontinuities by the top of the reservoir and deep reflector, reflections (Rs, Rm, Rd) from the reflector at shallow, middle and deep zones, multi-scattered waves (Ms) due to diffraction from the sharp edges of gas layers and multiple reflection (Mp) between the surface and reflectors (figure adapted form Zhou (2016)).

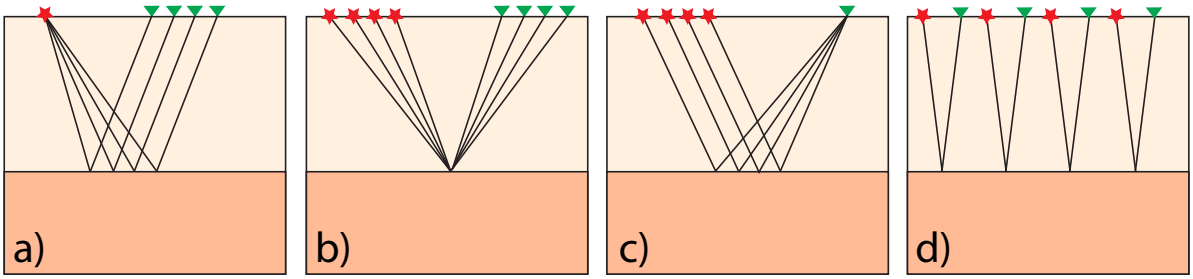

Figure 1.4 – Source and receiver layouts for different gathers. a) common shot gather, b) common mid point (CMP) gather, c) common receiver gather and d) common offset gather.

1.1.3 Seismic imaging workflow

In exploration seismology, seismic imaging relies on the measurement of seismic waves generated by controlled sources and aims to transform these observables into a well-resolved image of the subsurface down to several kilometres depth.

In order to achieve the final interpretable seismic image which accurately reflects the geology of the targeted area, one need to process the seismic data with an appropriate processing workflow. But before that, we need to remove some noise from the data. Here, the noise does not have a general strict definition; it is rather defined based on the key principles of the processing workflow to be performed. For instance, the first-arrivals are the input data in first-arrival traveltime tomography or early-arrival waveform inversion, while they are considered as noise in reflection tomography or migration. However, there are always some random background noises which are needed to be removed as well as other coherent noises such as multiples that may need to be eliminated. After de-noising, there are different approaches to build the final image of the subsurface, which can be broadly categorized as follow.

Scale separation based approaches

In reflection seismology, the reflected waves are generated by reflectors corresponding to the velocity or impedance contrasts between lithological or structural units (Fig. 1.5a). The attribute which describes these contrasts is generally referred to as the reflectivity (Fig. 1.5c). In terms of scale or resolution, these reflectors can be viewed as the short-wavelength components of the subsurface. A geometrical picture of these reflectors is useful for interpreters to identify the main lithological units and the deformation they underwent (faulting, folding, ...). The imaging process that generates such pictures is called migration, this denomination capturing the idea of moving the waves recorded at the surface to the reflectors that produced them (see Etgen et al. (2009) and Biondi (2006) for recent reviews of the different kinds of migration methods and Claerbout (1976), Berkhout (1985) and Yilmaz (2001) for some historical references). On the other hand, the traveltimes of the waves that propagate from the source or the receiver positions to the reflectors are mostly sensitive to the long wavelengths of

the wavespeed distribution (Fig. 1.5b), which are generally referred to as the velocity macro-model. This implies that an accurate velocity macro-model is necessary to correctly position in depth the reflectors by migration. When the spread of the receiver layout is of insufficient length to record reflected waves beyond the critical incidence (Fig. 1.6a), it is well acknowledged that there is spectral gap between the wavenumber component of the velocity macro-model and that of the reflectivity (Fig. 1.6b), namely, the intermediate wavenumbers belong to the null space of the imaging process (Claerbout, 1985; Jannane et al., 1989). In other words, with such acquisition, it is not possible to build a model of the subsurface which embeds a continuous broadband of wavelengths, from the mean property up to a high-cut wavenumber which would be controlled by the maximum temporal frequency of the seismic source. This explains why a conventional reflection seismic imaging workflow alternates two main tasks within an iterative process: the velocity macro-model building by a velocity analysis techniques (traveltime tomography, migration-based velocity analysis, ...) followed by migration to build the reflectivity. Generally, the output of one of these two tasks is used as an input of the other one. This two-step approach has been referred to as the sequential approach by Biondi and Almomin (2013).

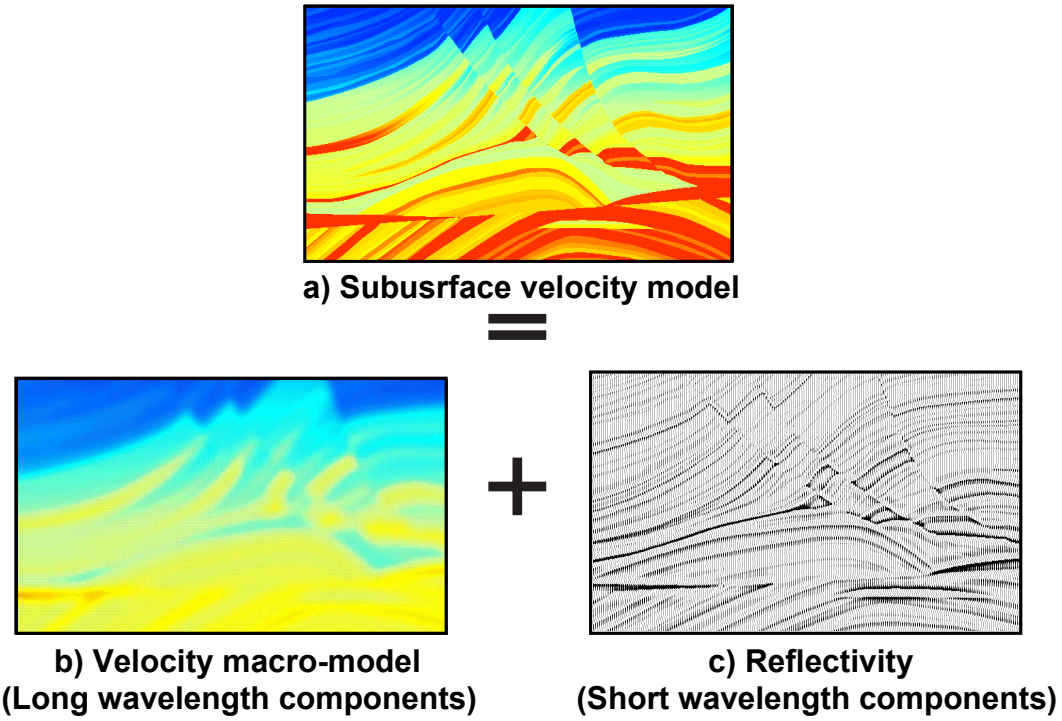

Figure 1.5 – Scale separation. The subsurface velocity model can be split into the velocity macro-model and reflectivity which include the long wavelength and short wavelength components of the subsurface velocity model, respectively (figure adapted form Chauris (2000)).

In conventional workflows the velocity analysis relies on maximizing the stacking power of reflection hyperbolas in common midpoint (CMP) gathers, Fig. 1.7, (Yilmaz, 2001). This

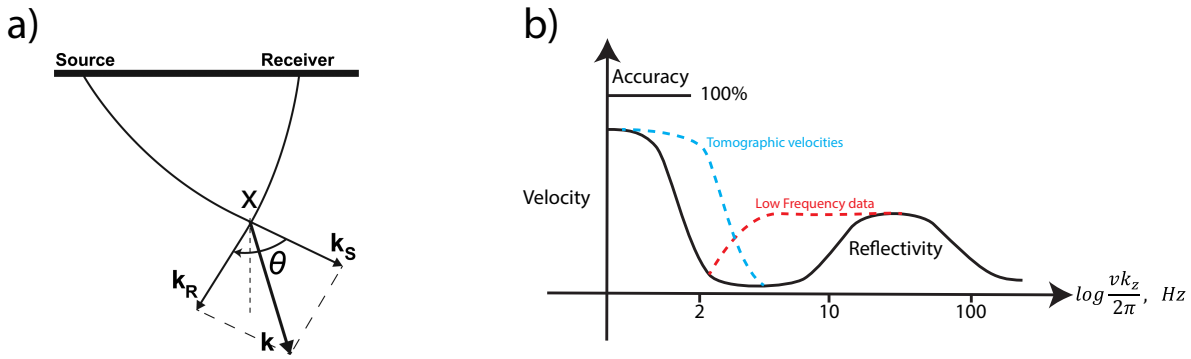

Figure 1.6 – Scale separation in seismic imaging. a) Relationship between the scattering angle θ and the local wavenumber at the diffractor point X. The wavenumber component $\mathbf{k}$ injected at the diffractor point is related to the local wavelength λ of seismic wave and the scattering angle θ through the expression $\|\mathbf{k}\| = \|\mathbf{k_s} + \mathbf{k_r}\| = \frac{2}{\lambda}\cos(\theta/2)$. In the seismic acquisitions with a short offset range (compared to the depth of target), only small θ are sampled if the background velocity be smooth. Therefore, only the high spatial frequency (short wavelength) components of the subsurface can be retrieved. In order to retrieve the long wavelength components, one needs to consider the propagation paths which involve the large scatterer angles. This raises the issue of scale separation. b) Schematic presentation of scale separation for a specific diffractor in (a) where the background velocity macro-model and reflectively include low and high frequency content of the subsurface model, respectively. In broad-band wide-azimuth seismic acquisitions, the medium is sampled with lower spatial frequencies (smaller θ) which narrow the gap between two scales (red curve in diagram). Also, high resolution tomographic approaches push upward the high end of the short wavelength component of the velocity (blue curve in diagram). Figure is adapted from Claerbout (1985); Biondi and Almomin (2013).

simple time processing generally results in smooth velocity models in time which can be used for time migration of stacked section (post-stack time migration). But, developments in seismic acquisition, computational power and theoretical aspects of the methods enabled us to retrieve more details of subsurface. This required a more careful and complicated imaging workflow. Among different methods for retrieving the velocity macro-model are reflection traveltime tomography (Aki et al., 1977; Bishop et al., 1985; Farra and Madariaga, 1988) and migration-based velocity analysis (MVA) (Gardner et al., 1974; Al-Yahya, 1989). Tomographic approaches seek to minimize the misfit between the picked traveltimes of the recorded reflected waves and the modelled traveltimes, through the resolution of a numerical optimization problem.

MVA methods broadly speaking seek to optimize the focusing of the energy in the prestack migration domain. The governing idea is that a seismic acquisition always generates some kinds of redundant information (Fig. 1.8), this redundancy being necessary to estimate wavespeeds and solve the velocity-depth ambiguity. This redundancy implies that the dimension of the data space (a space allowing a full parametrization of the acquisition) is always higher than of the

model space (a space allowing a full parametrization of the subsurface). For example, a 2D towed-streamer acquisition can be described by three attributes: the source position, the offset and time (Fig. 1.8a), while the subsurface is described by two parameters, depth and distance. One can build as many images of the reflectivity as available surface offsets to generate a migrated volume in the prestack domain (prestack means that the velocity analysis is performed before generating the final migrated image by stacking all the offset-dependent partial migrated images). For one fixed distance, some offset-depth panels, referred to as common-image gathers (CIGs), can be generated (Fig. 1.8b). If the velocity model that has been used to perform the migration is accurate, the reflectors should be flat in these panels as each class of offsets should reveal the same subsurface structure. Indeed, if these reflectors are flat in the CIGs, their stack will optimally focus them in the final migrated image.

Therefore, the aim of the migration-based velocity analysis is to minimize the residual move-out of the reflectors in the CIGs (Fig. 1.9). Some of MVA methods require picking of the residual moveouts in prestack migration volume, while others more automatic methods, that are recast as waveform inversion (Symes (2008) for a review) such as differential semblance optimization (DSO) (Symes and Carazzone, 1991; Symes, 1998; Chauris and Noble, 2001), migration-based traveltime inversion (MBTT) (Clément et al., 2001) or reflection waveform inversion (RWI), Fig. 1.10, (Xu et al., 2012; Brossier et al., 2015; Zhou et al., 2015), do not require this tedious picking task, although a careful preprocessing of the gathers is always necessary to filter out coherent noise and other stretching effects.

The RWI technique takes advantage of the reflectivity of subsurface (acquired by migration with an initial velocity model) to introduce the reflector as the secondary sources in the depth and inject the low wavenumber components at the diffractor points, Fig. 1.10. This allows updating the velocity macro-model by emphasizing the transmission paths of the reflected waves. Since the initial velocity is incorrect and consequently the reflectors are not well positioned, one needs to perform again migration and update the reflectivity. By alternating these steps in an iterative process, the workflow converges to the final subsurface model.

Scale integration based approaches

Recent developments in broad-band, wide-azimuth and long-offset seismic acquisitions made the gap between the discussed scales narrower. Broad-band sources push back the limits of low and high parts of the frequency bandwidth such that reflectivity and the velocity bands are extended at their low and high ends, respectively (Biondi and Almomin, 2013) (the red line in Fig. 1.6b). Moreover, long-offset acquisitions make recording of deeply-penetrating diving waves and post-critical reflections possible. These waves are typically transmitted across the heterogeneities during their propagation, leading to a forward-scattering type interaction (by opposition to back-scattering generated by reflection). This transmission regime from the source to the receiver combined with broadband data provides a suitable framework to reconstruct the intermediate wavelengths of the subsurface by waveform inversion techniques such as full waveform inversion (FWI) (see Tarantola (1984b), Lailly (1984), Gauthier et al. (1986), Mora (1987), Mora (1989) Mora (1988), Mora (1988) and Pratt (1999) for original publications and Virieux and Operto (2009) for a more recent review) and narrow the scale deficit mentioned

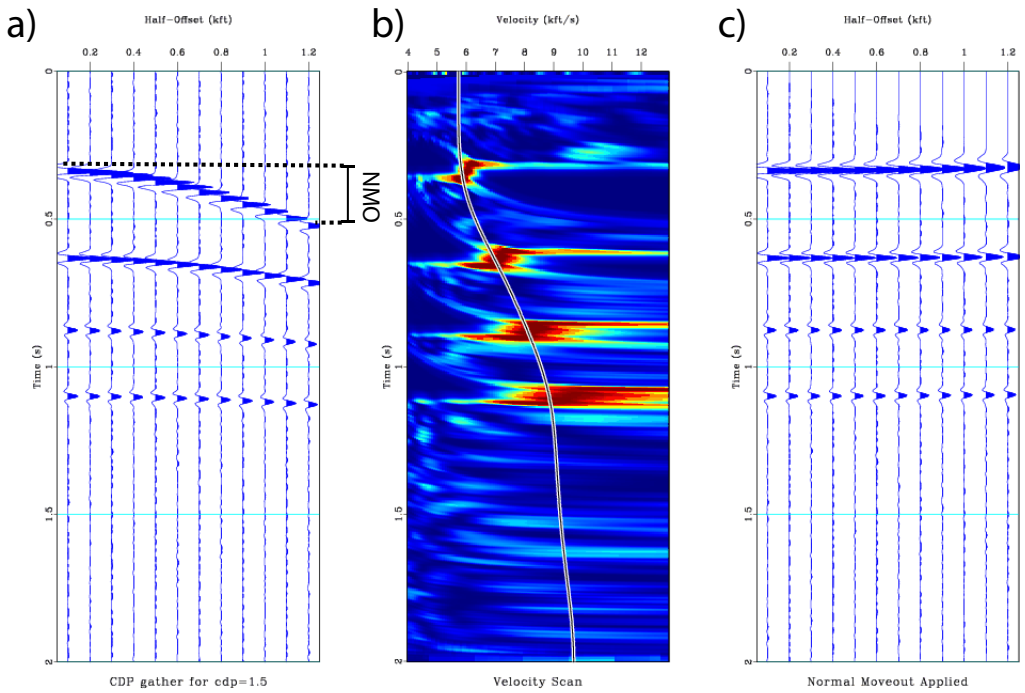

Figure 1.7 – Velocity analysis based on maximizing the stacking power in CMP gathers. a) A CMP gather for a layering media with four flat layers (no multiple). b) Velocity scan panel which shows the stack power of hyperbolas for different velocities. c) The CMP gather in (a) after NMO correction (figure from Madagascar development blog).

in the previous section accordingly. In other words, gathering all of the information carried out by diving waves, pre- and post-critical reflections makes in theory non-sequential or fully-integrated waveform inversion approaches possible for the simultaneous update all of the wavelength components of the subsurface up to a cut-off wavenumber controlled by the maximum frequency of the source. This prompts for example Mora (1989) to define these waveform inversion techniques under the form of the equation: inversion = tomography + migration. The issue with such waveform inversion techniques, beyond their computational cost which can limit the maximum frequency injected in the inversion, is their nonlinearity (the classical least-squares difference-based misfit function to be minimized contains many local minima). There are currently three main approaches to circumvent this issue: the first ones attempt to design new FWI misfit functions based on optimal transport ((Métivier et al., 2016c,a,b), matching filter (Luo and Sava, 2011; Warner and Guasch, 2016; Zhu and Fomel, 2016), correlation (van Leeuwen and Mulder, 2008) or instantaneous phase and envelope (Bozdağ et al., 2011; Luo and Wu, 2015) which exhibit better convexity properties. These approaches generally provide velocity model with a more limited resolution than classical FWI. However, these velocity models can be used as improved starting velocity models for classical FWI. The second one tend to add an extra dimension in the model space to expend the search space beyond the local-minima valleys

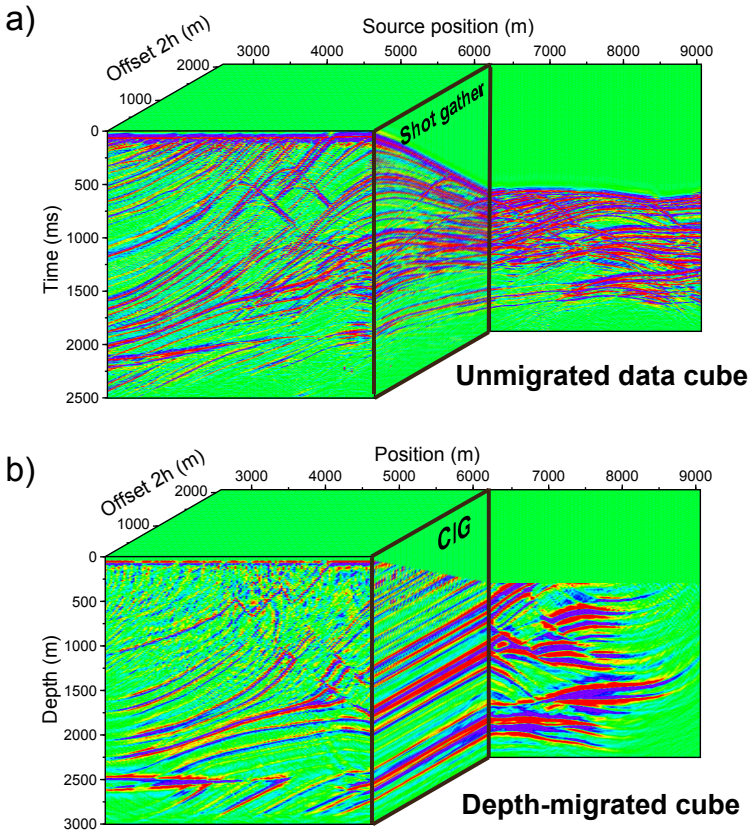

Figure 1.8 – A 2D dataset cube a) before migration b) after depth migration. Seismic acquisitions generate redundant data. For instance in (a) the recorded data in a 2D acquisition is presented with three attributes: source position, offset, time. In the migrated cube (b) the CIGs are the offset-depth panels extracted for a fixed position on the surface. If the migration velocity be a close estimation of the subsurface velocity, the CIGs includes flat reflectors (figure adapted form Chauris (2000)).

(Symes, 2008; van Leeuwen and Herrmann, 2013). The third one aims to improve the starting model for FWI through the development of new traveltime tomography methods, which are amenable to exploit a dense picking of high-frequency data (e.g. Lambaré, 2008; Woodward et al., 2008). The work carried out in this PhD thesis falls into the third category and seeks to develop a novel tomographic approach based on traveltimes and slopes of locally-coherent events.

To conclude this review of the different categories of techniques for subsurface imaging, it is worth mentioning that even the most modern long-offset acquisition devices do not allow one to record those diving waves which penetrate down to the deepest targeted structure. Therefore, while a fully integrated FWI approach can be sufficient to build a high-resolution velocity model in the upper part of the subsurface, one has to come back to the sequential approach to build the velocity model from reflected waves in the deep part. In the framework of waveform inversion

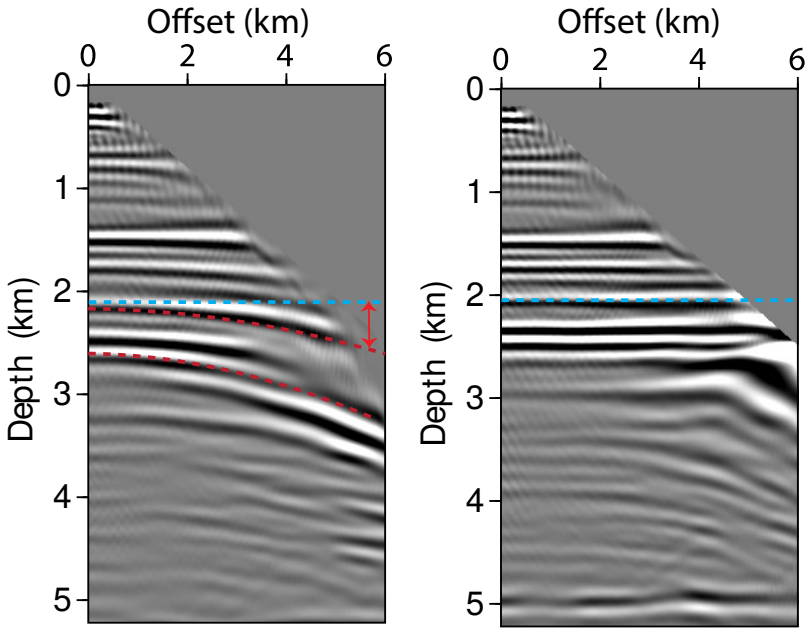

Figure 1.9 – Flatness of the reflectors in CIG. A CIG from a realistic data set after depth migration with a) incorrect velocity model and b) a close approximation of true velocity model. The red arrow designates the non-zero residual moveout (RMO). The migration-based velocity analysis aim at minimizing the residual moveouts for the reflectors (figure is adapted form Zhou (2016)).

techniques, reflection waveform inversion (RWI) mentioned in the previous section (Xu et al., 2012; Brossier et al., 2015) should be viewed as an adaptation of FWI to perform velocity model building from reflections only. Consistently with the sequential strategy, the governing idea is to use a reflectivity image as a prior information that is used as secondary sources to emphasize the transmission paths of the reflected waves in the sensitivity kernel of the inversion at the expense of the interactions of the wave with the reflectors (migration isochrones). Conversely, the upstream reflectivity (or impedance) image can be built by a least-squares reverse time migration which turns out to be equivalent to a waveform inversion of the short-spread reflections. Again, these two processes (impedance imaging followed by macro-velocity mode building) are alternated with an iterative workflow. An extension of RWI was recently proposed by Zhou et al. (2015) such that the diving waves can be included consistently with the reflected waves in the velocity-model building. The governing idea is that the diving waves will contribute to improve the reconstruction of the shallow part where they propagate, this improvement allowing for a more accurate reconstruction of the deep structure by reflection waveform inversion. It is important to stress that, in the upper part where both diving waves and reflected waves propagate, the long-wavelength information carried out by these two kinds of waves (diving and reflected waves) are complementary and not redundant because diving waves mostly propagate sub-horizontally while reflections propagate sub-vertically. Therefore, they do not see the same spectral components of the subsurface (see Mora (1989) for an illustration of the spectral

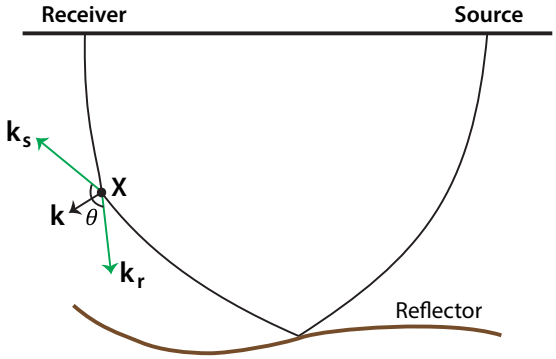

Figure 1.10 – Background velocity updating in the reflection waveform inversion. The reflectivity model acquired by migration with an initial velocity model, plays the role of secondary sources in the depth and provides a large scattering angle for the sampling of diffractor "X". Therefore, the wavenumber component $\|\mathbf{k}\| = \|\mathbf{k_s} + \mathbf{k_r}\| = \frac{2}{\lambda}\cos(\theta/2)$ injected at the diffractor point includes the low spatial frequencies which allows updating the velocity macro-model in the background. This emphasizes the transmission path of the reflected wave. This complements the discussion of Fig. 1.6 where only the reflectivity was updated through narrow-azimuth seismic data (see more details in Brossier et al. (2015)).

coverage achieved by different wavepaths).

In this thesis, I develop the traveltime and slope tomographic method for reflected waves. However, I will discuss in the perspective section the extension of the method to the joint inversion of first-arrival and reflection traveltimes and slopes, hence defining the tomographic counterpart of the waveform inversion approach of Zhou et al. (2015).

1.2 Velocity macro model building

The challenge of velocity macro-model building has motivated development of an impressive number of techniques. These methods can be classified according to different criteria such as the domain in which inversion is performed (unmigrated domain, migrated domain, extended migrated domain), the need or not of picking times or extracting attributes, the kinds of observables that feed the inversion (traveltimes, slopes, waveforms), the kinds of acquisition and related arrivals used by the inversion (towed-streamer versus stationary-receiver acquisitions, reflected waves versus diving waves) and the optimization criterion that drives the inversion toward the subsurface model (semblance, differential semblance, correlation, difference, double difference, optimal transport, matching filter) with many possible bridges between these different approaches.

In the following we review the bases of reflection traveltime tomography, classic stereotomography, and non-linear slope tomography. Through these reviews we try to cover the main principles of the approach developed in this thesis for velocity macro-model building.

1.2.1 Reflection traveltime tomography

Refection traveltime tomography as a tool to estimate the kinematic attributes of the subsurface (wavespeeds or related anisotropic dimensionless parameters such as the Thomsen's parameters (Thomsen, 1986)) was introduced as an alternative to stacking velocity analysis to improve the resolution of the velocity models. Resolution of the stacking velocity analysis is in the order of receiver array spread while traveltime tomography resolution is of the order of Fresnel zone width (Williamson, 1991). The stacking velocity analysis relies to a large extent on the assumption of laterally-homogeneous media. Moreover, stacking velocity analysis performs the velocity estimation in the time domain and, hence, requires some time-to-depth conversion trough for example the Dix formula for depth imaging. This approach can mainly recognize anomalies when their seismic response can be followed along the receiver spread, while each reflection time can separately influence the velocity update in reflection tomography. Therefore, dipping reflectors and the local anomalies do not violate any assumption in reflection tomography.

However, reflection tomography relies on the picking of primary reflections which can be a tedious task. On the other hand, the primary reflections generally might not always provide enough illumination of the full subsurface target. For example, multiple reflections from a deep reflector and the flanks of a salt body can be helpful to image them. These intrinsic limitations promote two ways out: 1) applying other methods which take into account the full waveform (no need to pick the observables) and both kinematic and dynamic information, 2) try to improve the basis of reflection tomography and overcome some of the limitations by reformulating the method.

Although some limitations of reflection traveltime tomography such as relying on the primary reflections seem unavoidable, there have been many successful developments of this technology, such as pre-stack depth migration (PreSDM) velocity analysis and slope tomographic approaches with possible connections with these two categories of methods. Since the central aim of this PhD is to develop a slope tomographic method, I review now in more details the principles of ray-based traveltime reflection tomography before introducing the main concepts of slope tomography.

Classic reflection traveltime tomography

Here, picking a collection of reflection traveltimes following laterally-coherent events is the first step. Picking is conventionally performed in the unmigrated pre-stack time domain (e.g. CMP gathers) (Bishop et al., 1985; Farra and Madariaga, 1988). These traveltimes are related to laterally-coherent or continuous events on seismic gathers, which can be interpreted as reflections from subsurface reflectors.

After picking, one can subdivide the subsurface into rectangular cells as in Fig. 1.11. Classic reflection tomography uses the observed traveltimes of the picked reflections for each offset to retrieve the velocity in each cell through solving a system of linear equations. For the simple example in Fig. 1.11 for which straight rays are assumed, the travel path from the source to

the receiver is formed by linear segments crossing the cells. Under this crude assumption, the observed traveltime for a given receiver is given by

$$T = \frac{d_2}{v_2} + \frac{d_8}{v_8} + \frac{d_{14}}{v_{14}} + \frac{d_9}{v_9} + \frac{d_4}{v_4} + \frac{d_5}{v_5}. \tag{1.1}$$

Here, d_j is the travelled distance by the ray in the jth cell which can be inferred by performing ray tracing from the associated source and receiver to the reflector (or vice versa), and v_j is the velocity of the jth cell. Reflection tomography seeks to find the velocity model that best fits the observed traveltimes. A similar equation can be written for all the picked events and offsets and the relationship between traveltimes

$$T_i = \sum_j d_{ij}/v_j = \sum_j d_{ij} s_j, \tag{1.2}$$

where T_i is the observed traveltime along the ith ray path (which is associated to a specific CMP and offset), d_{ij} is the travelled distance by ith ray in jth cell with velocity v_j or slowness s_j (inverse of velocity v_j). This equation can be recast in matrix form as

$$\mathbf{T} = \mathbf{DS}. \tag{1.3}$$

Here, vectors $\mathbf{T}_{N\times 1}$ and $\mathbf{S}_{M\times 1}$ include the observed traveltimes for N rays and M slowness cells, respectively, and matrix $\mathbf{D}_{N\times M}$ (sensitivity matrix) consists of the associated travel distance by rays in each cell. In real applications since each ray is not crossing all the cells, the matrix $\mathbf{D}$ is sparse (i.e. many of its coefficients are zero). The solution of this equation in a least-square sense is given by

$$\mathbf{S} = (\mathbf{D}^t\mathbf{D})^{-1}\mathbf{D}^t T. \tag{1.4}$$

For small problems the slowness in each cell can be calculated with direct solver (namely, Gaussian-elimination techniques). However, for large scale problems, iterative solvers such as those based on LSQR (Paige and Saunders, 1982) are classically used to tackle realistic case studies.

In realistic settings, the depth of the reflectors are unknown and the relationship between traveltime and slowness is non linear because the ray paths depend on the velocity model. This leads to development of non-linear tomographic algorithms (Bishop et al., 1985; Farra and Madariaga, 1988) in which one solves the following optimization problem in a least-squares sense

$$\min_{\mathbf{m}} \|\mathbf{t}_{obs} - \mathbf{t}_{clc}(\mathbf{m})\|_2^2. \tag{1.5}$$

Here, the unknown slownesses and the reflector depths are gathered in vector $\mathbf{m}$, and $\mathbf{t}_{obs}$ and $\mathbf{t}_{clc}$ are the observed and the calculated traveltimes by ray tracing, respectively. Note that, the calculated traveltime is a function of the subsurface slowness and the reflector depths. Zeroing the gradient of the misfit function (which is the necessary condition for the minimizer of the misfit function), leads to

$$\mathbf{J}^t(\mathbf{t}_{obs} - \mathbf{t}_{clc}(\mathbf{m})) = \mathbf{0}, \tag{1.6}$$

where $\mathbf{J} = \partial \mathbf{t}_{clc}/\partial \mathbf{m}$ is the so-called sensitivity or Fréchet derivative matrix. Through an

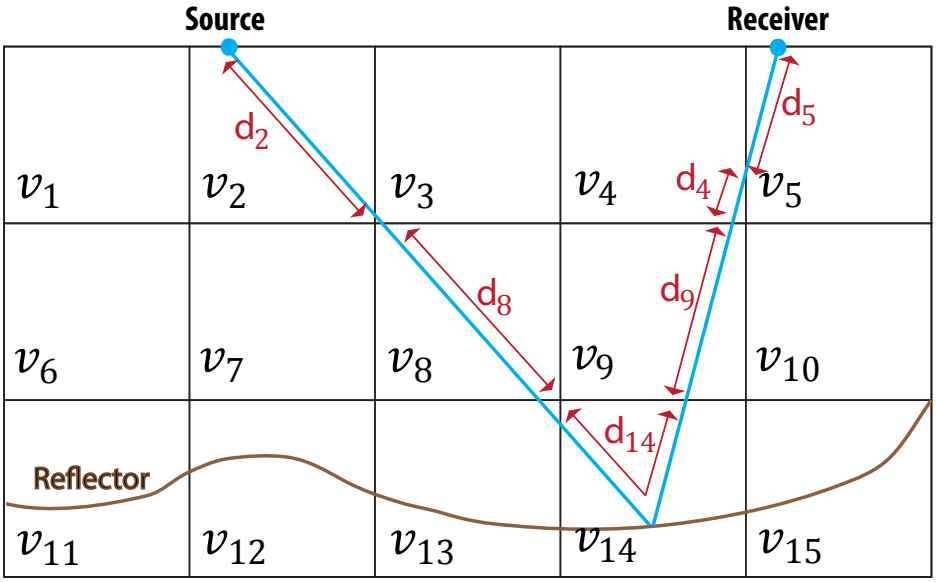

Figure 1.11 – Classic ray-based tomography. The subsurface is described by several velocity cells. The blue lines represent schematically the ray which samples the reflector at depth. Red arrows are the travelled distance by ray within each velocity cell. In real application the rays should respect the Snell's law where crossing the border of each velocity cell (figure adapted from Jones (2010)).

iterative algorithm starting from an initial guess for the subsurface slowness and the reflector depths, one can find the solution of this equation which fits the observed traveltimes.

In order to clarify this procedure, consider iteration k in which $\mathbf{m}_k$ is known and the desired $\mathbf{m}_{k+1}$ is unknown. If $\mathbf{\Delta m} = \mathbf{m}_{k+1} - \mathbf{m}_k$, by applying the first-order Taylor's expansion around $\mathbf{m}_k$, one can write

$$\mathbf{t}_{clc}(\mathbf{m}_k + \mathbf{\Delta m}) = \mathbf{t}_{clc}(\mathbf{m}_k) + \mathbf{J}_k \mathbf{\Delta m}, \tag{1.7}$$

where the Fréchet derivative matrix $\mathbf{J}_k$ is defined as $\mathbf{J}|_{\mathbf{m}=\mathbf{m}_k}$. Condition (1.6) implies $\mathbf{J}_k^t(\mathbf{t}_{obs} - \mathbf{t}_{clc}(\mathbf{m}_{k+1})) = \mathbf{0}$. Inserting the expression (1.7) in this condition leads to

$$\mathbf{\Delta m} = \left[\mathbf{J}_k^t \mathbf{J}_k\right]^{-1} \mathbf{J}^t \mathbf{\Delta t}. \tag{1.8}$$

Therefore, with the calculated perturbation $\mathbf{\Delta m}$, the updated $\mathbf{m}_{k+1}$ can be achieved through

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \mathbf{\Delta m}. \tag{1.9}$$

The Fréchet derivative matrix in each iteration is calculated based on the Fermat's principle [1] of least time for rays which shows the derivative of the traveltime with respect to the

1. According to the modern Fermat's principle in optic (Ghatak, 2005, section 2.1), "*the actual ray-path between two points is the one for which the optical path length is stationary with respect to variations of the path*".

slowness in a velocity cell (Fig.1.11) is approximately the travelled distance by the ray in that cell. Derivative of the traveltimes with respect to the depth of reflectors can be obtained in a close form directly form ray tracing (Bishop et al., 1985, their appendix B).

PreSDM tomography

In real applications, reflection tomography can be performed in a more reliable way in the (prestack) depth migrated domain because the picking is generally easier in the migrated domain; energy in migrated images are more focused and signal to noise ratio is higher. In the following, I briefly introduce the main principles of PreSDM tomography, Fig. 1.12.

In Fig. 1.12(a) the subsurface is discretized in a comparable manner to the classic approach in Fig. 1.11 but there are many sophisticated interpretative discretization schemes in which the subsurface can be divided into layers (by considering the geology of the area under exploration) and within each layer consider a fine grid.

After PreSDM with an initial guess for the subsurface velocity and calculation of CIGs [2], an automatic volumetric picking in the CIGs and common-offset migrated sections is performed in order to picks the continuous events and the geological dip for common imaging points (CIPs), respectively. For an accurate enough velocity model, the depth of a CIP should be same for all of the offsets (i.e. a flat continuous event in the associated CIG)(Fig. 1.12c). Therefore, the PreSDM tomography consider the flatness of the events in CIGs as a criterion to measure the accuracy of the velocity model. In case of performing the migration with wrong velocity model, there is no alignment for the picked events in CIGs (Fig. 1.12b). This deviation of the CIP depth for different offsets is called the residual moveout (RMO). Therefore, the flatness of the events in CIGs is equivalent to the zero RMOs (Fig. 1.12b-c). Considering the zero RMO as the criterion for accuracy of velocity models, the PreSDM tomography formulation reads as follow.

If the true subsurface slowness $\mathbf{s}$ be the perturbed version of initial slowness $\mathbf{s}_0$, one can write

$$\mathbf{s} = \mathbf{s}_0 + \mathbf{\Delta s}, \qquad \mathbf{\Delta s} \ll \mathbf{s}. \tag{1.10}$$

Therefore, the true depth of a picked RMO at offset h in a CIG can be defined as

$$z_h(\mathbf{s}) = z_h(\mathbf{s}_0 + \mathbf{\Delta s}), \tag{1.11}$$

where through the first order Taylor's series expansion of relation (1.11) around the initial slowness one can conclude

$$z_h(\mathbf{s}) = z_h(\mathbf{s}_0) + \frac{\partial z_h(\mathbf{s}_0)}{\partial \mathbf{s}} \mathbf{\Delta s}. \tag{1.12}$$

Stork (1992) derived the relation between the changes in source-receiver traveltime t_h for offset

A less accurate but original statement is: "*the actual path between two points taken by a beam of light is the one which is traversed in the least time.*"

2. CIGs can be calculated in both offset and angle domains (i.e. offset and reflection angle of rays which sample a CIP, respectively). For sake of brevity here we only consider the offset domain.

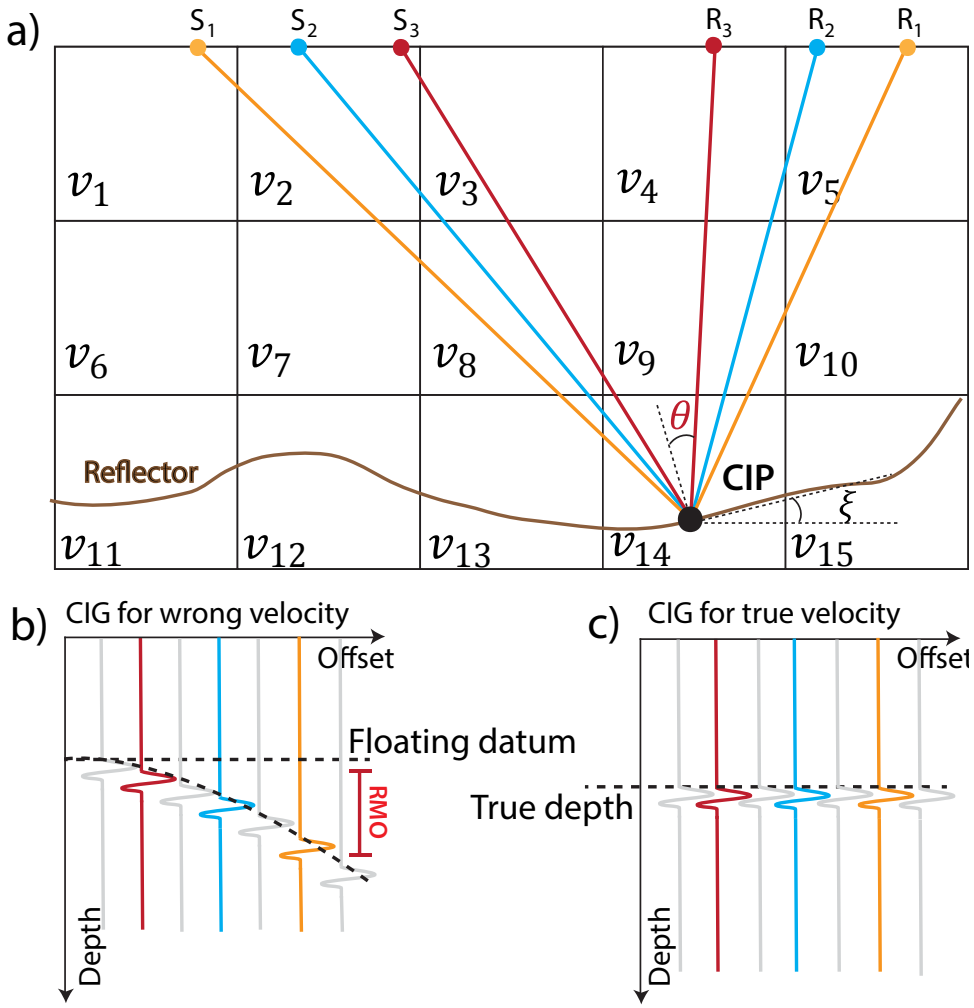

Figure 1.12 – Pre-SDM tomography. a) Subsurface is divided into several cells and a CIP is sampled with different rays. θ is the incident angle for the short-offset ray sampling the CIP and ξ is the dip of the reflector at the CIP. The associated offset domain CIGs after migration with b) a wrong velocity model and c) the true velocity model. Migration with wrong velocity model results in the different CIP depths at different offset. RMO for each offset in (b) is defined as the CIP depth deviation respect to a datum. Migration with true velocity model results in zero RMO for all the offsets (figure adapted from Jones (2010)).

h and the changes of picked RMO depth as

$$\frac{\partial z_h}{\partial t_h} = \frac{v_r}{2\cos\theta\cos\xi}, \tag{1.13}$$

where v_r is the local slowness above the reflector at the CIP. By applying the chain rule of

derivative, eq. (1.12) can be written as

$$z_h(\mathbf{s}) = z_h(\mathbf{s}_0) - \sum_j (\frac{\partial t_h}{\partial s_j} \Delta s_j) \frac{v_r}{2 \cos\theta_h \cos\xi}, \tag{1.14}$$

where $\partial t_h / \partial s_j$ is the change in traveltime (associated to the rays which sample the reflector at offset h) with respect to the change in slowness at cell j. $\partial t_h / \partial s_j$ is the linear tomography equation (eq. (1.2)) and can be calculated through ray tracing and the assumption that a perturbation of the subsurface properties does not change the ray path. θ_h is the incidence angle of the associated rays with offset h and ξ is the reflector dip at the corresponding CIP (Fig. (1.12)a). In eq. (1.14), perturbation $\mathbf{\Delta s}$ is the unknown which PreSDM tomography aims to retrieve by minimizing the RMOs.

By considering a floating datum such as the picked RMO depths at near-offset ($z_0(\mathbf{s}_0)$) of migration by slowness $\mathbf{s}_0$, one can define the minimization of the RMOs as the minimization of relative depth error $\Delta z_h = z_h(\mathbf{s}_0) - z_0(\mathbf{s}_0)$ rather than a static depth error (Woodward et al., 2008). By applying this criterion, the relative depth error for all the picked RMOs in CIGs is acquired. Considering eq. (1.14), this results in the following linear equation

$$\mathbf{L\Delta s} = \mathbf{\Delta z}. \tag{1.15}$$

Here the coefficients of the sensitivity matrix $\mathbf{L}$ depend on the ray incidence angles, local dips at CIPs and the velocities. The PreSDM tomography is formulated as a minimization problem,

$$\min_{\boldsymbol{\delta s}} \|\mathbf{L\Delta s} - \mathbf{\Delta z}\|. \tag{1.16}$$

In section 3.1 I shall review some methods which can be used to solve such a minimization problem. After solving this minimization problem the slowness $\mathbf{s}_0$ can be updated. Though a PreSDM with updated slowness one can check the flatness of the events in CIGs. If they are not flat, a new picking is required to repeat the PreSDM tomography. This loop continues to flatten the CIGs event. Fig.1.13 illustrates this PreSDM tomography workflow.

Generally, it is needed to apply some regularizations or smoothing on PreSDM tomography inversion to mitigate the ill-posedness of the problem. During the iterations the smoothing window length can be shorten to increase the resolution of the updated velocity model.

Resolution of the reflection tomography depends on the seismic wavelength, acquisition geometry, and the complexity of the subsurface. Generally, when no multi-pathing [3] and caustics are generated during ray tracing and the picked RMOs are noise free, increasing the pick densities increases the vertical resolution until the velocity layers stop generating pick-able RMO (Woodward et al., 2008). The horizontal resolution depends on the distance between the CIGs. However, CIG distances smaller than the Fresnel zone of dominant frequency wave-path can not increase the horizontal resolution. In shallow part of the subsurface, where the velocity is slower, the Fresnel zone is narrower and this increases the resolution. While, in deep parts of the subsurface, higher velocities result in wider Fresnel zones and consequently a lower resolution (Woodward et al., 2008). If the size of velocity anomaly is smaller than the Fresnel zone

3. Multi-pathing occurs when more than one ray-path connect one source and a subsurface point.

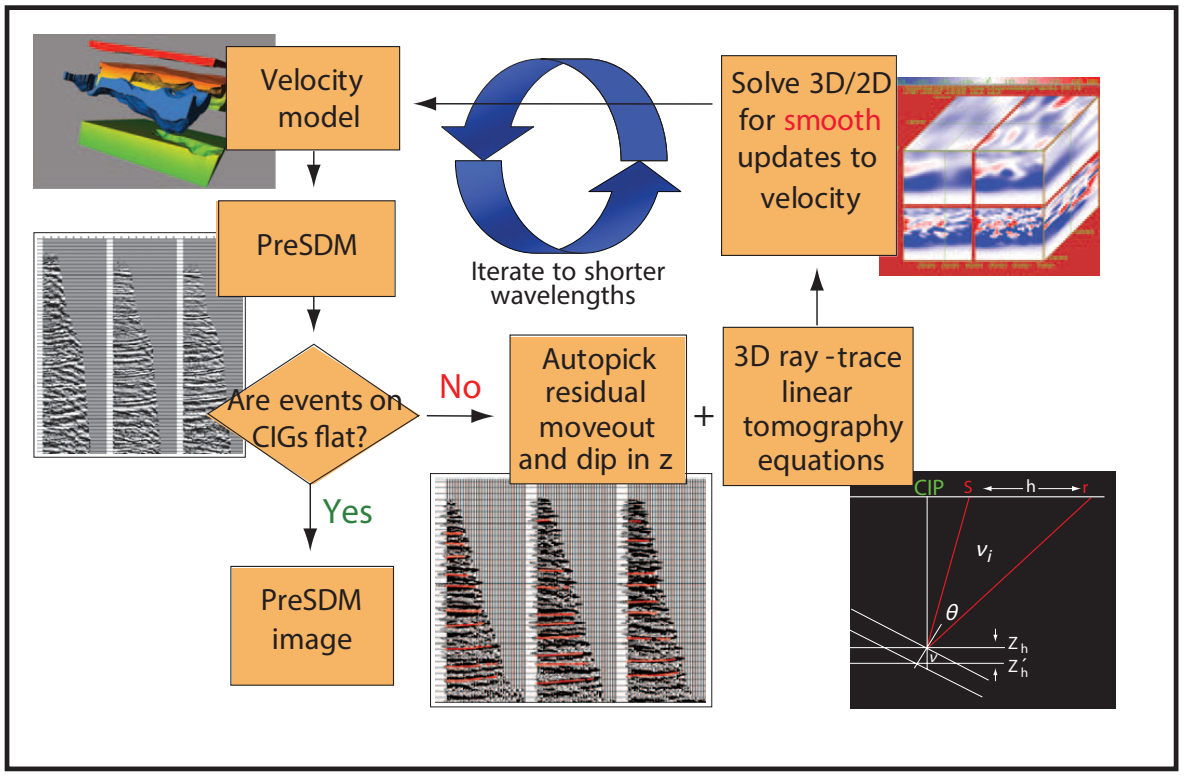

Figure 1.13 – PreSDM tomography workflow. Perform a Pre-SDM on data with an initial velocity. Through an automatic picking tool pick the RMOs on CIGs and reflector dips on the stacked section. Perform ray tracing to find the associated source-receivers for the picked RMO at each offset. Then, calculate linear equations relating the perturbations in RMO to the velocity perturbations (eq. 1.15 or 1.16). With acquired velocity perturbations, update the initial velocity. Repeat the loop to flat the CIP gathers. In each iteration the smoothing window length can be reduced to insert higher frequency (shorter wavelength) of the subsurface in the inversion (figure from Woodward et al. (2008)).

size, the seismic waves generates diffraction. In this case traveltime tomography can not resolve the anomaly and waveform tomographic approaches (such as FWI and diffraction tomography), which consider the amplitude information, become necessary (Jones, 2010).

Although there are numerous successful applications of reflection traveltime tomography, still the picking step is challenging in the sense that it requires to follow some continuous events. In the following, I review the bases of slope tomography, which is a remedy to this issue.

1.2.2 Two-dimensional classic stereotomography

Stereotomography is a slope tomographic tool for velocity macro-model building (Billette, 1998; Billette and Lambaré, 1998). This tool is based on the controlled directional reception

(CDR) method (Riabinkin, 1957; Riabinkin et al., 1962; Sword, 1986, 1987; Biondi, 1992), and it has been introduced as an alternative to reflection traveltime tomography. The key idea behind slope tomography is that, instead of tracking continuous events, we pick locally-coherent events associated to reflections or diffractions. Each local event is described by its two-way traveltime and its slopes in the common-shot and common-receiver gathers. In this section, I review the CDR and stereotomography which is the basis of our current research. In the following all the discussions are based on the primary reflections.

In order to explain the slope tomographic approaches, I first define the concept of slope. Slowness vectors (gradient of traveltime) are perpendicular to the wavefronts and their norm are equal to the inverse of wave propagation velocity. This vector can be interpreted in the seismic gathers. Footprint of a reflection or diffraction on recorder seismic gathers is a series of coherent events. These events show the variation of traveltime of the arrival wavefronts in the acquisition direction. This means, for a horizontal receiver array, the slope of a local coherent event (in shot gathers) is the horizontal component of the slowness vector of the arrival wave at receiver position. If r be the central receiver of the local coherent, the associated slope is called p_r. Considering the reciprocity rule between source and receiver, one can define p_s on a common receiver gather which is the slope of a local coherent event at source s (Fig. 1.14).

A reflecting or diffracting point in the subsurface [4] generates a pair of local coherent events in shot and receiver gathers. This pair can be fully parameterized by five quantities $(T_{sr}, p_s, p_r, \mathbf{s}, \mathbf{r})$: the two-way traveltime of the event, the slopes of the event in the shot and receiver gathers and the associated central source and receiver coordinates, respectively. These information can be also extracted uniquely from common offset and common midpoint gathers. The corresponding slopes in these domains are $\frac{p_s + p_r}{2}$ and $p_r - p_s$, respectively.

In order to illustrate how slopes constrain the velocity updates, let's consider a scatterer in a medium of unknown velocity which generates local coherent events recorded in the shot and receiver gathers with $(T_{sr}, p_s, p_r, \mathbf{s}, \mathbf{r})$. One performs ray tracing from the source and receiver toward the scatterer (or vice versa) for an initial guess of the subsurface velocity. The velocity in the background is true if the calculated traveltime and the slopes at the source and receiver be equal to (T_{sr}, p_s, p_r) (Fig. 1.15a). Without considering the slope information, there is no control on the slowness vector. In other words, there are many scatterers along the isochrone (iso-traveltime surface) which satisfy the observed T_{sr} (Fig. 1.15a). From a different point of view, the velocity in the background is correct if two descending rays from the source and receiver positions with the horizontal components of slowness p_s and p_r, respectively, intersect at the scatterer position after total traveltime T_{sr} (Fig. 1.15b). This forms the basics idea behind the CDR method.

In the context of the CDR method, the picked traveltime and slopes of locally-coherent events in the shot and receiver gathers form the observables. In this method, for each locally-coherent event, two rays are propagated in the velocity model from the source and receiver (descending rays, Fig. 1.15b). The take-off angles of the rays are given by the observed slopes. The sum of the two ray traveltimes is calculated at different depths, and, the ray tracing is stopped when this sum equals to the observed traveltime. If the velocity is the true one, the end-point of

4. Here and in the following I call this point a scatterer or scattering point.

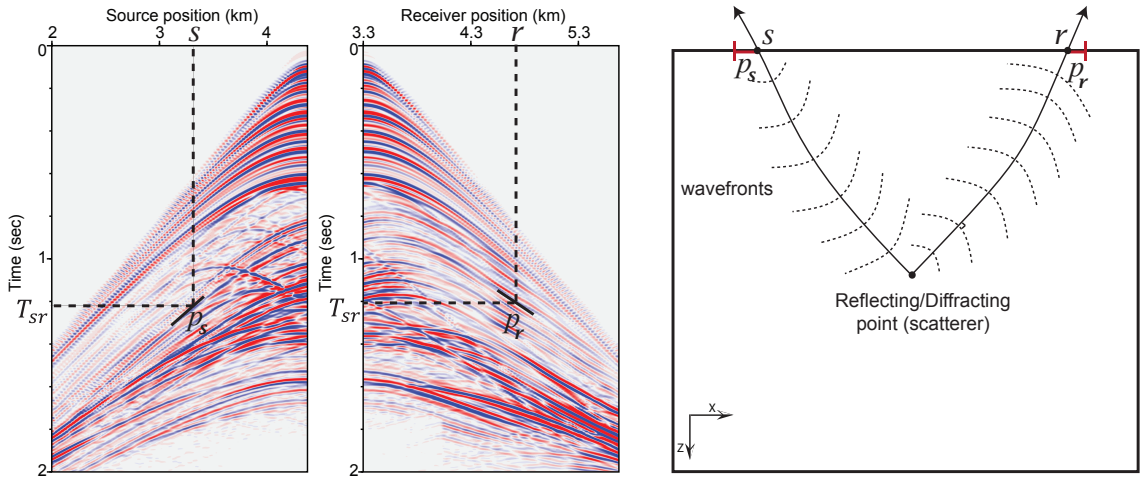

Figure 1.14 – Local coherent events in stereotomography. Definition of horizontal component of the slowness vector at the source and receiver positions in ray theory, right panel, and recorded shot and receiver gathers, left panels. Each local coherent event which is generated by a primary reflection/diffraction can be associated to a reflection/diffraction point (scatterer). The local coherent events are fully characterized with the related central source and receiver positions $(\mathbf{s}, \mathbf{r})$, slopes (p_s, p_r) and two-way traveltime T_{sr} (figure adapted from Billette (1998)).

the two rays coincide (i.e. $x_{error} = 0$). Otherwise, the velocity is updated to reduce the misfit function $\sum_i^{picks} \|x^i_{error}\|$. This can be achieved by linking the velocity perturbation to x_{error} and solving a non-linear optimization problem. Billette (1998) recast the CDR method into a more general formulation of slope tomography in which he took advantage of the Hamiltonian formulation of the ray theory (Farra and Madariaga, 1987; Lambaré et al., 1996), paraxial ray theory (Červený et al., 1977; Farra and Madariaga, 1987) and general inverse problem theory (Tarantola, 1987).

Model and data space in stereotomography

Stereotomography as an inverse problem consists of a data and model spaces. A semi-automatic volumetric picking in shot and receiver gathers generates the stereotomographic data set $\mathbf{d}$ as

$$\mathbf{d} = [(\mathbf{s}, \mathbf{r}, T_{s,r}, p_s, p_r)_n]_{n=1}^{N}, \tag{1.17}$$

where N is the total number of the picked values and $\mathbf{s}$ and $\mathbf{r}$ are the position of the associated source s and receiver r for the picked traveltime and slopes values.

Billette (1998) performed ray tracing from the scatterers in depth toward the sources and receivers in a given velocity model to model the data. The two rays are traced with a prior guess of the take-off angles θ_s and θ_r and are stopped when their traveltimes equal to the prior guess of the one-way traveltimes T_s and T_r. Since there is no guarantee that these two rays end at the

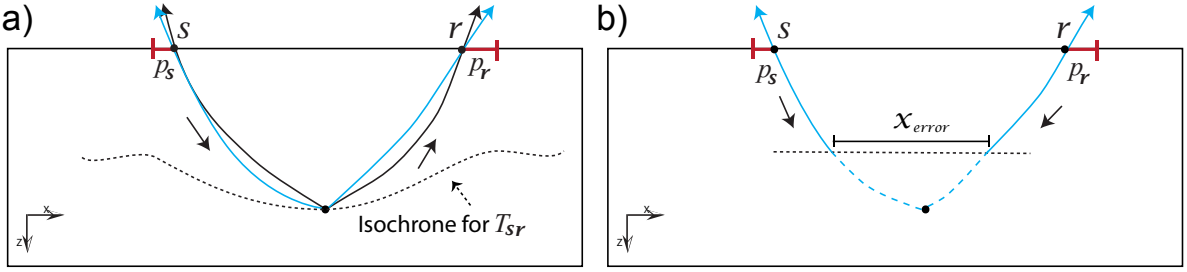

Figure 1.15 – The role of slopes in tomography. a) For a give scatterer at depth the velocity is correct if the slopes at the source and receiver and the traveltime be equal to the observed values. Here the black and blue rays (generated by different velocities in the background) satisfy the observed traveltime but only the blue one satisfies the slopes. b) Two rays are propagated from the source and receiver with the observed slopes. They cross each other (i.e. $x_{errtr} = 0$) at the scatterer true position if the velocity in the background be true (this is the basis of CDR method). At this position the summation of traveltime for each ray should be the observed traveltime.

source and receiver positions, respectively, as they should, the source and receiver coordinates need to be incorporated in the data space. Alternatively, one may use a more computationally-demanding two-point ray tracing to remove on the one hand the one-way traveltimes and the take-off angles from the model space and on the other hand the shot and receiver coordinates from the data space.

By parametrizing the background velocity on M cardinal cubic B-spline nodes with coefficients $[c_m]_{m=1}^{M}$, the above implementation of the stereotomography forward problem leads to the following model space

$$\mathbf{m} = \{(\mathbf{x}, \theta_s, \theta_r, T_s, T_r)_n|_{n=1}^{N}, [c_m]_{m=1}^{M}\}. \tag{1.18}$$

Here the calculated data can be formulated as $\mathbf{d}_{clc} = g(\mathbf{m})$ where the non-linear operator g stands for the modelling operator. This definition of the model and data spaces provides the most general definition for the inverse problem in slope tomography.

Resolution of the inverse problem relies on the minimization of least-squares misfit between picked ($\mathbf{d}_{obs}$) and modelled data ($\mathbf{d}_{clc}$) as

$$\min_{\mathbf{m}} \quad \frac{1}{2}(\mathbf{d}_{obs} - g(\mathbf{m}))^t \mathbf{C}_d^{-1}(\mathbf{d}_{obs} - g(\mathbf{m})) + (\mathbf{m} - \mathbf{m}_{prior})^t \mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}_{prior}), \tag{1.19}$$

where $\mathbf{m}_{prior}$ is the *a priory* model, and $\mathbf{C}_d$ and $\mathbf{C}_m$ are the *a priori* covariance matrix in the data and model space, respectively. The superscript t stands for transpose operator. This non-linear minimization problem can be solved with local optimization methods such as the Newton method, where the parameters are iteratively updated along the Newton descent direction until

convergence. The update model at iteration $k+1$ is given by

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \left(\frac{\partial^2 C}{\partial \mathbf{m}^2}(\mathbf{m})\right)^{-1} \frac{\partial C}{\partial \mathbf{m}}(\mathbf{m}), \tag{1.20}$$

where the descent direction is the product of the inverse of the Hessian (the second derivative of the misfit function) and the steepest descent direction (the opposite of the gradient of the misfit function). By introducing $\mathbf{G}_k$ as the Fréchet derivative (sensitivity matrix) of the data with respect to the model parameters at iteration k,

$$\mathbf{G}_k = \left[\frac{\partial g(\mathbf{m})}{\partial \mathbf{m}}\right]_{\mathbf{m}=\mathbf{m}_k} \tag{1.21}$$

the updates in eq. (1.20) can be written as

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \left(\mathbf{G}_k^t \mathbf{C}_d^{-1}(\mathbf{d}_{obs} - g(\mathbf{m}_k))\mathbf{G}_k + \mathbf{C}_m^{-1}\right)^{-1} \mathbf{G}_k^t \mathbf{C}_d^{-1}(\mathbf{d}_{obs} - g(\mathbf{m}_k)) + \mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}_{prior}). \tag{1.22}$$

Here the Fréchet derivative matrix is a sparse matrix (Billette, 1998) and the non-zero elements can be calculated by paraxial ray tracing (Červený et al., 1977).

Resolution of this multi-parameter inverse problem requires coping with its non-linearity and ill-posedness. Among the different approaches that address this issue are: Billette et al. (2003) proposed a multi-scale approach in which they initialize the velocity model on a coarse grid of cardinal B-spline nodes and increase the node numbers during the iterations to insert progressively shorter wavelengths in the updated velocity model. Alerini et al. (2007) proposed a hierarchical approach to insert the picked values in the optimization process according to their traveltimes. Gosselet and Bégat (2009) combined the borehole and stereotomographic picks to overcome the velocity-depth ambiguity in the inversion.

Beside the improvements in handling the inversion, there are different approaches which try to improve the quality of the picking procedure. A stereotomographic data set with less noisy picks that are well distributed in depth, facilitates convergence of inversion toward a reliable solution. Chauris et al. (2002a); Nguyen et al. (2008) proposed to perform picking in the depth-migrated domain, Lavaud et al. (2004); Neckludov et al. (2006) in the post-stack time domain, and Lambaré et al. (2007) in the pre-stack time migrated domain. These domains provide more focused local coherent events (higher signal to noise ratio) and this results in less erroneous picks by automatic picking tools. Picking in depth-migrated domain greatly improves the control on depth distribution of the picks and makes their interpretation and QC easier.

Chauris et al. (2002a,b) established the link between the local coherent events in the pre-stack depth migrated domain and their counterpart in the pre-stack unmigrated time domain. In other words, they established a link between MVA methods and stereotomography. This relation allows development of new non-linear slope tomographic methods such as Guillaume et al. (2008) where picking is performed in the migrated domain while the inverse problem is performed with data-domain observables (two-way traveltimes and source and receiver slopes). This approach does not require to perform migration at each iteration. In the following I discuss the principles of these approaches.

1.2.3 Non-linear slope tomography through kinematic invariants

I first review the link between local coherent events in the pre-stack depth migrated domain and the unmigrated (or data) domain. Then I introduce some recent approaches of non-linear slope tomography.

Local coherent events in pre-stack migrated depth vs. pre-stack unmigrated time domain

In the context of MVA methods, Chauris et al. (2002a) replaced the criterion of flattening the continuous events in CIGs with zeroing the RMO of local coherent events in CIGs. A local coherent event in 2D pre-stack migrated domain is described by five parameters (Fig. 1.16): the position and offset of the central trace in the common offset gather (COG) and CIG, x and h respectively, the depth of the local event (z), the apparent geological migrated dip ξ (measured in the COG), and the residual slope φ (measured in the CIG). Since we seek to minimize the residual slope φ in the migrated domain, a natural definition of the cost function is given by

$$C(\mathbf{s}) = \frac{1}{2} \sum_{picks} (w \tan \varphi)^2, \tag{1.23}$$

where w stands for a weighting factor to be determined and $\mathbf{s}$ is the slowness.

An equivalent definition in the data domain would be

$$C(\mathbf{s}) = \frac{1}{2} \sum_{picks} (p_s^* - p_s)^2, \tag{1.24}$$

where p_s^* and p_s stand for the observed and calculated slopes in the receiver gather at the source s position, respectively (note that from the picking in depth domain there is no direct access to the slope values). This misfit function involves only one observable (p_s^*) because the two focusing equations, Chauris et al. (2002a, their equations 10 and 11)[5] resulting from the migration imaging principle constrain the two-way traveltime and the receiver slope during the inverse problem. From the two focusing equations, Chauris et al. (2002a) found the relationship

$$\tan \varphi = \frac{p_s^* - p_s}{2s \cos\theta \cos\xi}, \tag{1.25}$$

where θ is the incidence angle for the ray connecting the source to the local coherent event at depth. Therefore, the weighting factor in the cost function (1.23) is $w = 2s \cos\theta \cos\xi$, and the

5. In the framework of migration of common shot gathers, focusing equations imply that the summation of the one-way traveltimes of the rays which sample a scatterer should be equal to the picked traveltime and horizontal component of the slowness vector at receiver should be equal to the observed slope at the receiver position. The ray traced from the receiver position with the observed slope intersects the isocrhone defined by the picked two-way traveltime at the scatterer position. The ray connecting this scatterer position and the shot position reaches this latter with a slope that does not match the picked slope if the velocity model is not correct, justifying the misfit function given in eq. (1.24).

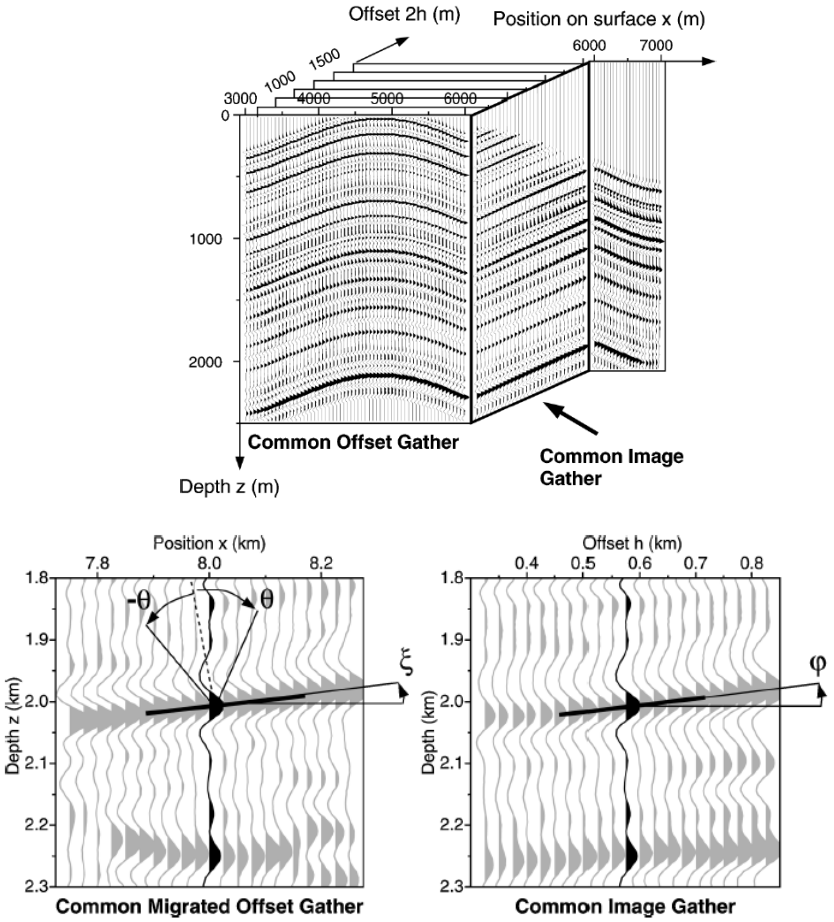

Figure 1.16 – Local coherent event in pre-stack depth migrated domain. Here each local coherent event is characterized by (x, z, h, ξ, φ): position of the central trace in the COG, depth of the local event, offset of the central trace, apparent geological dip and residual slope, respectively (figure adapted from Chauris et al. (2002a))

cost function is finally given by

$$C(\mathbf{s}) = \frac{1}{2} \sum_{picks} (w \tan \varphi)^2 = \frac{1}{2} \sum_{picks} (p_s^* - p_s)^2. \tag{1.26}$$

In this equality expression, the last term makes it physically meaningful to derive the gradient of cost function through the paraxial ray tracing.

It is important to note that, compared to stereotomography, the data space has been reduced to one observable class (the source slope) and the model space has been reduced to only one parameter class, namely, velocity (or slowness), which can contribute to mitigate the ill-posedness of the inverse problem. Here, the optimization involves a loop over these three steps: 1) performing PreSDM and compute the CIGs and COGs, 2) picking local coherent events for

measuring φ and ξ, respectively, 3) calculation of the gradients and the velocity update.

Slope tomography based on kinematic invariants

Guillaume et al. (2008) took advantage of the described relation between the depth and time picks to introduce a more efficient velocity model building tool. They used migration/demigration of local coherent events and the concept of kinematic invariants to estimate RMOs after the velocity update performed in each inversion iteration. This means there is no need to perform PreSDM in each iteration.

They first perform a PreSDM with an initial velocity model to build the migrated cube and pick the local RMOs in CIGs and apparent geological dips in the COGs. Then they perform specular offset ray tracing upwards from the picked event positions (this is kinematic demigration) while the background velocity is same as the one for the performed PreSDM. This leads to the associated kinematic invariants for each pick, i.e. source and receiver position, slopes at the source and receiver positions and two-way traveltime. These are called invariants because they do not depend on the velocity model as long as migration and demigration are performed in the same velocity model.

Here the cost function is defined as the $l_2$norm of the weighted residual slopes φ in Fig. 1.16 (such as eq. 1.23) and its gradient with respect to the velocity can be calculated by paraxial ray tracing (similar to Chauris et al. (2002a)). The gradient provides the necessary information to update the velocity model. In order to compute the cost function, calculation of φ is required. This can be achieved by performing kinematic migration for kinematics such that the kinematic invariants traveltime, source and receiver position and slope at the receiver position are fitted (no fit for the slope at the source) (Montel et al., 2010). Then by measuring the difference between the calculated (by kinematic migration) and observed (by kinematic demigration) slopes at the source position one can retrieve φ through eq. (1.25) and calculate the cost function through the relation (1.23). Therefore, the local RMOs are achieved without performing PreSDM. This procedure is repeated to satisfy the convergence criterion. A similar approach can be applied for the picks from pre-stack time migrated domain with performing demigration in time (Lambaré et al., 2007). In chapter 3 and 4 I shall mention some other applications regarding this technique.

All the discussed methods in this introduction, are introduced for isotropic assumption for the medium. In the following I introduce the anisotropic definition in exploration seismology and its footprint on the recorded seismic data.

1.3 Seismic anisotropy

The seismic anisotropy in a elastic media can be defined as the dependency of seismic wave velocity to the propagation direction. Here, the seismic waves include all types of seismic waves (e.g. S-waves) with arbitrary frequencies. Seismic anisotropy implies existence of a certain structure on the scale of seismic wavelength (Tsvankin, 2001a). In other words, seismic

anisotropy, similar to seismic heterogeneity, is scale-dependent; a medium can be heterogeneous and anisotropic for a short wavelength while it is homogeneous and isotropic for larger wavelength (Helbig, 1994).

Since the intrinsic anisotropy of the material in microscopic and submicroscopic scales rarely persist over distances comparable with seismic wavelength, the observed anisotropy effect on seismic record should be due to other elements. For the sedimentary sequences the anisotropy cause mainly relates to the following issues (Thomsen, 1986; Tsvankin, 2001a)

— intrinsic anisotropy due to preferred orientation of anisotropic mineral grains or the shapes of isotropic minerals;

— preferred orientation of the shape of isotropic minerals (such as flat layering platelets);

— thin bedding of isotropic or anisotropic layers (the layers may be horizontal or tilted);

— preferred orientation of the fractures or micro-cracks.

Each of these ordering arrangements has axial symmetry and the subsurface anisotropy formation can be a combination of these factors. Each anisotropic symmetry is characterized by a specific structure of the stiffness matrix [6], with the number of independent elements decreasing for higher-symmetry systems (Tsvankin, 2001a). The least symmetric system is the triclinic system with 21 independent elements in the stiffness tensor. The most symmetry is for the transversely isotropic (TI) media (hexagonal symmetry) with only 5 independent elements in the stiffness tensor. The TI media are the result of thin (compared to the dominant wavelength) layering in which the symmetry axis can be vertical (VTI), tilted (TTI) (Fig. 1.17a-b) or horizontal (HTI). For instance, VTI and TTI media can adequately describe the elastic properties of the shale in sedimentary basins and flanks of the salt domes (Tsvankin, 1997), respectively. HTI is the simplest representation for cracked media where an isotropic medium contains a single system of parallel vertical circular cracks (small fractures and micro-cracks). In case of non-circular cracks, existence of the second crack system, and presence of anisotropy or layering in the background media reduce the symmetry to orthorhombic or less (Tsvankin, 1997)(Fig. 1.17c).

Anisotropy has different effects on the seismic wave propagations such as: velocity of wave propagation is not same in all the directions, wavefronts are non-spherical, rays are not perpendicular to the wavefronts, displacement vectors are not parallel (in P-waves) and perpendicular (in S-waves) to the wavefronts (Helbig, 1994). These influence the recorded data in different ways. For instance, due to anisotropy the two shear waves with different polarizations arrive at the receiver with different times (shear wave splitting), conflict between depth estimation for a reflector with different waves and non-hyperbolic reflections events in CMP gathers (Helbig, 1994). Due to these differences between body waves in isotropic and anisotropic media, the three body-waves in anisotropic media are called quasi P-wave (qP) and quasi sheer waves (qS1, qS2) where quasi indicates that these waves only have a resemblance to the isotropic P and S-waves (Crampin, 1981).

Most of studies on anisotropy in exploration seismology are focused on the symmetric anisotropy such as TI and orthorhombic media. This choice can be because of two reasons,

6. Refer to the section 2.2 for a brief review on the elastic wave equation and the role of stiffness tensor.

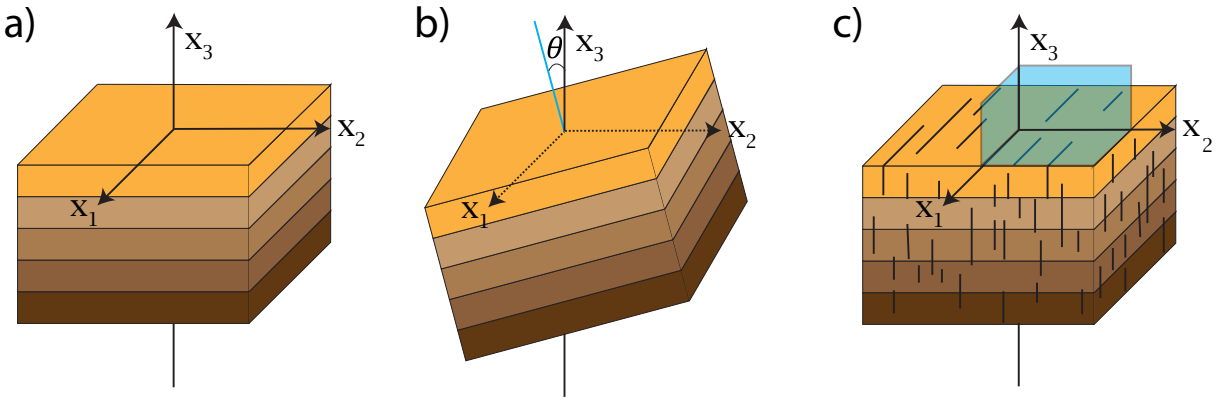

Figure 1.17 – Three symmetric anisotropic system: a) VTI with a vertical symmetry axis (x_3), b) TTI with a tilted symmetry axis (the blue line), and c) an orthorhombic media caused by parallel vertical cracks in a layering background with three mutually orthogonal planes of mirror symmetry (blue planes).

first, the formulation, parametrization and analysis of the wave propagation in these media are far simpler than in general anisotropic one, second, in many applications these assumptions about the subsurface properties are close enough to the reality and satisfy the exploration expectations. Moreover, the intrinsic limitations of surface seismic data (e.g. limited resolution specially at depth) prevent recognition of the different sources of anisotropy (e.g. cracks, layering, lithology, etc) for a general anisotropic media. Since in this study we deal with 2D TTI media, in the following I describe the parameters which describe such a medium.

In the context of depth imaging techniques, the most used parameterization for a 2D TTI medium is presented by Thomsen (1986) in which the medium is parametrized by vertical velocity (v_v) (velocity along the symmetry axis), two dimensionless anisotropic parameters ϵ and δ and the tilted angle θ of the symmetry axis (Fig. 1.17b). On the other hand, for time processing the introduced parametrization by Alkhalifah and Tsvankin (1995) is more interesting. They parametrized the medium with the NMO velocity and the dimensionless η parameter which give the expression controls the non-hyperbolic residuals of the reflection curves.

In order to visualize the effect of the anisotropy on the wave propagation, Fig. 1.18 shows the modelled wave propagation in elastic homogeneous isotropic (panel a) and VTI media (panel b). Here, the velocity in the isotropic medium is equal to the vertical wave speed of the VTI medium and for the VTI medium $\epsilon = 0.2$ and $\delta = 0.1$. In the isotropic case, the wave propagation velocity in all the directions is same (circular wavefronts), while in the VTI medium the non-circular wavefront implies different velocity in different directions. Here, δ has the leading role in the near-vertical directions and ϵ has dominant role for the horizontal propagations. By deviating from the principle axes the combination of both parameters controls the propagation speed and trajectory. Fig. 1.18(c-d) shows the calculated seismograms recorded at the surface. The footprint of anisotropy on longer offsets are more visible. The difference between the first arrival times for the isotropic and VTI media in zero offset seismograms is negligible while for

larger offsets the foot print of anisotropy is more evident (compare the first-arrival times at B and B'). This simple example shows that an accurate subsurface imaging in anisotropic media requires the anisotropic properties of the subsurface.

In the context of PreSDM tomography, this difference in arrival time, due to the anisotropy, affects the RMOs measurements. This means even with true vertical velocity in the background one can not flat the events in CIGs unless with a wrong depth. This non-flatness or mis-positioning of the events in CIGs result in less focused migrated section or wrong depth for the reflectors, respectively. In order to solve this problem, one needs to define the anisotropic properties among the tomographic model parameters. For instance, if one assumes that the subsurface is a 2D VTI medium, the model parameters for PreSDM tomography are (v_v, ϵ, δ). Adding these parameters increase the ill-posedness of the inverse problem and intrinsic limitations of the reflection data leads to ambiguity between these parameters. There are different methods to overcome these problems which I shall mention some of them in section 3.3. Also, in this section I shall discuss in detail the extension of slope tomographic approaches to anisotropic medium.

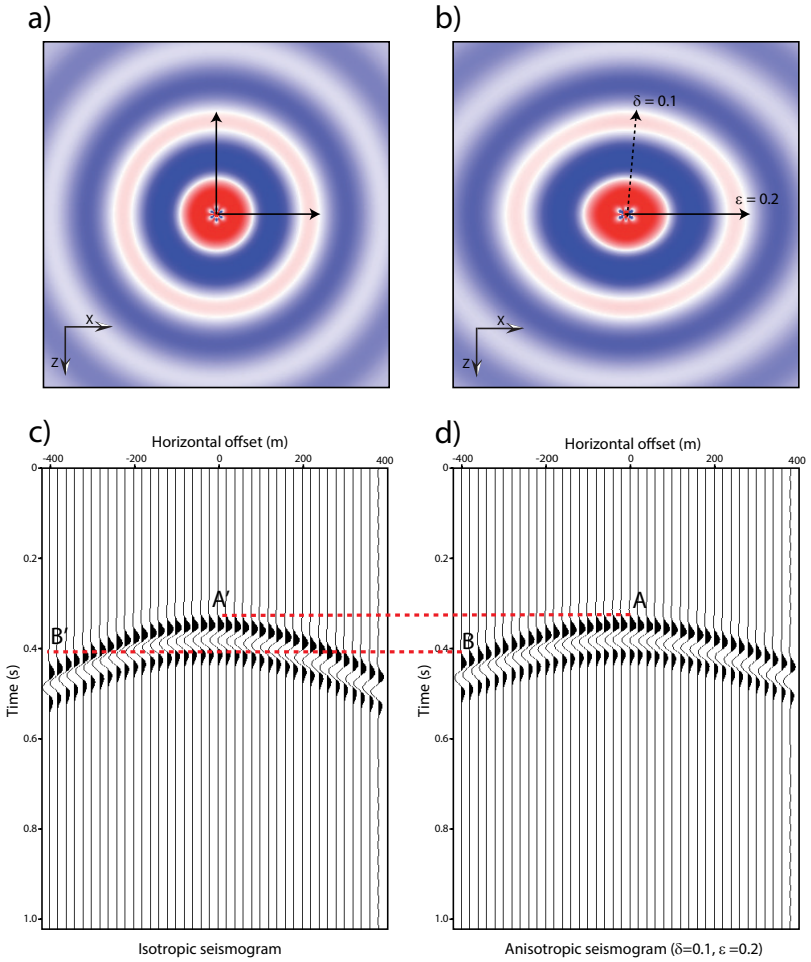

Figure 1.18 – Footprint of anisotropy on wave propagation. The pressure wavefield in elastic homogeneous a) isotropic and b) VTI media ($\epsilon = 0.2$, $\delta = 0.1$). The source is located at the center. The calculated pressure seismograms on the surface for c) elastic isotropic media (a) and d) elastic VTI media (b). At zero offset the anisotropy footprint is negligible (compare A and A') while for longer offsets (B and B') the difference between the first-arrival times is 10msec (figure from Gholami (2012)).

1.4 Motivation

In sequential seismic imaging workflows based on the scale separation between the long and short wavelength components of the subsurface, the velocity macro-model plays a major role in depth imaging by migration techniques or FWI. Among the widely used approaches to build this velocity macro-model are reflection traveltime tomography and MVA methods. Repetitive picking of continuous events in traveltime reflection tomography and the high computational cost of the MVA methods make implementation of these techniques for some applications challenging.

Slope tomography, as an alternative approach for velocity macro-model building, is attractive since it avoids the repetitive picking of continuous events and simultaneously takes advantage of the efficiency of the traveltime reflection tomography. The key idea is to use local coherent events in shot and receiver gathers and consider their slopes as the extra constraints during the velocity update. Stereotomography, as the most general formulation of slope tomography, proposes updating all the ray attributes associated with each local coherent event as well as the velocity in the background. This extends the search space during the inversion, the goal of which is to fit a set of observables (position of the associated source-receiver pair, two-way traveltime, slopes at the source and receiver positions). This extended search space can be helpful to absorb experimental errors or bypass some local minima but also generate a significant null space. A semi-automatic picking tool provides the stereotomographic data set, but low signal to noise ratio for the coherent events in the shot and receiver gathers can contaminate the stereotomographic data set with noise.

There are different remedies for the problem of picking in shot and receiver gathers which among them is picking in pre-stack depth migrated domain. This led to emergence of slope tomographic approaches based on the kinematic invariant calculations by demigration. Compared to the classic stereotomography [7], these methods eliminate the ray parameters from the model space which may mitigate the inversion ill-posedness and non-uniqueness of the solutions.

The classic stereotomography and other slope tomographic approaches rely on ray tracing to perform the forward problem. Although ray tracing gives an access to the ray attributes (one-way traveltimes and slowness vectors), they can suffer from some limitations in complex media (shadow zones, strong velocity gradients) resulting from non uniform ray-field sampling. The inversion is based on the explicit building of the Fréchet derivative matrix, which can be expensive and tedious to manipulate for large scale problems. Moreover, extension of these approaches to anisotropic media makes the formulation complicated in both the forward and inverse problems. The motivation of this thesis is to overcome these issues by reformulating the classic stereotomography in anisotropic media with eikonal solvers for the forward problem and the adjoint state method for the inverse problem. My approach will be referred to as the *anisotropic adjoint slope tomography*. Eikonal solvers are referred to numerical techniques to compute first arrival traveltime maps in computational mesh such as Cartesian grid in the case of this thesis (e.g. Vidale, 1988b; Podvin and Lecomte, 1991; Hole and Zelt, 1995; Noble et al., 2014). Therefore, the sampling of the traveltimes is fully controlled although the accuracy

7. Here and in the following chapters, "classic stereotomography" refers to the slope tomographic approaches which are based on Billette (1998); Billette et al. (2003) formulation for the inverse and forward problems.

and the relevance of the solution may need some quality control depending on the level of heterogeneity of the velocity model. The adjoint state method provides a matrix-free framework to compute the gradient of a functional (Plessix, 2006; Chavent, 2009). In this thesis, I fully develop this approach for 2D TTI media, implement it on massively-parallel computers and assess it with various synthetic experiments and a real data case study.

1.5 Thesis context

This manuscript is organized based on the published or under-publishing materials, and this led to some repetitive discussions about the bases of the methods.

Chapter 2; Adjoint slope tomography: Forward problem. In our approach the forward modelling engine relies on eikonal solvers. In chapter 2, I introduce an eikonal solver for 2D TTI media (Tavakoli F. et al., 2015). This solver is based on the fast sweeping method taking advantage of the fixed-point iteration technique to take into account the anelliptic terms in the TTI eikonal equation. We remove the point source singularity error by factorization technique. This results in an accurate and efficient approach for traveltime calculation. Through different examples such as BP TTI salt model I assess the accuracy of the proposed method.

Chapter 3; Adjoint slope tomography: Inverse problem. This chapter mainly consists of two parts: isotropic adjoint slope tomography and the extension to TTI media. In the first part (which is published in Tavakoli F. et al. (2016, 2017b)) I develop the reformulation of the isotropic classic stereography through the implementation of an eikonal solver for the forward modelling and the adjoint state method for the gradient calculation. Use of these numerical tools required to define new slope tomographic model and data spaces and derive the expression of the adjoint state equations. I assess the approach with several synthetic examples including the Marmousi model. In the second part, I extend the isotropic formulation to TTI media (Tavakoli F. et al., 2017a). This requires modifying the model parameters and solving the adjoint state equations associated with the TTI eikonal equation. Through two synthetic examples I evaluate the capability of the method in retrieving the anisotropic properties of the subsurface.

Chapter 4; Real data application: This chapter is devoted to the real data application. Processed recorded data and the picks were provided by Compagnie Générale de Géophysique (CGG). The picking is performed in the depth migrated domain. Through the proposed method, I update the vertical velocity while ϵ and δ are considered as passive parameters during the inversion. By performing reverse time migration and computing CIGs I assess the accuracy of the updated vertical velocity model. This chapter and the second part of the chapter 3 will be the basis of another paper to be submitted for publication.

Publications:

Tavakoli F., B., Ribodetti, A., Virieux, J., and Operto, S. (2015). An iterative factored eikonal solver for TTI media. In *SEG technical program expanded abstracts 2015*, volume 687, pages 3576–3581

Tavakoli F., B., Ribodetti, A., Operto, S., and Virieux, J. (2016). Adjoint stereotomography. In *SEG Technical Program Expanded Abstracts 2016*, pages 5269–5273

Tavakoli F., B., Operto, S., Ribodetti, A., and Virieux, J. (2017b). Slope tomography based on eikonal solvers and the adjoint-state method. *Geophysical Journal International*, 209(3):1629–1647

Tavakoli F., B., Operto, S., Ribodetti, A., and Virieux, J. (2017a). Anisotropic slope tomography based on eikonal solver and adjoint-state method. In *79th EAGE Conference and Exhibition 2017*

Chapter 2

Adjoint Slope Tomography: Forward problem

Chapter overview: In this chapter I introduce an efficient finite difference (FD) numerical algorithm for solving the 2D eikonal equation in a medium with vertical transverse isotropy (VTI) or tilted transverse isotropy (TTI). This solver will be used as the forward modelling engine for adjoint slope tomography. First, I review some basic concepts about wave equation and its ansatz solution in the frame of the high frequency approximation. I briefly discuss the resolution of the high frequency wave equation with ray tracing methods, and then I derive the eikonal equation for VTI and TTI acoustic media. In the last section I introduce our TTI eikonal finite-difference solver. In this solver, the fast sweeping method (FSM) is considered for solving non-linear partial differential equations alongside the fixed-point iterations for including the anelliptic term of VTI/TTI eikonal equation in an iterative scheme. Moreover, I mitigate the propagated errors due to point source implementation by using a factorization method removing the point source upwind singularity. Different numerical examples on TTI models show that incorporation of the iterative fixed-point scheme and the factorization method yields enough accurate solutions for long offsets and complex subsurface models. This chapter in a shorter format is published in SEG annual conference 2015 (Tavakoli F. et al., 2015). Coincidently, this method is also covered by Waheed et al. (2015) for 3D media.

2.1 Introduction

Some geological structures such as layered sediments and shale overlying salt flanks lead to mechanical anisotropy which strongly affects seismic wave propagation. Such media can be described as vertical transversely isotropic (VTI) and tilted transversely isotropic (TTI) models. Considering the crucial role of first-arrival traveltimes computation in seismic modelling and imaging, there is a strong need for accurate and efficient methods for computing first-arrival traveltimes in VTI and TTI media.

There are many methods dealing with computing the seismic wave first-arrival traveltimes, the two categories of the most popular methods being based on ray tracing and eikonal solvers. Some defects of ray tracing based methods, such as a complex formulation and numerical implementation, in blocky and anisotropic media, and presence of shadow zones in the ray fields in rapidly-varying structures, make the eikonal solvers a reliable alternative. However, because eikonal solvers naturally do not provide explicit information about the wave trajectories, still ray tracing remains the method of choice for many applications.

The idea of computing first-arrival traveltimes by solving the eikonal partial differential equation through a finite difference method, which estimates first-arrival traveltimes at fixed positions through an Eulerian formalism rather than by integration along ray paths through a Lagrangian counterpart, was first proposed by Vidale (1988a, 1990). The limitation of the Vidale's method in presence of sharp velocity contrast (Fig.2.1a) was overcame by Podvin and Lecomte (1991) through a new FD operator. They took advantage of Huygens' principles to consider different wave propagation modes, like transmitted, diffracted and head waves, in to FD discretization. Hole and Zelt (1995) overcame the Vidale's approach limitation and the costly calculation of Podvin and Lecomte (1991) through modifying the Vidale's finite difference scheme by incorporating the head wave operators and considering reverse propagations (Fig.2.1b). Then Afnimar and Koketsu (2000) defined a new 3D operators to cope with head waves multi-pathing problem arising by waves travelling along an irregular interface. Tracking wavefronts, rather than rays, allows for a better control on the sampling of the medium and may fill also shadow zones with diffracted wavefronts when considering first-arrival times.

To cite few contributions among many trying to acquire more accurate first-arrival traveltime solutions with Eulerian approaches, I can mention using the celerity domain (Pica, 1997; Zhang et al., 2005) or using an implementation of Lax-Friedrichs numerical Hamiltonian (Kao et al., 2004). Aside the Fast Marching Method (FMM) (Lelièvre et al., 2011), the Fast Sweeping Method (FSM) has gained popularity thanks to its highly efficient computational performance (Zhao, 2005; Noble et al., 2014; Belayouni, 2013) especially in heterogeneous media we encounter in the Earth.

In an Eulerian frame, the eikonal equation is a non-linear partial differential equation (PDE) with boundary conditions such as a source term which is often considered as a point singularity. As a result, the upwind source singularity introduces initial errors impacting the computation of traveltimes. A constant velocity zone around the source point could be considered (Zhang et al., 2006; Benamou et al., 2010) which may be a problem if the source be nearby an heterogeneity. A local refinement around the source point may mitigate this intrinsic problem (Kim and Cook, 1999) with subtle mesh structure for numerical implementation. Fomel et al. (2009) has proposed a rather simple and efficient method using a factorization of the traveltime as the product of an analytical known solution in a given reference medium and an unknown function introducing a new PDE to be solved.

Inserting anisotropic rheology in the eikonal equation leads to a rather complicated non-linear PDE. Many anisotropic eikonal solvers are restricted to simple anisotropic types such as elliptic anisotropy (Qian and Symes, 2001; Luo and Qian, 2012) or may suffer from drawbacks such as heavy computational cost (Eaton, 1993) with a strong possible decrease in accuracy especially at the source (Qian et al., 2007). Therefore, a fixed-point iterative technique has been

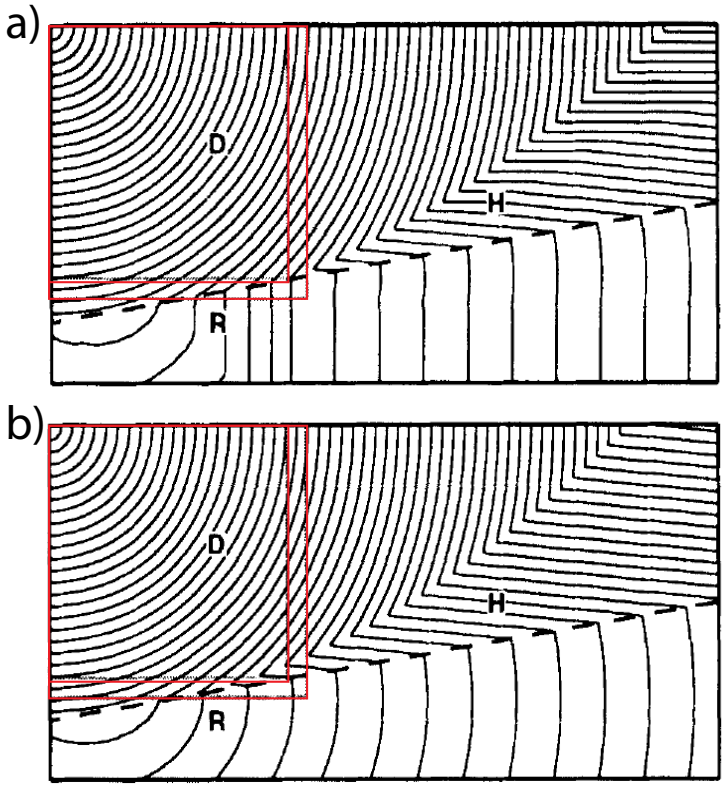

Figure 2.1 – First-arrival traveltime contours for a 2D velocity model with a sharp velocity contrast (a low velocity layer on a high velocity half-space). The labelled phases are direct waves (D), refracted waves (R) and critically refracted head waves (H). The red boxes show the expanding square used in finite difference scheme of Vidale (1988a). Calculated first-arrival traveltime contours with a) Vidale's original approach where the refracted waves are distorted, and b) modified Vidales's approach by Hole and Zelt (1995) where the head waves are calculated correctly (figure from Hole and Zelt (1995)).

proposed for improving the solution accuracy when considering complex anisotropic rheology while mitigating the computational cost (Alkhalifah, 2011; Ma and Alkhalifah, 2013; Waheed et al., 2014). Alternative approach to the fixed-point strategy has been proposed recently by Han et al. (2015, 2017). They define a local stencil which requires solving a quartic equation. Further comparisons should be performed for efficiency comparison. Also, Le Bouteiller et al. (2017) propose a novel approach using time-dependent Hamilton-Jacobi equation. Their solver is based on discontinuous Galerkin scheme which takes advantage of the factorization technique to calculate accurately first-arrival traveltimes in 2D TTI media.

In this study, I combine the factorized eikonal equation presented in Luo and Qian (2012) and the fixed-point iteration method designed by Waheed et al. (2014) in order to deal with high powers of spatial derivatives existing in the TTI eikonal equation. I propose a new 2D TTI eikonal solver that removes source singularity and computes accurately the first-arrival

traveltimes in complex TTI media.

In the following, I first briefly review the elastodynamic wave equation, and then I introduce the high frequency approximation for the anisotropic acoustic wave equation, from which I derive the eikonal equation. Finally, I present my proposed eikonal solver and assess its efficiency and accuracy for long-offset simulation in homogeneous VTI/TTI media and the heterogeneous TTI and BP TTI salt models.

2.2 Wave propagation

Seismic records are the result of the interaction between the seismic waves and the Earth structures, and brings to the surface some indirect information about the subsurface geology. Numerical solution of the seismic wave equation simulates this interaction, from which seismic imaging techniques seek to reconstruct the subsurface properties. Therefore, efficient numerical tools for seismic wave modelling form the cornerstone of many seismic imaging algorithms.

If we consider the Earth as a 3D elastic solid medium, the propagated seismic waves from an external source $f(\mathbf{x},t)$ can be formulated as (Červený, 2001; Chapman, 2004)

$$\rho\ddot{u}(\mathbf{x},t) - \nabla.\tau(\mathbf{x},t) = f(\mathbf{x},t), \quad \mathbf{x} \in \mathbb{R}^3, \quad \mathrm{t} \in [\mathrm{t}_0, \mathrm{t}_1] \in \mathbb{R}, \tag{2.1}$$

where the symbol " ¨" denotes the second order time derivative and ρ is the mass density model. Here, $u(\mathbf{x},t)$ and $\tau(\mathbf{x},t)$ stand for the displacement vector and the stress tensor at position $\mathbf{x}$ and elapsed time t, respectively. Based on the second Newton's law, this expression presents the balance between the momentum, internal forces and external forces.

The Hooke's law establishes a linear relation between the stress tensor τ and strain tensor ε as

$$\tau_{ij} = c_{ijkl}\varepsilon_{kl}, \tag{2.2}$$

where $i,j,k,l \in \{1,2,3\}$ and the Einstein convention is used for the summations. c_{ijkl} stands for the stiffness tensor and definition of the strain tensor ε reads

$$\varepsilon_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i}). \tag{2.3}$$

Here and in the following, spatial derivatives are compactly described by a comma in front of an indices, and for the sake of brevity I omit the coordinate dependency of c_{ijkl}. Because of the symmetric structure for stress and strain tensor (i.e. $\tau_{ij} = \tau_{ji}$ and $\varepsilon_{kl} = \varepsilon_{lk}$), the stiffness tensor is symmetric such that $c_{ijkl} = c_{jikl} = c_{ijlk} = c_{klij}$. Accordingly, I can rewrite eq. (2.2) as

$$\tau_{ij} = c_{ijkl}u_{k,l}. \tag{2.4}$$

By inserting this relation into eq. (2.1) I derive a more common form of the linear elastodynamic wave equation as

$$(c_{ijkl}\ u_{k,l}(\mathbf{x},t))_{,j} + f_i(\mathbf{x},t) = \rho\ddot{u}_i(\mathbf{x},t). \tag{2.5}$$

The analytical or semi-analytical solutions of the elastodynamic wave equation can be found only for some canonical problems (e.g. wave propagation in isotropic homogeneous or laterally-homogeneous media (Aki and Richards, 1980)), and, hence the solution of this equation in realistic anisotropic heterogeneous media relies on some approximations. The main difficulty of solving this elastodynamic wave equation is that the wavefield cannot, in general, be described with a few identified P or S-wave arrivals in realistic heterogeneous media because of the wavefield complexity (Červený, 2001). However, one can compute the solution of the elastic wave propagation, eq. (2.5), with numerical methods. Here, I only discuss the resolution of this equation in the frame of the high frequency (HF) approximation of the wave propagation, but before that I review some assumption that one can consider on the Earth properties.

From a pragmatic point of view, the elastic assumption for the medium is not mandatory in many seismic imaging applications in exploration geophysics, where the subsurface can be assumed to be acoustic (Alkhalifah, 2000, 2003). In acoustic media, there is no S-wave excitation from the source and the acousto-dynamic equation can be derived from the elastodynamic counterpart, simply by setting the S-wave velocity to zero. This simplifies and speeds up the resolution of the wave equation by providing a simpler structure for the stiffness tensor and removing the effect of the S-wave propagation.

On the other hand, one can approximate the Earth interior by an isotropic medium or an anisotropic one with particular symmetry. This leads to a considerable simplification in the stiffness tensor structure and makes the resolution of the elastic wave equation less challenging. The stiffness tensor for a general heterogeneous anisotropic medium includes 21 coefficients (triclinic symmetry) while this number reduces to 2 and 5 for elastic isotropic and transversely isotropic (TI) media (section 1.3), respectively. Therefore, these assumptions on the Earth anisotropy alongside considering the acoustic wave propagations can greatly facilitate the problem of investigating the seismic wave propagation in the Earth.

Despite the fact that anisotropic acoustic wave equation does not express any physical phenomena (Alkhalifah, 2000), it produces accurate traveltimes of quasi-P waves in weak to moderate anisotropic media, even at long offsets (Tsvankin and Thomsen, 1994; Tsvankin, 1995). However, the full numerical solution of the acoustic wave equation in TTI media generates artefacts resulting from spurious S waves generated out of the symmetry axis at the source or by conversion at lithological interfaces (Grechka et al., 2004). These spurious S waves lead to an unstable mode in VTI media where the Thomsen's parameters δ and ϵ are such that $\delta > \epsilon$. The footprint of these spurious S waves can be limited by considering smooth media to prevent conversion and considering locally elliptic media around the sources (Duveneck and Bakker, 2011). Alternatively, more sophisticated boundary conditions have been proposed to filter out them (Métivier et al., 2014c). The TTI acoustic wave equation also accurately predicts amplitudes of qP waves in weak to moderate anisotropic media (Operto et al., 2009). This has contributed to popularize the application of waveform inversion based imaging techniques such as full waveform inversion under the VTI acoustic approximation (e.g., Warner et al., 2013).

A popular approximation that is used to solve the seismic wave equation relies on the high frequency assumption, which assumes that the propagated wavelength is small compared to the wave travelling distance and the dimensions of the subsurface heterogeneities (Virieux and Lambaré, 2015). This allows to describe the wave as an impulse, itself defined by a traveltime

and amplitude, and calculate them by applying some conditions. Ray theory is the cornerstone of high-frequency body-wave seismology (Chapman, 2004). The ray methods provide some invaluable insights on the physics of wave propagation by separating the wavefield into individual elementary waves and tracking the paths followed by these waves (Červený, 2001). In the following I discuss in more details the high-frequency solution of the wave equation.

2.2.1 High frequency solution of wave equation

As a first consequence of the high frequency assumption for the elastic wave propagation, P and S-wave propagations can be considered separately even inside heterogeneous media. Moreover, this assumption simplifies the waves propagation problem such that it includes only three ingredients: the propagation path described by the ray trajectories perpendicular to the wavefronts, the traveltime and amplitude along the rays. These closely related concepts can be calculated through numerous methods and selection of an appropriate procedure to derive traveltimes and rays depends on many factors such as (Červený, 2001, chapter 3): dimension of the model under consideration (1D, 2D, 3D), discretization and complexity of the model (e.g. a smooth model without interfaces, blocky models with sharp interfaces, a regular gridded model), the required accuracy and efficiency for the traveltimes and rays calculation, the type of the traveltime (e.g. first-arrivals only, diffracted waves), the desired output information (e.g. only traveltimes, ray trajectories, Green function, synthetic seismograms) and etc. Here I follow the discussion for a smooth medium without sharp interfaces. The main goal here is deriving the equations whose unknown are traveltimes.

In the absence of external force term $f(\mathbf{x},t)$ in eq. (2.5), a solution (the ansatz) for the displacement vector $u(\mathbf{x},t)$ in an anisotropic smoothly inhomogeneous medium is

$$u_i(\mathbf{x},t) = U_i(\mathbf{x})F(t-T(\mathbf{x})), \tag{2.6}$$

where F is the oscillating term for an analytical high frequency wave. I consider this term as a simple time harmonic $F = \exp(-i\omega(t-T(\mathbf{x}))$ with frequency ω. $T(\mathbf{x})$ is the traveltime of the wave in the medium and $\overrightarrow{U}(\mathbf{x})$ stands for the amplitude of this wave. $T(\mathbf{x}) = const.$ is the definition of the wavefront. Note that both traveltime $T(\mathbf{x})$ and amplitude $\overrightarrow{U}(\mathbf{x})$ are frequency independent. Inserting this solution into eq. (2.5) results in

$$N_i(\overrightarrow{U})\ddot{F} - M_i(\overrightarrow{U})\dot{F} + L_i(\overrightarrow{U})F = 0 \tag{2.7}$$

where, by omitting the coordinate dependency,

$$\begin{aligned} N_i(\overrightarrow{U}) &= c_{ijkl}T_{,l}T_{,j}U_k - \rho U_i, \\ M_i(\overrightarrow{U}) &= c_{ijkl}T_{,j}U_{k,l} + (c_{ijkl}T_{,l}U_k)_{,j} \\ L_i(\overrightarrow{U}) &= (c_{ijkl}U_{k,l})_{,j}. \end{aligned} \tag{2.8}$$

For a high frequency wave propagating in a smooth medium, the dominant terms in eq. (2.7) (due to the power of ω) are the first two terms. I will now make the approximation that the third

term can be neglected in eq. (2.7). The resulting approximate solution of eq. (2.7) is obtained by solving equations $N_i(\overrightarrow{U}) = 0$ (which includes traveltime information) and $M_i(\overrightarrow{U}) = 0$ (which includes amplitude information). These equations result in the eikonal and transport equations, respectively [1]. Here I only discuss the eikonal equation.

I can rewrite equation $N_i(\overrightarrow{U}) = 0$ (Christoffel equation or dispersion relation) as

$$(\Gamma_{ik} - \delta_{ik})U_k = 0, \tag{2.9}$$

where δ_{ik} denotes the Kronecker delta function and Γ_{ik} is the Christoffel matrix

$$\Gamma_{ik} = \frac{c_{ijkl}}{\rho} T_{,j} T_{,l} = \frac{c_{ijkl}}{\rho} p_j p_l. \tag{2.10}$$

Here, $p_j = T_{,j}$ and $p_l = T_{,l}$ are the slowness vector components such that $\nabla T(\mathbf{x}) = \mathbf{p}(\mathbf{x})$. Note that all the elements in the Christoffel matrix are function of the spatial coordinates. The Christoffel matrix, Γ_{ik}, in general, has three eigenvalues $G_1(\mathbf{x}, \mathbf{p})$, $G_2(\mathbf{x}, \mathbf{p})$, $G_3(\mathbf{x}, \mathbf{p})$ and three corresponding eigenvectors $\mathbf{g}^{(1)}(\mathbf{x}, \mathbf{p})$, $\mathbf{g}^{(2)}(\mathbf{x}, \mathbf{p})$, $\mathbf{g}^{(3)}(\mathbf{x}, \mathbf{p})$. Each of these eigenvalues and related eigenvectors describes one of the qS1, qS2 or qP wave propagation modes [2]. The eigenvalues include the information about wave traveltimes and the eigenvectors give the wave polarizations. Considering the Christoffel equation (2.9), the following first order non linear partial differential equation (PDE) is the eikonal equation

$$G_m(\mathbf{x}, \mathbf{p}) = 1 \quad m = 1, 2, 3, \tag{2.11}$$

where, by considering $\mathbf{g} = (g_1, g_2, g_3)$,

$$G_m(\mathbf{x}, \mathbf{p}) = \Gamma_{ik} g_i^{(m)} g_k^{(m)} = \frac{c_{ijkl}}{\rho} p_j p_l g_i^{(m)} g_k^{(m)}. \tag{2.12}$$

This relation depends both on $\mathbf{x}$ and $\mathbf{p}$ and, in this phase space, it is expressed under the form of an Hamiltonian

$$H(\mathbf{x}, \mathbf{p}) = (G_m(\mathbf{x}, \mathbf{p}) - 1) = 0 \quad m = 1, 2, 3. \tag{2.13}$$

The solutions of this dispersion relation are the first-arrival (fastest arrival) traveltimes for S1, S2 and P-waves. As an example to clarify these relations, for an isotropic medium the stiffness tensor is defined with Lamé elastic moduli $\lambda(\mathbf{x})$ and $\mu(\mathbf{x})$ such that

$$c_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}). \tag{2.14}$$

1. There are different approaches which try to take into account the third term in the resolution of eq.(2.7) (Červený, 2001): This term can be considered as a source term for the wave equation (2.7) which implies some sort of generalized Born approximation. Also, it is possible to combine this term with the second term which results in the frequency-dependent transport equation, or one can combine the third term with the first term which yields frequency-dependent eikonal or frequency-dependent ray equations.
2. In the following I omit the prefix quasi (q).

Considering this structure for the stiffness tensor, the Christoffel matrix is

$$\Gamma_{ik} = \frac{\lambda + \mu}{\rho} p_i p_k + \frac{\mu}{\rho} p_k p_k, \tag{2.15}$$

which results in the following eikonal equations for S1, S2 and P-waves, respectively,

$$G_1 = G_2 = \beta^2 p_k p_k = 1, \qquad G_3 = \alpha^2 p_k p_k = 1, \tag{2.16}$$

where

$$\alpha^2 = (\lambda + \mu)/\rho, \qquad \beta^2 = \mu/\rho. \tag{2.17}$$

Here α and β are the P and S-waves velocities, respectively. The eigenvector $\mathbf{g}^{(1)}$ associated with G_1 is a unit vector perpendicular to the wavefronts and the eigenvector vectors related to G_1 and G_2 are mutually perpendicular unit vectors which are locates in the plane perpendicular to $\mathbf{g}^{(1)}$.

Conventional ray tracing

Considering the general Hamiltonian (2.13), the characteristics are the trajectories along which $H(\mathbf{x}, \mathbf{p}) = 0$ and traveltimes are computed by numerical integration along these trajectories. This makes the characteristic system of the Hamiltonian (2.13) to be

$$\frac{dx_i}{d\chi} = \frac{dH}{dp_i}, \qquad \frac{dp_i}{d\chi} = -\frac{dH}{dx_i}, \qquad \frac{dT}{d\chi} = p_k \frac{dH}{dp_k}, \tag{2.18}$$

where χ is a real-valued parameter along the ray. Rays are defined as the characteristics of the eikonal equation and this system of ordinary differential equations defines the ray tracing system. This system generally results in the coordinates of the points and slowness along the ray trajectory ($x_i(\tau)$ and $p_i(\tau)$, respectively) as well as traveltime, phase and group velocity and the angle between wavefront normal and the polarization vector.

Remember, the ray tracing system (2.18) is defined based on the assumption that the medium is smoothness (no sharp interface) and in order to perform ray tracing in a blocky model, it is needed to supplement the ray definition with the Snell's law at the intersection of rays and interfaces. In this case, one needs to distinguish between the incident rays and those generated due to interaction of the incident rays and the interface (such as reflected rays).

There are mainly two approaches for ray tracing: 1) initial-value ray tracing and 2) boundary-value ray tracing. In initial-value ray tracing the direction of the ray is known at some points of the ray and these can be used as the initial conditions to solve the rays system. In the boundary-value ray tracing the direction of the ray is unknown. Instead, there are some conditions which rays should respect. For instance, the ray should connect two specific points. There are numerous methods in these two categories which in the follow I only mention a few of them.

The initial-value ray tracing approach includes different methods such as: a) Numerical solution of the first-order ordinary differential equations of ray system with some initial conditions

(Červený et al., 1988). b) Analytical solution of the ray tracing system for some canonical models such as an isotropic model with constant gradient slowness. c) Semi-analytical approaches such as cell ray tracing in which the whole model is subdivided to several cells and within each cell the rays are solved analytically. In the simplest formulation, the velocity within the individual cells can be constant and the interfaces coincide with the cell borders (Aki et al., 1977). A more general formulation can be achieved by imposing the continuity of velocity at the cell borders (Virieux et al., 1988).

A well-known application of the boundary-value ray tracing is in two-points ray tracing in which one searches for the ray(s) connecting two points in a medium. The two main methods for performing two-point ray tracing are the shooting and ray bending method. The shooting method is an iterative procedure in which through the initial-value ray tracing one solves a boundary-value ray tracing (see Bulant, 1996, for a 3D algorithm). Fig. 2.2(a) shows application of shooting method in a smooth medium in which the rays are calculated by initial-valued ray tracing with take-off angles from the source as the initial conditions. Here, one searches for the proper ray passing the two target points by changing the take-off angles. b) The next method for performing two-points ray tracing is the bending method. In this method, one first guesses an initial hypothetical ray path connecting the two points and then perturbs iteratively this ray to find the one which respects the boundary conditions (passing two points) and the physics of ray propagation (Fermat's principle (in smooth media) and Snell's law (in layered media)) (Fig. 2.2b). Note that, the initial ray is not required to be an actual ray. In bending approach the search for the desired ray can be performed with different methods which among them are bending method based on the paraxial ray approximation (Farra, 1992), minimization of traveltime (Um and Thurber, 1987) and structural perturbation (Keller and Perozzi, 1983; Farra and Madariaga, 1987; Farra, 1989; Virieux, 1989, 1991; Virieux and Farra, 1991). Reviewing the details of these method is out of the scope of this study. I refer the readers to Červený (2001, chapter 3) for an extensive references and review on these methods.

Direct calculation of traveltimes and wavefronts

In some applications, the traveltimes or the wavefronts in the entire model are required. In this case, the conventional ray tracing may not be efficient and it is difficult to sample all the model with a bundle of rays from source (Vinje et al., 1993). Two widely used approach to attain these goals are direct resolution of eikonal equation and wavefront reconstruction (Fig. 2.3). Eikonal solvers directly solve the eq. (2.13) and result in the first-arrival traveltimes maps, or equivalently the first-arrival wavefronts, (single-valued maps) in whole the discretized model (refer to section 2.1 for some key references). In the wavefront reconstruction techniques, the governing idea is to use ray tracing to reconstruct a new wavefront from the older wavefront. Here the motivation is to keep under control the sampling of the wavefront. This technique has been applied successfully for 3D smooth media to build the multi-valued traveltime [3] and amplitude maps (Vinje et al., 1992, 1993, 1996; Lambaré et al., 1996; Lucio et al., 1996). Here, one can have direct access to the all the ray variables (e.g. slowness vectors and ray trajectories)

3. If several rays connect two points of the medium with different traveltimes, the traveltime is multi-valued.

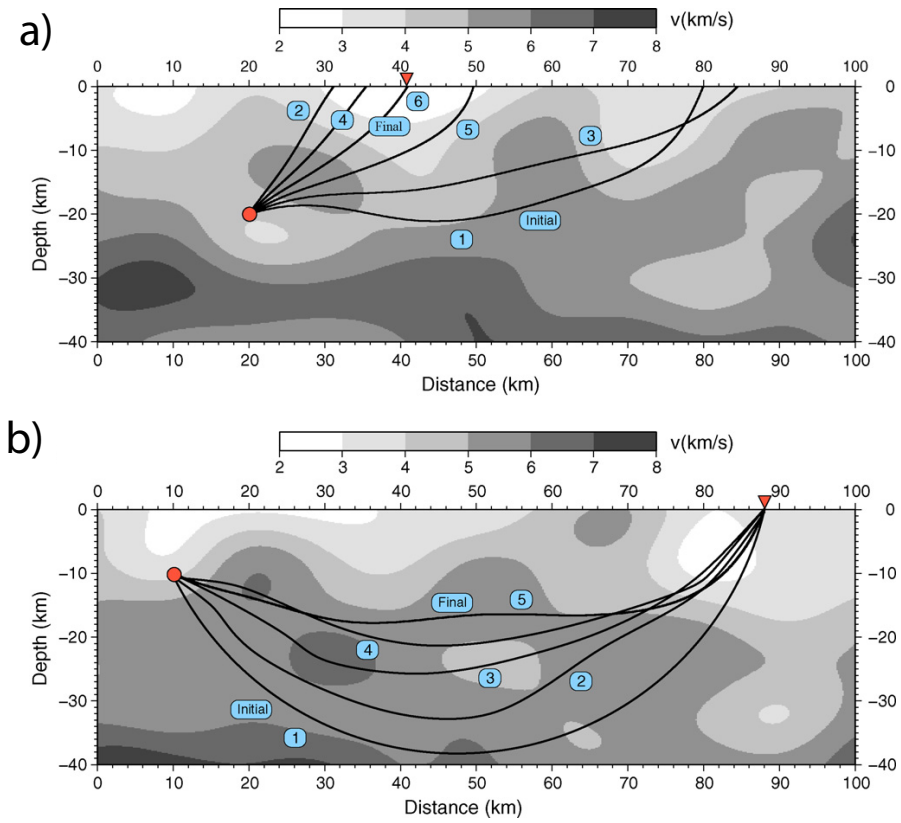

Figure 2.2 – Tow-points ray tracing between a source (circle) and receiver (triangle). a) Shooting method: one iteratively updates the initial path from source by changing the initial take-off angle of the ray such that the final ray connects the source and receiver. b) Bending method: an initial path is perturbed iteratively to converge to a physically meaningful ray (here, the one which satisfies the Fermat's principle) for the medium (figure from Rawlinson et al. (2010)).

(Lambaré et al., 1996; Lucio et al., 1996) while, in general, eikonal solvers do not provide a direct access to the ray related variables.

The ray traveltimes (the traveltimes which are calculated along the rays) and the first-arrival traveltime can be equal. For example, in a homogeneous acoustic model in presence of a point source, the direct wave traveltimes are the first-arrival traveltimes. For a non-smooth acoustic medium these two quantities describe totally different informations. For instance, beyond the caustics, where the ray traveltime is multivalued, the first-arrival traveltimes are coincide with the fastest branches of the P-waves not the most energetic.

Concerning the sampling issue of the medium, eikonal solvers can provide the first-arrival traveltime maps for a gridded model through a rather simpler formulation (compared to the ray tracing based approaches) while there is no shadow zone for the calculated map. In the context of smooth models, the shadow zones are the parts of the model where the rays do not exist or they have a very small geometric amplitude , i.e. with a small change in the ray take-off angle

there is a large change in the ray trajectory (Vidale, 1988a). This can be observed mostly in presence of low velocity zones and application of the conventional ray tracing.

Which approach?

Selecting a suitable approach depends on the application. For instance, in tomographic applications, ray tracing results in accurate traveltimes but may lack efficiency if a uniform sampling of the medium is requested. On the other hand, in some applications like migration, where we back propagate the energy inside the medium, using the first-arrival traveltime may distorts or change the final migrated section amplitude (Červený, 2001; Gray and May, 1994; Thierry et al., 1999).

In the stereotomography approach, where the final goal is the reconstruction of a smooth velocity model for subsurface, the one-way ray traveltimes from a source/receiver to a scatterer (Fig. 1.14) can be equivalent to the first-arrival traveltime. This suggests that in stereotomography the ray tracing forward modelling engine can be replaced with an eikonal solver. Our primary motivations for moving toward a slope tomographic method based on the eikonal solvers are: avoiding shadow zones in the calculated traveltimes with rays, easier implementation of eikonal solvers compared to the ray tracing (specially in anisotropic media) and efficient calculation of the first-arrival traveltimes on the subsurface grid (no need for interpolation). However, calculation of rays (which require resolution of ordinary partial differential equations) is cheaper than solving eikonal equation (which involves solving a partial differential equation). I try to compensate this extra computational cost by modifying the framework of stereotomography which is based on the ascending rays (rays travel from the scatterers toward the sources and receivers). In the new framework, I solve the eikonal equations for a descending propagation regime (propagation from sources and receivers to the scatterer). This allows to scale the total computational cost to the number of sources and receivers rather than scatterers (details in chapter 3).

Since I aim at developing a VTI/TTI slope tomographic method based on the eikonal solvers, in the following I first define the acoustic VTI/TTI eikonal equation and then I shall discuss how I solve them.

2.3 Eikonal equation for acoustic VTI media

The acoustic TTI/VTI assumption is relevant in many situations to describe subsurface media at the exploration geophysics scale. The 2D VTI/TTI eikonal equation will be of interest for the slope tomographic approach I want to promote. Here, I derive the acoustic VTI eikonal equation from Hamiltonian (2.13) (Alkhalifah, 2000, 2003), that requires to build the stiffness tensor and solve the dispersion relation (2.9).

The stiffness tensor c_{ijkl} is mostly represented by Voigt notation, $C_{\alpha\beta}$, as a square 6×6 matrix. According to the Voigt notations outlined in table 2.1, the stiffness tensor for a 3D VTI elastic medium is

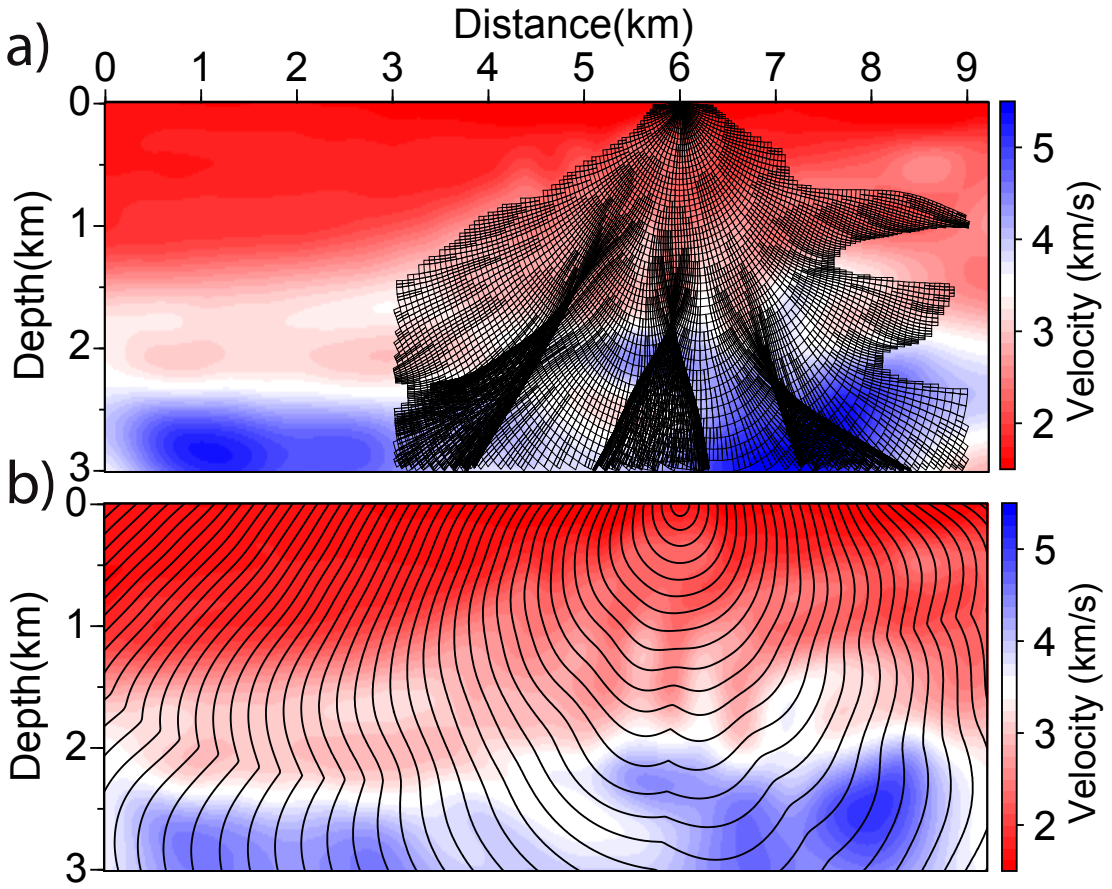

Figure 2.3 – Direct calculation of traveltimes and wavefronts for a smooth velocity model for a source on the surface. Solution of a) the wavefront reconstruction method (Lambaré et al., 1996) and b) an eikonal solver. The Eikonal solver only provides the first arrivals while the wavefront reconstruction solution includes a multi-valued traveltime map. Note that both methods sample well the medium.

Tensor index	ij	kl	11	22	33	32=23	31=13	12=21
↓								
Voigt notation	α	β	1	2	3	4	5	6

Table 2.1 – Voigt notation for stiffness tensor indexes.

$$C_{\alpha\beta} = \begin{pmatrix} C_{11} & C_{11}-2C_{66} & C_{13} & 0 & 0 & 0 \\ C_{11}-2C_{66} & C_{11} & C_{13} & 0 & 0 & 0 \\ C_{13} & C_{13} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & & C_{66} \end{pmatrix}. \tag{2.19}$$

Therefore, following eq. (2.10) I can write the Christoffel matrix for 3D VTI elastic medium elements as

$$\begin{aligned}
\Gamma_{11} =&(C_{11}p_1^2 + C_{66}p_2^2 + C_{55}p_3^2)/\rho,\\
\Gamma_{22} =&(C_{66}p_1^2 + C_{11}p_2^2 + C_{55}p_3^2)/\rho,\\
\Gamma_{33} =&(C_{55}(p_1^2 + p_2^2) + C_{33}p_3^2)/\rho,\\
\Gamma_{12} =&(C_{11} + C_{66})p_1p_2/\rho,\\
\Gamma_{13} =&(C_{13} + C_{55})p_1p_3/\rho,\\
\Gamma_{23} =&(C_{13} + C_{55})p_2p_3/\rho.
\end{aligned} \tag{2.20}$$

For a 2D acoustic medium [4] $C_{44} = C_{66} = 0$ (since these coefficients describe S-wave propagations) and $p_2 = 0$, and consequently the associated Christoffel matrix reduces to

$$\Gamma_{ik} = \begin{pmatrix} \frac{C_{11}}{\rho}p_1^2 & 0 & \frac{C_{13}}{\rho}p_1p_3 \\ 0 & 0 & 0 \\ \frac{C_{13}}{\rho}p_1p_3 & 0 & \frac{C_{33}}{\rho}p_3^2 \end{pmatrix}. \tag{2.21}$$

Finally, according to eqs. (2.9) and (2.11), zeroing the determinant of $(\Gamma_{ik} - \delta_{ik})$ gives the the eikonal equation in 2D acoustic VTI media as follow

$$\frac{C_{11}}{\rho}p_1^2 + \frac{C_{33}}{\rho}p_3^2 - p_1^2p_3^2\frac{C_{33}C_{11} - C_{13}^2}{\rho^2} = 1, \tag{2.22}$$

where C_{11} and C_{33} govern the anisotropy properties in the vertical and horizontal propagation, respectively.

In order to simplify the interpretation of the eikonal equation (2.22) and, in general, wave propagation, one can parametrize this equation based on different combinations of the stiffness tensor coefficients (Thomsen, 1986; Tsvankin, 1997; Alkhalifah, 2000). Thomsen (1986) parameterized the VTI media with one elastic moduli v_v (vertical P-wave velocity) and two dimensionless anisotropic parameters, ϵ and δ. For a 2D acoustic medium, these parameters are given by

$$v_v = \sqrt{C_{33}/\rho}, \tag{2.23}$$

$$\epsilon = \frac{C_{11} - C_{33}}{2C_{33}}, \tag{2.24}$$

and

$$\delta = \frac{C_{13}^2 - C_{33}^2}{2C_{33}^2}. \tag{2.25}$$

The effect of ϵ is mostly significant along horizontal propagation paths, while its effect is negligible along the vertical direction, and δ, which does not include C_{11}, control the near-vertical anisotropy.

4. In the following, unless otherwise stated, the formulations and discussions are based on acoustic assumption for the medium.

By parametrizing eq. (2.22) with Thomsen parameters and considering $p_1 = \partial_z T$ and $p_3 = \partial_x T$, I rewrite the eikonal equation (2.22) as

$$v_v^2(1+2\epsilon)(\partial_x T)^2 + v_v^2(\partial_z T)^2 - 2v_v^4(\epsilon-\delta)(\partial_x T \partial_z T)^2 = 1. \tag{2.26}$$

Alkhalifah and Tsvankin (1995) suggested an alternative parametrization based on two elastic moduli, v_v and v_{nmo}, and a dimensionless parameter, η. This parametrization reads

$$v_{nmo} = v_v\sqrt{1+2\delta}, \tag{2.27}$$

and

$$\eta = 0.5(\frac{v_h^2}{v_{nmo}^2} - 1) = \frac{\epsilon-\delta}{1+2\delta}, \tag{2.28}$$

where v_h is the horizontal velocity

$$v_h = v_0\sqrt{1+2\epsilon}. \tag{2.29}$$

They concluded that this parametrization is sufficient for all the time-related processing, like NMO correction. Considering these parametrizations, I can write a general form of VTI acoustic eikonal equation as

$$A(\partial_x T)^2 + C(\partial_z T)^2 + E(\partial_x T)^2(\partial_z T)^2 = 1, \tag{2.30}$$

where

$$\begin{aligned} A =& v_v^2(1+2\epsilon) = v_h^2, \\ C =& v_v^2 = \frac{v_{nmo}}{1+2\delta}, \\ E =& -2v_v^4(\epsilon-\delta) = -2v_v^2 v_{nmo}^2\eta. \end{aligned} \tag{2.31}$$

[5] In the next section, I introduce the 2D TTI eikonal equation and the algorithm to solve it.

2.4 Solution of TTI Eikonal equation

A heterogeneous 2D TTI medium can be considered locally as a VTI medium with a tilted symmetry axis (section 1.3). If these local tilt angles be $\theta(\mathbf{x})$ (positive for clockwise rotation and negative for anticlockwise), I can write the 2D heterogeneous TTI eikonal equation as (Alkhalifah, 2003; Waheed et al., 2014)

$$A(\partial_{\hat{x}} T)^2 + C(\partial_{\hat{z}} T)^2 + E(\partial_{\hat{x}} T)^2(\partial_{\hat{z}} T)^2 = 1, \tag{2.33}$$

5. The 2D isotropic eikonal equation can simply be derived by setting $\epsilon = 0$ and $\delta = 0$,

$$\|\nabla T\|^2 = \frac{1}{v^2}, \tag{2.32}$$

where v is the velocity in medium.

where coefficient A, C and E are defined locally with eq. (2.31), and the symbol "$\widehat{\cdot}$" represents the local rotated coordinate system such that

$$\begin{aligned} \partial_{\widehat{x}} T &= \partial_x T \cos\theta - \partial_z T \sin\theta, \\ \partial_{\widehat{z}} T &= \partial_x T \sin\theta + \partial_z T \cos\theta. \end{aligned} \tag{2.34}$$

Here I omit the coordinate dependency of θ for the sake of brevity. I set the traveltime T to zero at the source position, which is a Dirichlet boundary condition for PDE (2.33). I propose an algorithm based on the fast sweeping method (FSM) (Zhao, 2005) in which I approximate the gradient of the traveltimes with a finite difference operator. Using such a method raises two difficulties: 1) The eikonal equation for TTI/VTI media embeds some terms which involve products of first-order spatial derivatives raised to a power greater than two. This leads to numerical difficulties to derive accurate solutions when the first-order derivatives are discretized. 2) The zero traveltime at the source results in singularity at the source position and consequently erroneous calculated traveltimes. I shall proceed through an iterative procedure (fixed-point iterations) and factorization method to overcome these two issues.

2.4.1 Implementation of fixed-point iteration technique

In order to cope with the products of traveltime derivatives raised to power higher than two in the 2D TTI eikonal equation, I use the fixed point iteration technique which inserts iteratively the information of high power terms into the main solver.

By inserting eq. (2.34) into eq. (2.33), I can rewrite TTI eikonal equation (2.33) as

$$\tilde{a}(\partial_x T)^2 - 2\tilde{c}(\partial_x T)(\partial_z T) + \tilde{b}(\partial_z T)^2 = D(T), \tag{2.35}$$

where the left and right-hand-side terms are elliptic and anelliptic components of the equation, respectively. Here

$$\begin{aligned} \tilde{a} &= A\cos^2\theta + C\sin^2\theta, \\ \tilde{b} &= A\sin^2\theta + C\cos^2\theta, \\ \tilde{c} &= (A - C)\cos\theta\sin\theta, \end{aligned} \tag{2.36}$$

and the right-hand-side term $D(T)$ reads

$$D(T) = 1 - E\left(\partial_x T\cos\theta - \partial_z T\sin\theta\right)^2\left(\partial_x T\sin\theta + \partial_z T\cos\theta\right)^2. \tag{2.37}$$

$D(T)$ includes all the products with a power higher than two of the traveltime derivatives. For elliptic anisotropy (EA) eikonal equation $\epsilon = \delta$ and this results in $E = 0$ and consequently $D(T) = 1$. Inspired by the similarity between the left-hand-side term (elliptic term) in eq. (2.35) and elliptically-anisotropic eikonal equation [6], Waheed et al. (2014) proposed an iterative approach based on fixed-point iteration method (Kelley, 1995) to take into account the anelliptic term in the resolution of the TTI eikonal equation.

6. Note that the left-hand-side of eq. (2.35) does not represent an elliptically-anisotropic eikonal equation because ϵ and δ can have different values. But, by abuse of language, I refer to this term as elliptic eikonal equation.

According to the fixed-point iteration method, in order to solve a non-linear equation like $f(x) = 0$, one can consider it as $f(x) + x = x \equiv g(x) = x$. Therefore, solution of the equation $g(x) = x$ which is a fixed-point of g can be considered as the solutions of $f(x) = 0$. Through the recursive process $x_n = g(x_{n-1})$ and an initial value for the fixed-point, this algorithm converges to a fixed-point for g which is solution of $f(x) = 0$.

This suggests that the eikonal equation (2.35) can be formulated as

$$\tilde{a}(\partial_x T_n)^2 - 2\tilde{c}(\partial_x T_n)(\partial_z T_n) + \tilde{b}(\partial_z T_n)^2 = D(T_{n-1}), \tag{2.38}$$

where $n \geq 1$ indicates the fixed-point iteration number. Considering this iterative method, I introduce the following quantities

$$a_n \to \frac{\tilde{a}}{D(T_n)}, b_n \to \frac{\tilde{b}}{D(T_n)}, c_n \to \frac{\tilde{c}}{D(T_n)} \tag{2.39}$$

and rewrite eq. (2.35) for iteration n of the fixed-point technique as

$$\sqrt{a_{n-1}(\partial_x T_n)^2 + b_{n-1}(\partial_z T_n)^2 - 2c_{n-1}(\partial_x T_n)(\partial_z T_n)} = 1. \tag{2.40}$$

Solving this eikonal equation for iteration n yields the first-arrival traveltime map T_n and this enables us to update the coefficients a_n, b_n, c_n through eq. (2.39). By iterating this procedure one can update the traveltime maps to satisfy the convergence criterion. I considered $D(T_0) = 1$ as the initial guess to build the initial coefficients a_0, b_0, c_0. Fig. 2.4 shows the flowchart of this approach based on the fixed-point iterations.

Waheed et al. (2014) took advantage of the FSM to solve the elliptic eikonal equation (2.40) in each iteration of the fixed-point algorithm where the traveltime gradients are approximated by finite differences. This algorithm handles efficiently the presence of the anelliptic term in the TTI eikonal equation, but it suffers from source singularity error propagation. In the following, I discuss an elliptic eikonal solver (Luo and Qian, 2012) in which the source singularity is removed by a factorization method.

2.4.2 Factorized eikonal solver

According to the factorization method proposed by Fomel et al. (2009); Luo and Qian (2012), solution of eq. (2.40) can be written as

$$T = T_0 \tau, \tag{2.41}$$

where T_0 is a known analytical solution of an elliptic eikonal equation in a homogeneous medium, whose properties match those of the real medium at the source position. The τ map should be computed at each point of the medium and will correct the traveltimes to account for the differences between the real and the homogeneous media. In the factorization method, T_0 captures the source singularity and τ is a smooth function around the source. Therefore, the

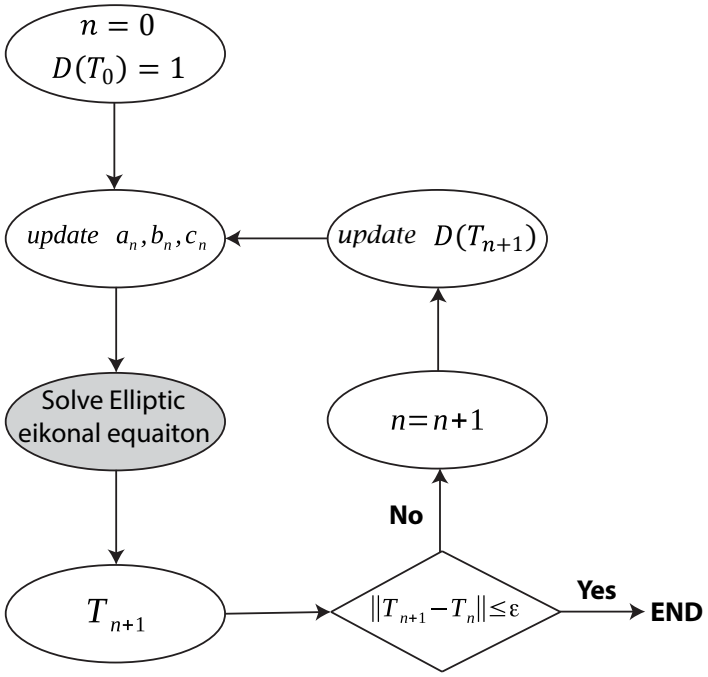

Figure 2.4 – Fixed-point iteration flowchart. Initialize the anelliptic term with $D(T_0) = 1$. Then update the coefficients of the elliptic eikonal equation and solve it. Iterate this procedure until the convergence criterion is satisfied. Here ε designates a small number (10E-4).

spatial gradient of traveltime can be expressed as

$$\begin{aligned} \partial_x T \equiv T_x &= T_{0_x}\tau + \tau_x T_0, \\ \partial_z T \equiv T_z &= T_{0_z}\tau + \tau_z T_0, \end{aligned} \tag{2.42}$$

with obvious notations for derivatives. By inserting eq. (2.42) into eq. (2.40) and omitting the subscripts n and $n-1$, I obtain the Hamiltonian for the factorized elliptic eikonal equation (Luo and Qian, 2012) as

$$H(x, z, \tau_x, \tau_z) = \sqrt{\begin{array}{l} \tau^2(aT_{0_x}^2 - 2cT_{0_x}T_{0_z} + bT_{0_z}^2) + 2T_0\tau(aT_{0_x}\tau_x \\ -c(T_{o_x}\tau_z + T_{0_z}\tau_x) + bT_{0_z}\tau_z) + T_0^2(a\tau_x^2 - 2c\tau_x\tau_z + b\tau_z^2) \end{array}} - 1 = 0. \tag{2.43}$$

Here T_0 can be analytically calculated from the homogeneous elliptic eikonal equation

$$\begin{cases} \sqrt{a(\mathbf{x}_0)T_{0_x}^2 - 2c(\mathbf{x}_0)T_{0_x}T_{0_z} + b(\mathbf{x}_0)T_{0_z}^2} = 1, \quad \mathbf{x} \in \Omega \setminus \{\mathbf{x}_0\}, \\ T_0(\mathbf{x}_0) = 0, \end{cases} \tag{2.44}$$

where $\mathbf{x}_0 = (x_0, z_0)$ denotes the source position in the computational domain Ω and T_0 is defined as

$$T_0(x,z) = \sqrt{\frac{b(\mathbf{x}_0)(x-x_0)^2 + 2c(\mathbf{x}_0)(x-x_0)(z-z_0) + a(\mathbf{x}_0)(z-z_0)^2}{a(\mathbf{x}_0)b(\mathbf{x}_0) - c(\mathbf{x}_0)^2}}. \tag{2.45}$$

This definition of T_0 satisfies the required smoothness for τ around the source, in other words,

$$\lim_{(x,z)\longrightarrow(x_0,z_0)} \frac{T(x,z)}{T_0(x_0,z_0)} = 1. \tag{2.46}$$

In order to numerically calculate τ in eq. (2.43), Luo and Qian (2012) proposed a first-order FSM on a rectangular mesh. For a grid step size h they discretized eq. (2.43) thorough four triangles (Fig.2.5). This discretization provides an upwind scheme respecting the causality condition in the fast sweeping method. For example, a discretization involving the triangle

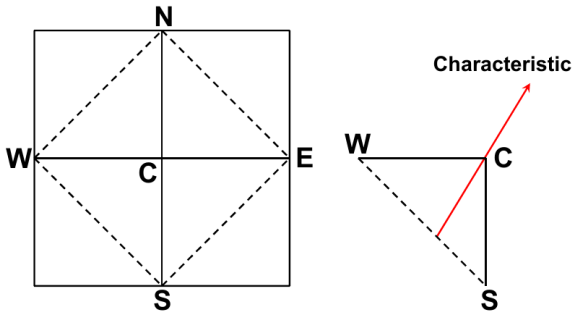

Figure 2.5 – Discretization scheme. Rectangular mesh (left) and the finite difference stencil (right). Each grid point is the central vertex for four different stencils (from Luo and Qian, 2012).

$\triangle WCN$ at vertex C can be performed with a linear Taylor expansion of τ_x and τ_z through the expression

$$\nabla\tau(C) = (\tau_x(C), \tau_z(C)) \approx \left(\frac{\tau(C)-\tau(W)}{h}, \frac{\tau(C)-\tau(S)}{h}\right). \tag{2.47}$$

Inserting equation (2.47) into the factorized elliptic eikonal equation (2.43) leads to the following discretized elliptic eikonal equation

$$\sqrt{k_1\tau^2(C) + k_2\tau(C) + k_3} = 1, \tag{2.48}$$

where $k_i|_{i=1}^3$ are defined as

$$
\begin{aligned}
k_1 &= a{T_{0_x}}^2 - 2cT_{0_x}T_{0_z} + b{T_{0_z}}^2 + 2\frac{T_0}{h}(T_{0_x}(a-c) + T_{0_z}(b-c)) + \frac{T_0^2}{h^2}(a-2c+b),\\
k_2 &= 2\frac{T_0}{h}(\tau(W)(cT_{0_z} - aT_{0_x}) + \tau(S)(cT_{0_x} - bT_{0_z})) - 2\frac{T_0^2}{h^2}(a\tau(W) - c(\tau(W)+\tau(S)) + b\tau(S)),\\
k_3 &= \frac{T_0^2}{h^2}(a\tau^2(W) - 2c\tau(W)\tau(S) + b\tau^2(S)).
\end{aligned}
$$

Therefore, for a given $\tau(W)$ and $\tau(S)$ in triangle $\triangle WCN$, resolution of the quadratic equation (2.48) provides the value of $\tau^h(C)$ [7]. The other triangles of the discretization scheme (Fig.2.5) lead to similar quadratic equation. The value of $\tau^h(C)$ inferred from the resolution of the quadratic equation can be an acceptable candidate if it is real and satisfies the causality condition. I recall that the causality condition can be stated as follow (Luo and Qian, 2012): the characteristic of the function T passing through vertex C of each triangle should be between the edges of the triangle (Fig. 2.5). For instance, for triangle $\triangle WCN$ the characteristic vector passing C is defined as

$$
(a(C)T_x^h(C) - c(C)T_z^h(C),\ b(C)T_z^h(C) - c(C)T_x^h(C)), \tag{2.49}
$$

where $\nabla T^h(C) = (T_x^h, T_z^h) = \nabla\tau^h(C)T_0(C) + \tau^h(C)\nabla T_0(C)$. In order to satisfy the causality condition, vector (2.49) should be between edges $\overrightarrow{WC}$ and $\overrightarrow{SC}$. When equation (2.48) yields no real solution for $\tau^h(C)$, one can convey the information about τ from the neighbours to the central vertex C along the edges of the associated triangle, and compute a viscosity solution for $\tau^h(C)$ through the method of characteristics. The characteristic system of Hamiltonian $H(x, y, \tau_x, \tau_z)$ (eq. (2.43)) reads

$$
(\frac{dx}{dt}, \frac{dz}{dt}) = (\frac{\partial H}{\partial \tau_x}, \frac{\partial H}{\partial \tau_z}). \tag{2.50}
$$

These equations can be used to present a viscosity solution for eq. (2.48). For example, in triangle $\triangle WCN$ by forcing the characteristics to be along edge $\overrightarrow{WC}$ I can write

$$
\frac{dz}{dt} = \frac{\partial H}{\partial \tau_z} = 0, \tag{2.51}
$$

and in result $b(C)T_z^h(C) = c(C)T_x^h(C)$. Inserting this into the factored eikonal equation (2.43) yields $\sqrt{\frac{ab-c^2}{b}}(T_0\tau_x + T_{0_x}\tau) = 1$. Finally, by discretizing through triangle $\triangle WCN$ I derive the following viscosity solution for $\tau^h(C)$ (Luo and Qian, 2012)

$$
\tau^h(C) = \frac{T_0(C)\tau(W) + h\sqrt{\frac{b(C)}{a(C)b(C)-c^2(C)}}}{T_0(C) + T_{0_x}(C)(x_C - x_W)}. \tag{2.52}
$$

7. Here and in the following discussions, the superscript h only emphasizes that the values are discretized on a regular mesh with grid interval h.

Each of the four triangle stencils can provide at most two candidates for $\tau^h(C)$ which satisfy the causality condition, but only the minimum one will be selected as the final candidate for the value of τ at the central vertex. The algorithm (1) associated with the solver of the factorized elliptic eikonal, equation (2.43), is outlined below.

Algorithm 1: Factorized elliptic eikonal solver for the FD stencils in Fig. 2.5 (at each grid point)

for $\triangle WCN$ **do**
  Discretize eq. (2.43).
  Solve the quadratic eq. (2.48) $\Rightarrow \tau_{c1}, \tau_{c2}$.
  if $\tau_{c1} \in I\!R^+$ **and** $\tau_{c2} \in I\!R^+$ **then**
    if τ_{c1}, τ_{c2} satisfy casualty **then**
      $T_{WN}(C) = \min\{T_0(C)\tau_{c1}, T_0(C)\tau_{c2}\}$
    else if τ_{c1} satisfies casualty **then**
      $T_{WN}(C) = T_0(C)\tau_{c1}$
    else if τ_{c2} satisfies casualty **then**
      $T_{WN}(C) = T_0(C)\tau_{c2}$
    else if τ_{c1}, τ_{c2} do not satisfy casualty **then**
      Calculate viscosity solution along to edges $\Rightarrow \tau_{WC}, \tau_{NC}$.
      if $T_0(C)\tau_{WC} > T(W)$ **and** $T_0(C)\tau_{NC} > T(N)$ **then**
        $T_{WN}(C) = \min\{T_0(C)\tau_{WC}, T_0(C)\tau_{NC}\}$
      else if $T_0(C)\tau_{WC} > T(W)$ **then**
        $T_{WN}(C) = T_0(C)\tau_{WC}$
      else if $T_0(C)\tau_{NC} > T(N)$ **then**
        $T_{WN}(C) = T_0(C)\tau_{NC}$
      else
        $T_{WN}(C) = \infty$
      end if
    end if
  else
    Calculate viscosity solution along to edges $\Rightarrow \tau_{WC}, \tau_{NC}$.
    if $T_0(C)\tau_{WC} > T(W)$ **and** $T_0(C)\tau_{NC} > T(N)$ **then**
      $T_{WN}(C) = \min\{T_0(C)\tau_{WC}, T_0(C)\tau_{NC}\}$
    else if $T_0(C)\tau_{WC} > T(W)$ **then**
      $T_{WN}(C) = T_0(C)\tau_{WC}$
    else if $T_0(C)\tau_{NC} > T(N)$ **then**
      $T_{WN}(C) = T_0(C)\tau_{NC}$
    else
      $T_{WN}(C) = \infty$
    end if
  end if
end for
for $\triangle ECN$, $\triangle ECS$, $\triangle WCS$ **do**
  Do as above $\Rightarrow T_{EN}(C), T_{ES}(C), T_{WS}(C)$
end for
$T(C) = \min\{T_{WN}(C), T_{EN}(C), T_{ES}(C), T_{WS}(C)\}$

FSM implementation

In order to cover all the possible propagation directions (the characteristics) I rely on the sweeping scheme of FSM. The FSM method includes a Gauss-Seidel iteration with alternating sweeping ordering which covers all the possible directions for the characteristics:

$$\begin{aligned} &\textbf{sweep1} : i = 1 : n_1,\ j = 1 : n_2; \qquad &&\textbf{sweep2} \quad : i = 1 : n_1,\ j = n_2 : 1; \\ &\textbf{sweep3} : i = n_1 : 1,\ j = 1 : n_2; \qquad &&\textbf{sweep4} \quad : i = n_1 : 1,\ j = n_2 : 1, \end{aligned} \tag{2.53}$$

where i and j are the indexes of each grid point on the regular $n_1 \times n_2$ mesh which discretizes the subsurface model. I sweep repeatedly the computational domain with these four sweeps and in each sweep I apply the factorized elliptic eikonal solver explained in the Algorithm (1). This results in four candidates at each grid point and I always pick the minimum value as the final traveltime value. Algorithm 2 represents this procedure in detail.

Algorithm 2: FSM algorithm for elliptic factored eikonal solver. Here $\mathbf{x}_0$ designates the source position. Note that in this algorithm $T_0(\mathbf{x})$ is the traveltime map for the whole computational domain in iteration 0, not the factored term in eq. (2.41).

Initialization
$n = 0$
if $\mathbf{x} = \mathbf{x}_0$ **then**
  $T_0(\mathbf{x}_0) = 0$
else
  $T_0(\mathbf{x}) = \infty$
end if
while Diff$<$ 10e-4 **do**
  $n = n + 1$
  for $i = 1 : n_1\ j = 1 : n_2$ **do**
    $T_n(\mathbf{x}) \rightarrow$ Perform Algorithm 1 $\Rightarrow T_{sweep1}(\mathbf{x})$
  end for
  for $i = 1 : n_1\ j = n_2 : 1$ **do**
    $T_{sweep1}(\mathbf{x}) \rightarrow$ Perform Algorithm 1 $\Rightarrow T_{sweep2}(\mathbf{x})$
  end for
  for $i = n_1 : 1\ j = 1 : n_2$ **do**
    $T_{sweep2}(\mathbf{x}) \rightarrow$ Perform Algorithm 1 $\Rightarrow T_{sweep3}(\mathbf{x})$
  end for
  for $i = n_1 : 1\ j = n_2 : 1$ **do**
    $T_{sweep3}(\mathbf{x}) \rightarrow$ Perform Algorithm 1 $\Rightarrow T_{sweep4}(\mathbf{x})$
  end for
  $T_n(\mathbf{x}) = T_{sweep4}(\mathbf{x})$
  Diff = $\|T_n(\mathbf{x}) - T_{n-1}(\mathbf{x})\|_\infty$
end while

In summary, the proposed algorithm for the resolution of the 2D TTI eikonal equation involves the following steps. I decompose the TTI eikonal equation in to an elliptic eikonal equation and a right-hand side term that accounts for anellipticity. Through the fixed-point iteration technique, I update iteratively the solution of the elliptic eikonal equation with the anellipticity-related correction. In each iteration of the fixed-point method, I solve an elliptic eikonal solver based on the FSM and factorization method to calculate the traveltime maps. These traveltime maps are used to compute the corrections required by the anellipticity. This loop continues until a stopping criterion of iterations is satisfied.

I apply the FSM algorithm (Algorithm 2) on a simple homogeneous isotropic model to demonstrate the role of the factorization and FSM methods in the traveltime calculation. Here, since the model is isotropic, there is no fixed-point iteration (the anelliptic term is one). I consider a 5 km $\times 5$ km subsurface model with a grid interval of 50 m, a velocity of 1000 m/s in the background and a point source at the center. Fig. 2.6 shows the difference between the analytical solution and the calculated traveltimes with the eikonal solver using factorization (Fig. 2.6b) and without factorization (Fig. 2.6c). Here the factorization method efficiently removes the singularity at the source and results in 10^5 times smaller error for calculated traveltimes. Fig. 2.7 shows the calculated traveltimes for each sweep separately. Each sweep includes a specific range of characteristics or propagation directions.

2.4.3 Source implementation in practice

The zero traveltime at the source position is a Dirichlet boundary condition for the eikonal equation. This can be extended for situations in which the source is not at any grid point (Fig. 2.8). In this case, I analytically calculate the traveltime for the four surrounded grid points and consider their traveltimes as the Dirichlet boundary condition for the eikonal solver. This requires a careful calculation of the traveltimes from the source to the surrounded grid points. By considering a homogeneous TTI model inside the associated grid cell, I can use the analytical phase velocity definition in homogeneous TTI media (Tsvankin, 1997) to calculate the wave propagation velocity and consequently traveltimes in four different directions (arrows in Fig. 2.8). This definition reads

$$\frac{V^2(\beta,\theta)}{v_0^2} = \frac{1}{2} + \epsilon \sin^2(\beta-\theta) + \frac{1}{2}\sqrt{(1+2\epsilon\sin^2(\beta-\theta) - 2(\epsilon-\delta)\sin^2(2(\beta-\theta)))}, \quad (2.54)$$

where $V(\beta,\theta)$ is the phase velocity of the wave, θ and β are tilted angle in the grid cell and the take-off angle of rays (arrows in Fig. 2.8) toward the neighbour grid points, respectively. As mentioned by Fomel et al. (2009), compared with source positioning on the grid points, here the extra boundary conditions for the eikonal solver result in extra iterations for FSM to converge.

2.4.4 Numerical examples

In this section I assess the accuracy of the proposed eikonal solver with different numerical examples involving long-distance propagation. These tests include a homogeneous isotropic,

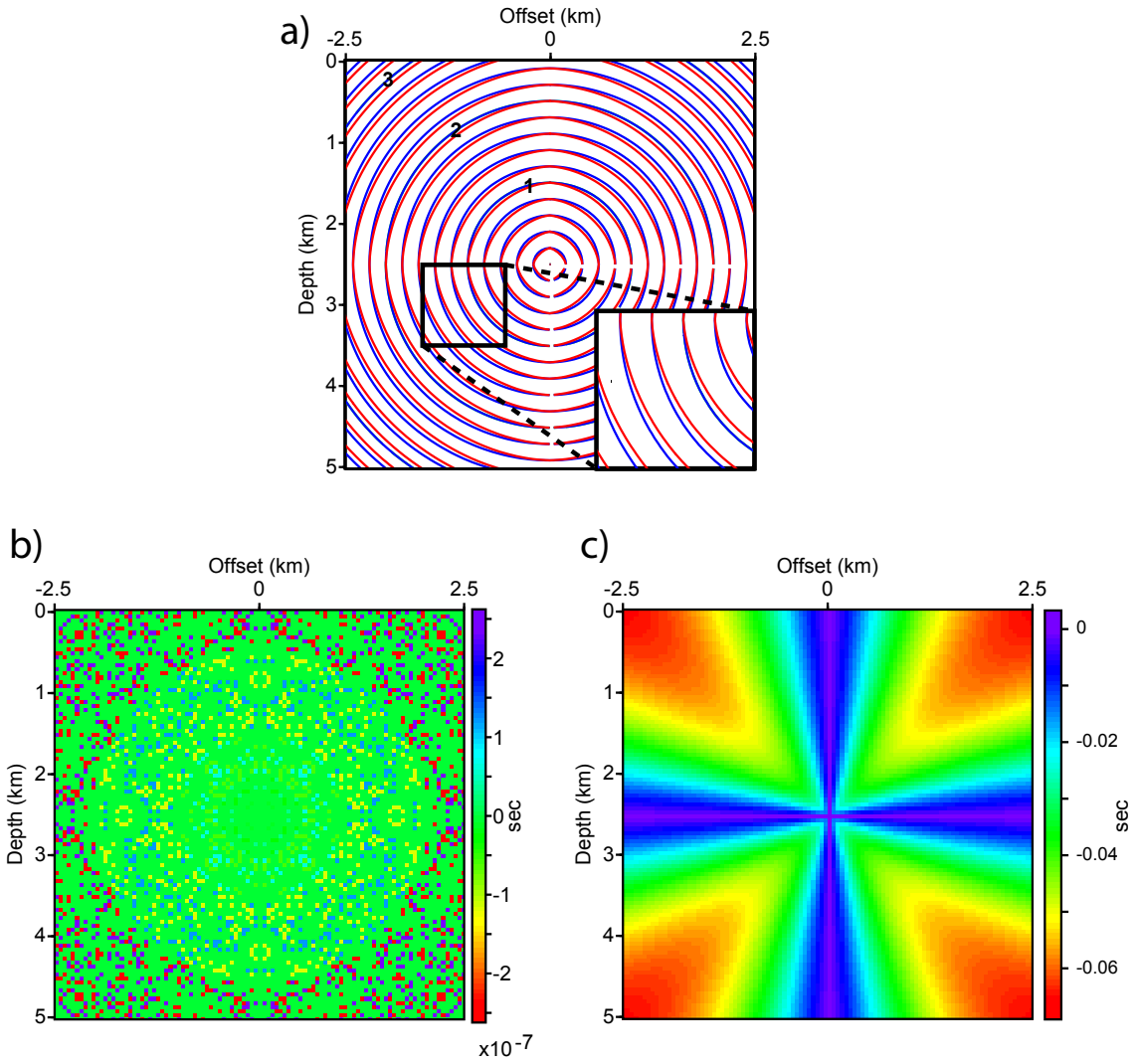

Figure 2.6 – Factorization effect on traveltime calculation for a homogeneous isotropic model. a) First-arrival traveltime contours for analytical solver (green) and eikonal solver with factorization (blue) and without factorization technique (red). The blue and green contours are completely matched. Difference between the analytical and eikonal solver solutions b) with factorization and c) without factorization technique. Factorization technique removes the source singularity error from the eikonal solver solution. The magnitude of the maximum error is 10^5 times smaller than the solution of the eikonal solver without factorization.

homogeneous VTI and three TTI models (i.e. homogeneous, constant gradient velocity and BP salt model (Shah, 2007)). Then, I discuss about the grid interval effect on the algorithm accuracy. The tests are performed in double precision Fortran 90 code on a PC with Intel Xeon CPU 3.60GHz. For convergence criterion in the FSM and fixed-point iteration I consider $\|T_n - T_{n-1}\|_\infty < \varepsilon$ where ε is a threshold and n is the iteration count of the FSM or fixed-point methods. Here, I set $\varepsilon = 10^{-3}$ and $\varepsilon = 10^{-4}$ for the fixed-point iterations and the FSM, respec-

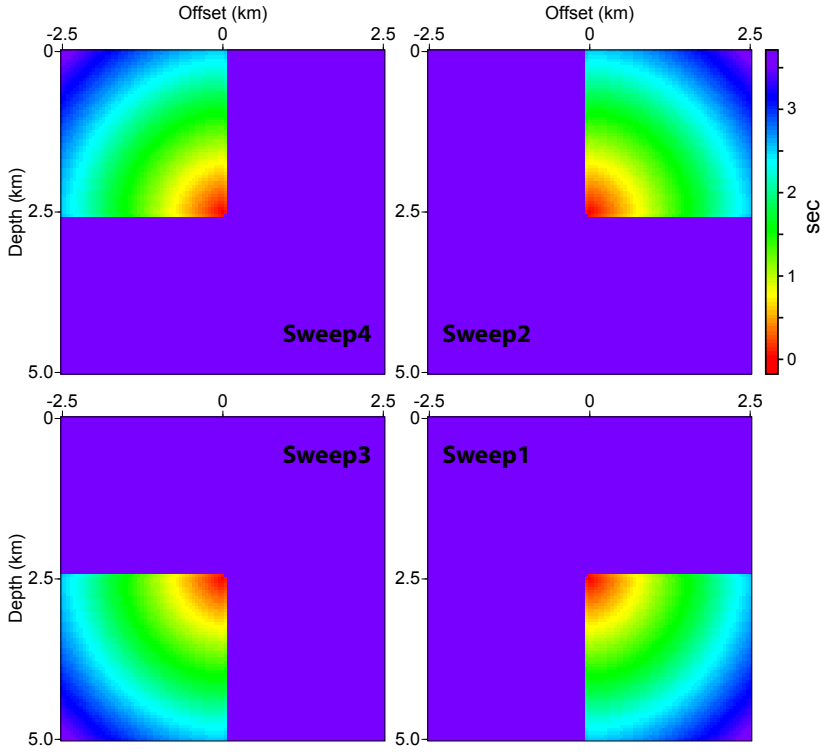

Figure 2.7 – FSM solution for each sweep for the example in Fig. 2.6. Each sweep calculates traveltimes within a specific propagation direction range.

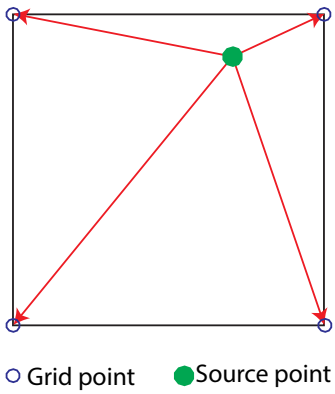

Figure 2.8 – Point source implementation when the source is not on the grid point. I calculate analytically the traveltime for the surrounded grid points and consider them as boundary conditions for the eikonal solver.

tively. I also use a threshold of 10^{-4} for the fixed-point iterations to assess the computational overhead generated by this loop. In order to understand this threshold is well chosen or not, I

calculate the following parameter (as a local convergence rate) in the last iteration of fixed-point

$$\alpha = \frac{log((s_n - s_{n-1})/(s_{n-1} - s_{n-2}))}{log((s_{n-1} - s_{n-2})/(s_{n-2} - s_{n-3}))}, \tag{2.55}$$

where s_i are the elements of given numerical series (here $\|T_n - T_{n-1}\|_\infty$) and α is called the convergence rate. By this definition the threshold is suitable if α be close to 1.

Homogeneous isotropic and VTI examples

Isotropic model
I consider a 20 km $\times 4$ km isotropic homogeneous model with a velocity of 1000 m/s and grid interval of $h = 50$ m. The point source is located at the upper-left corner of the model, $x_0 = z_0 = 0$ km. I validate the solution of the eikonal solver against the analytical solution. Fig. 2.9(a) shows the difference map between the two solutions where the eikonal solver accurately removes the point source singularity. Here the maximum error between numerical and analytical solution is less than 2×10^{-6} second. For the isotropic example there is no fixed-point iteration because the anelliptic term is zero, and therefore there is no value for convergence rate of fixed-point iterations.

VTI model
As a second example, I consider a 20 km $\times 4$ km homogeneous VTI model with a grid interval of $h = 50$ m. The subsurface parameters are $v_v = 3000$ m/s, $\epsilon = 0.15$ and $\delta = 0.05$, and the source is located at $x_0 = z_0 = 0$ km. I compare the analytical solution of Carcione et al. (1988) with our numerical solutions. Fig. 2.9(b) shows the superimposition of different wavefronts from the proposed eikonal solver and analytical solver where two solutions are almost identical. For this model, four iterations of the fixed-point method are sufficient to satisfy the convergence criterion, Fig. 2.10.

Table 2.2 presents the convergence rate, number of iterations performed by the fixed-point method and the maximum number of iterations in the FSM for the traveltime calculations in the homogeneous isotropic and VTI media. The computational cost for a threshold 10^{-4} is around 1.5 times more expensive than the one obtained with a threshold 10^{-3}, hence highlighting the linear relation between the computational cost and the number of fixed-point iterations.

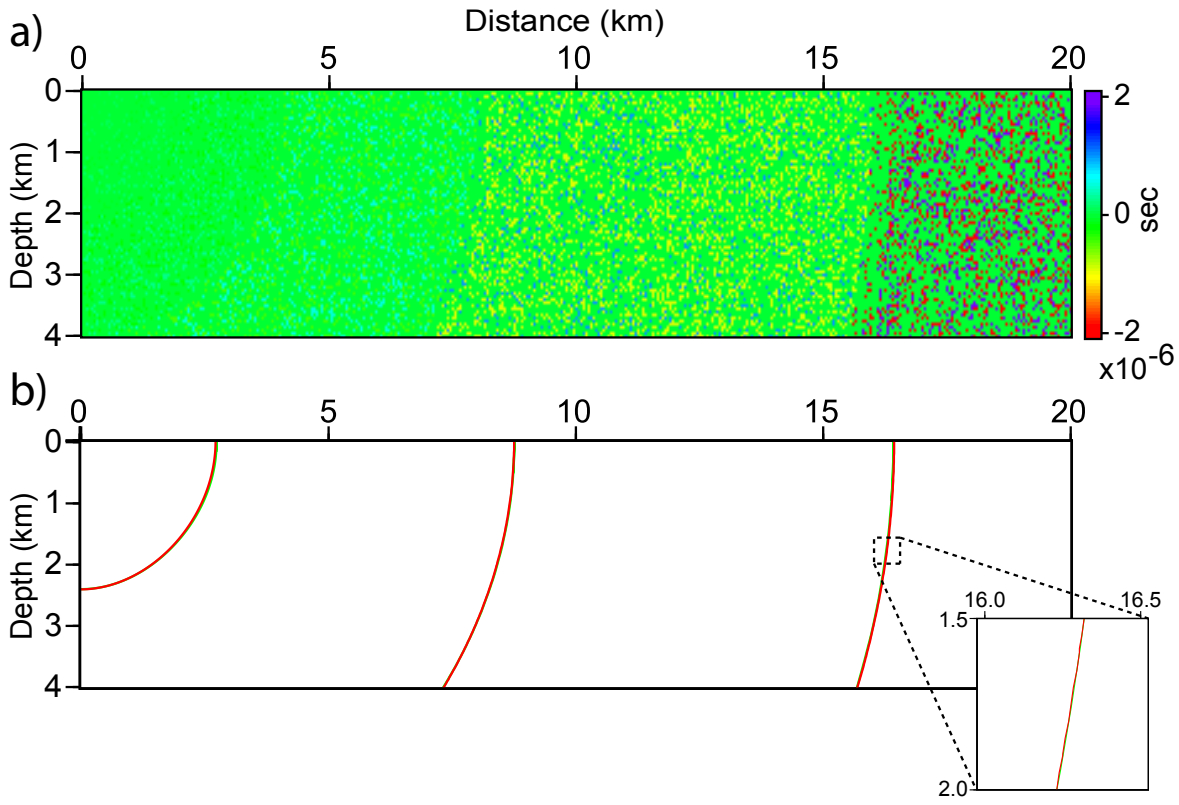

Figure 2.9 – Homogeneous isotropic and VTI examples. (a) Difference between analytical and numerical solution in second for a homogeneous isotropic model. The difference is increasing with distance while remaining below 2 microsecond. (b) Superimposition of traveltime contours at 0.8s, 2.56s and 4.8s, computed analytically (green) and with the eikonal solver (red) for a homogeneous VTI model. The box on the right corner of (b) shows a zoom on the traveltime contours.

Threshold	10^{-3}		10^{-4}		
MODEL ($n_1 \times n_2$)	FPI-(FSM) iteration #	Run time (sec)	FPI-(FSM) iteration #	Run time (sec)	α
ISO (81×401)	0-(2)	0.106	0-(2)	0.106	-
VTI (81×401)	4-(2)	1.058	6-(2)	1.575	1.053

Table 2.2 – Homogeneous isotropic and VTI examples. Computational cost and iteration numbers for homogeneous isotropic (ISO) and homogeneous VTI examples (VTI) in Fig. 2.9 for two convergence criteria for the fixed-point iterations. Columns with header FPI-(FSM) show the number of fixed-point iterations and the maximum number of FSM iterations. Computational costs are in second (Run time). The convergence rate α is calculated for threshold 10^{-4}.

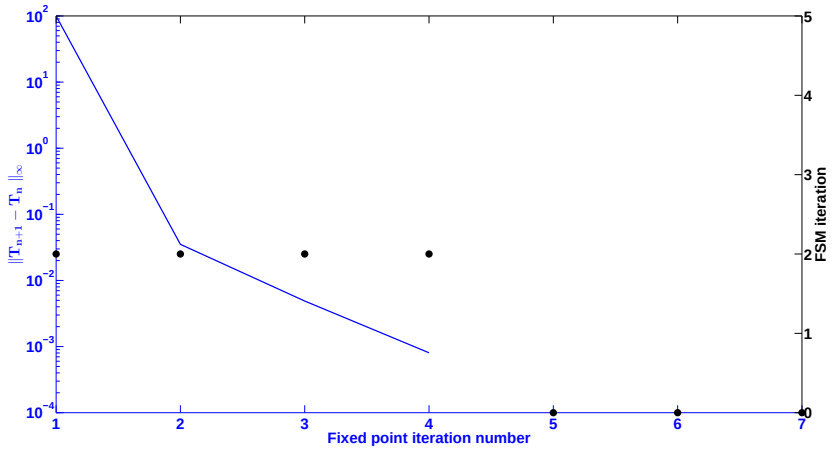

Figure 2.10 – Homogeneous VTI examples. Convergence diagram for the homogeneous VTI model of the example in Fig. 2.9(b). Black dots represent the number of FSM iterations in each fixed-point iteration. In each iteration of the fixed-point there are 2 FSM iterations.

TTI examples

Here I consider three examples including different TTI models (homogeneous, constant gradient and BP salt) of dimension 31.25 km $\times$ 6.25 km. These examples help to assess the effect of the structural complexity of the subsurface on the accuracy and computational cost of the proposed eikonal solver. I validate the the solutions of this eikonal solver against the full wavefield solutions computed with a $\mathcal{O}(\Delta t^2, \Delta x^4)$ TTI acoustic finite-difference time-domain (FDTD) code based on the staggered-grid stencil of Saenger et al. (2000) and the SMART absorbing boundary conditions of Métivier et al. (2014b). For all three tests, the source is located at $(z_0, x_0) = (125, 1000)$ m.

Homogeneous TTI model
For the homogeneous TTI (H-TTI) model I use $v_v = 3500$ m/s, $\theta = 30°, \epsilon = 0.15, \delta = 0.05$ and a grid interval of $h = 50$ m. Fig. 2.11 shows the superimposition of the eikonal solutions, red curves, on several snapshots of the FDTD wavefields. Parasite reflections from the boundaries are visible in the snapshots due to aggressive amplitude clipping. Convergence diagram, Fig. 2.12, shows that the algorithm satisfies a convergence criterion of 10^{-3} after five iterations of the fixed-point iteration algorithm.

Constant gradient TTI model
For this example I consider a constant gradient velocity TTI model (CG-TTI) where v_v increases with depth, from 2000 to 4000 m/s, and the tilted angle, ϵ and δ models are homogeneous ($\theta = 30°, \epsilon = 0.15, \delta = 0.05$). Fig. 2.13 presents the kinematic agreement between the eikonal solution and the FDTD snapshots at for four different traveltimes. The fixed-point iteration algorithm satisfies the stopping criterion of iteration for a threshold of 10^{-3} after six iterations (Fig. 2.14).

TTI BP salt model
In this example I calculate the first-arrival traveltimes in the complex TTI BP salt (BP-salt) model (background of panels in Fig. 2.15). Fig. 2.15 shows the superimposition of the wavefront computed with the eikonal solver with the FDTD wavefields at four different traveltimes. The FDTD and eikonal solutions closely coincide even in the complex part of the model and at long distance from the source. In order to have a more comprehensive assessment of the traveltime accuracy, I superimposed in Fig. 2.16 the eikonal traveltime curves on two sets of FDTD seismograms that are recorded by two receiver lines at 200 m and 4000 m depth. The good match between the two solutions can be also observed here. Here, four iterations of the fixed-point iteration method have been sufficient to satisfy a convergence criterion of 10^{-3}. However, the complexity of the geological features makes the first iteration of the fixed-point iteration method to perform more FSM iterations compared to the two previous TTI examples (Fig. 2.17).

Generally, these three TTI examples suggest that the proposed eikonal solver is accurate enough at near and far-offsets, regardless of the propagation direction, for imaging applications. Table 2.3 contains the convergence rates and the number of iterations for traveltime calculation of the examined TTI models for two convergence threshold of 10^{-3} and 10^{-4}. All the convergence rates are close to one which shows the stability of the fixed-point strategy. Decreasing the convergence threshold from 10^{-3} to 10^{-4} results in five and four extra fixed-point iterations in

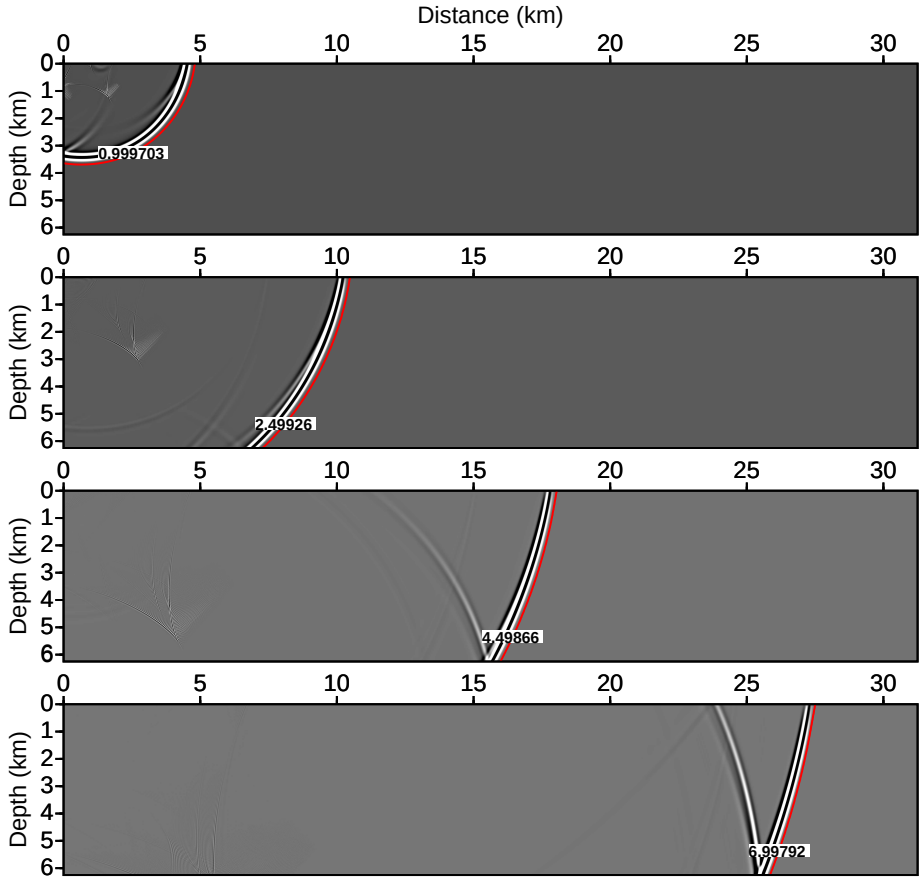

Figure 2.11 – Homogeneous TTI example. Superimposition of traveltime contours from eikonal solver, red curves, on the FDTD wavefields at times 1s, 2.5s, 4.5s and 7s. For all the offsets the eikonal solutions and wavefields coincide.

traveltime calculation for H-TTI and CG-TTI models, respectively, while the fixed-point iteration method performs only one extra iteration for the BP-salt model. In the following, I explain why the computational overhead resulting from the decrease of the convergence threshold is not in a same range for different models.

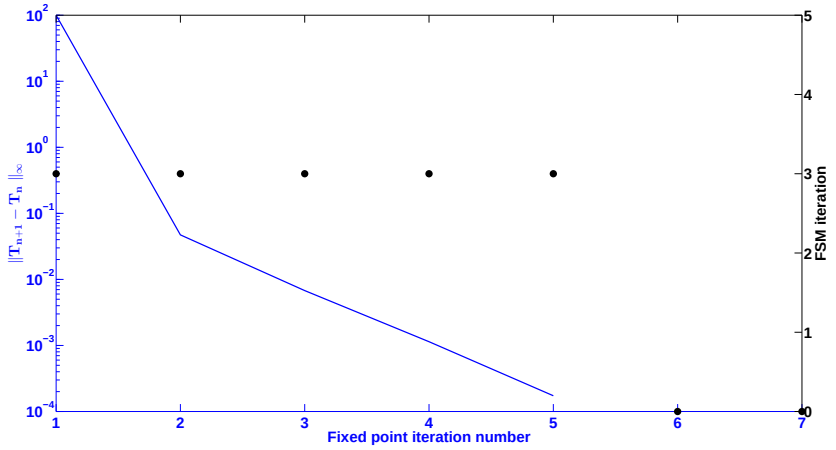

Figure 2.12 – Homogeneous TTI example. Convergence diagram for the homogeneous TTI model from the test in Fig. 2.11. Black dots represent the number of FSM iteration in fixed-point iterations. Algorithm requires 5 iterations to satisfy the convergence criterion 10^{-3} for fixed-point iteration. In each fixed-point iteration the algorithm performs 3 FSM iterations.

Threshold	10^{-3}		10^{-4}		
MODEL ($n_1 \times n_2$)	FPI-(FSM) iteration#	Run time (sec)	FPI-(FSM) iteration#	Run time (sec)	α
H-TTI (126×626)	5-(3)	1.2	10-(4)	3.8	0.8
CG-TTI (126×626)	6-(3)	1.7	10-(3)	2.4	1.21
BP-salt (126×626)	4-(4)	1.6	5-(4)	2.2	0.981

Table 2.3 – TTI examples. Computational cost and iteration numbers of homogeneous TTI (H-TTI), (CG-TTI) and BP salt TTI (BP-salt) examples in Figs 2.11, 2.13 and 2.15, respectively, for two convergence criteria for fixed-point loop. Columns with header FPI-(FSM) show the number of fixed-point iterations and the maximum number of FSM iterations. Computational costs are in second (Run time). The convergence rate α is calculated for convergence criterion 10^{-4}.

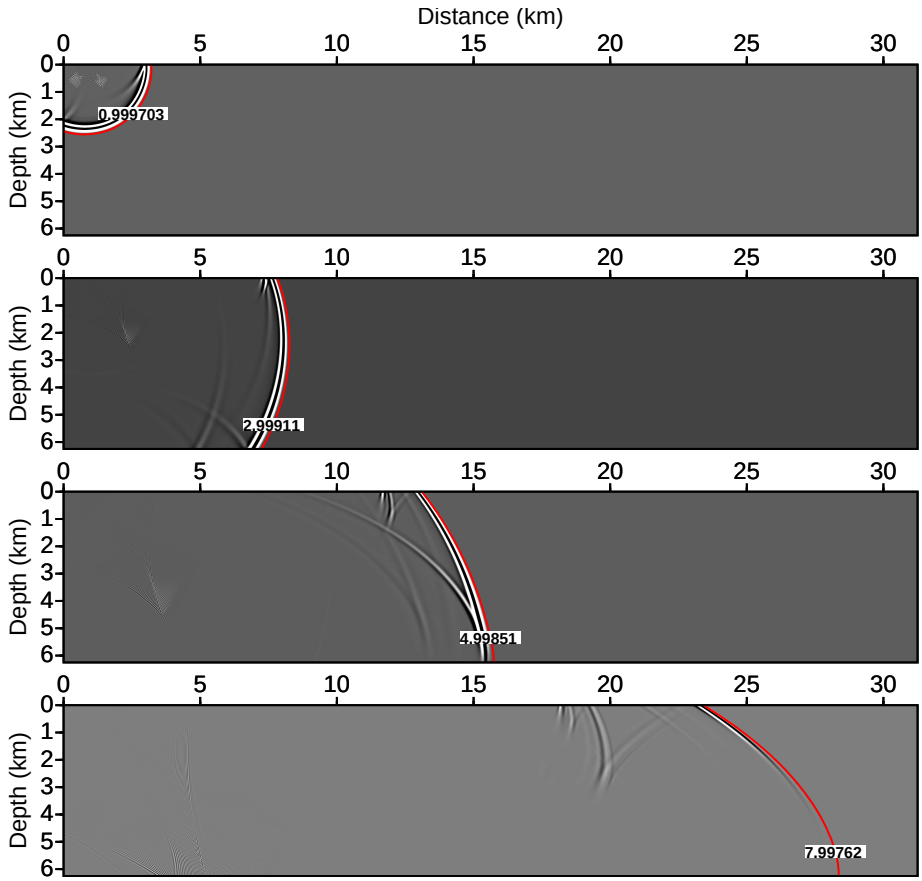

Figure 2.13 – Constant gradient TTI example. Superimposition of traveltime contours from eikonal solver, red curves, on the FDTD wavefields at times 1s, 3s, 5s and 8s. For all the offsets the contours from eikonal solver are matched with the wavefields.

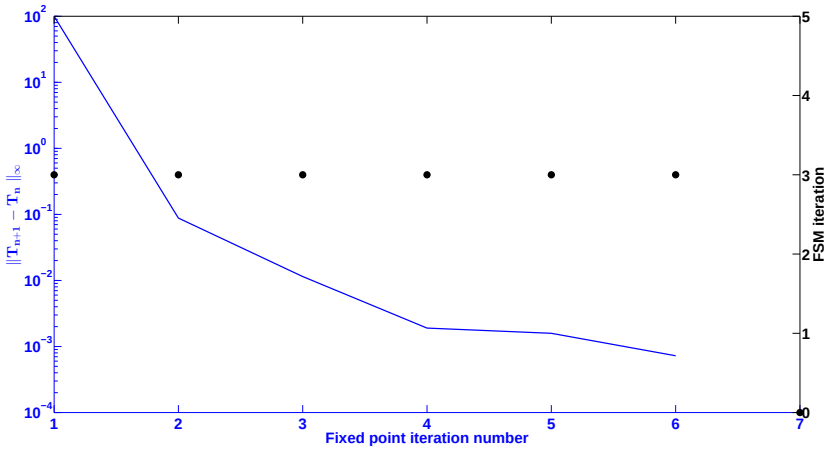

Figure 2.14 – Constant gradient TTI example. Convergence diagram for the constant gradient velocity TTI model from the test in Fig. 2.13. Black dots represent the number of FSM iteration in fixed-point iterations. With 6 iterations of the fixed-point the algorithm satisfies the convergence criterion 10^{-3}. In each fixed-point iteration the algorithm performs 3 FSM iterations.

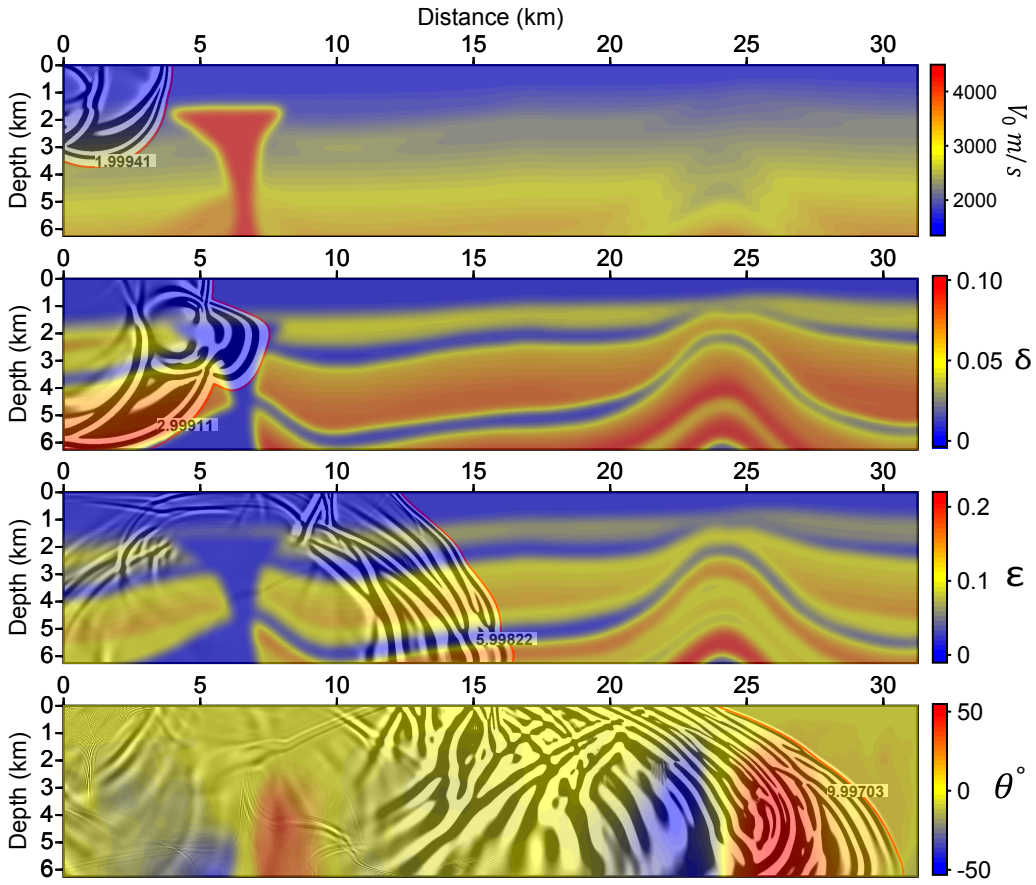

Figure 2.15 – TTI BP salt example. Superimposition of traveltime contours from eikonal solver, red curves, on the FDTD wavefields at times 2s, 3s, 6s and 10s for the TTI BP salt model. Even in complex parts of the model and large-offsets two solutions are almost identical.

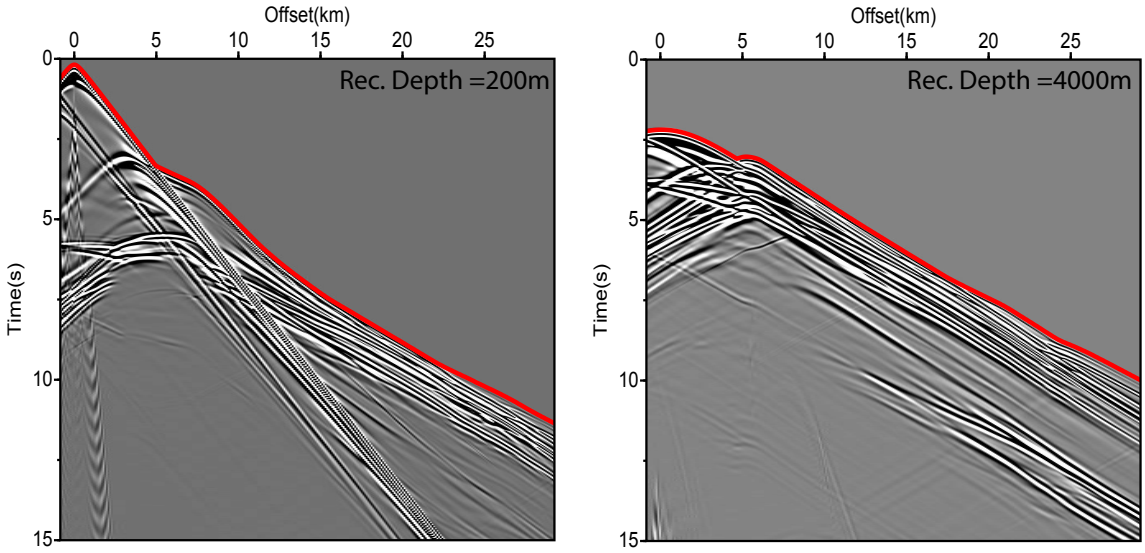

Figure 2.16 – TTI BP salt example. Traveltimes computed with the eikonal solver (red line) superimposed on the FDTD seismograms without saturation for BP salt model. The receiver lines are at depth 200m and depth 4000m. The first-arrival traveltimes calculated by eikonal coincide with the wavefields at different offsets.

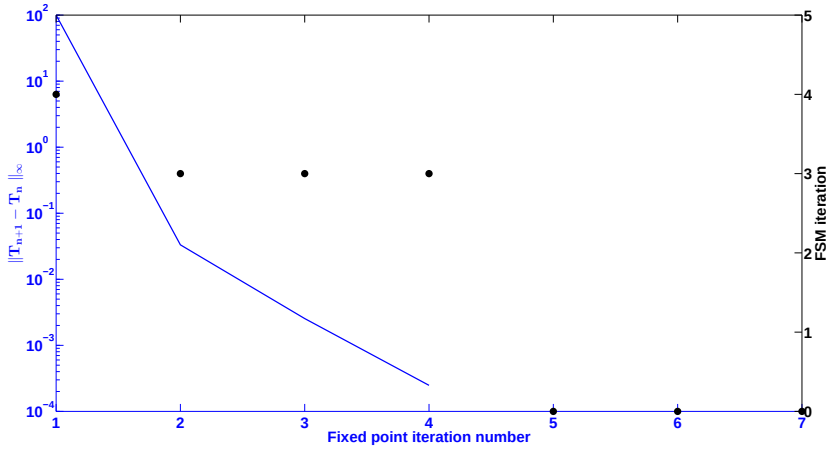

Figure 2.17 – TTI BP salt example. Convergence diagram for the TTI BP salt model for the test in Fig. 2.15. Black dots represent the number of FSM iteration in each of the fixed-point loop. The algorithm performs 4 iterations to satisfies the convergence threshold 10^{-3} of fixed-point. In the first fixed-point iteration the algorithm performs 4 FSM iterations and for the others the number of FSM iteration is 3. Here, compared to the previous TTI examples, the complex subsurface models result in more iteration for the FSM during the first iteration of the fixed-point.

Grid interval effect

In order to investigate the influence of the grid interval on the computational overhead resulting from the decrease in convergence threshold, I perform several simulations in some VTI and TTI models with different grid intervals. I consider the following models: homogeneous VTI (H-VTI), constant gradient velocity VTI (CG-VTI) and the three previous TTI models, i.e., homogeneous TTI (H-TTI), constant gradient velocity TTI (CG-TTI) and BP salt (BP-salt). The size of all these models is 31.25 km $\times$ 6.25 km and the subsurface parameters for H-VTI and CG-VTI are same as H-TTI and CG-TTI, respectively, only for VTI models the tilted angle is zero.

I discretize each of the mentioned models with four different grid intervals: 125, 50, 25 and 12.5 m while for all of them a point source is located at $(z_0, x_0) = (125, 1000)$ m. Then, I apply the eikonal solver to assess how the grid interval influences the convergence of the solver. Table 2.4 and 2.5 include the convergence rates and the number of fixed-point iterations and the maximum number of FSM iterations for different discretization of these VTI models (i.e. H-VTI and CG-VTI) and TTI models (i.e. H-TTI, CG-TTI and BP-salt), respectively. I used 10^{-4} as the threshold for the fixed-point iteration and 10^{-4} for the FSM.

MODEL	Mesh size	Grid step size (m)	FPI-(FSM) iteration#	α
H-VTI	51×251	125	6-(2)	1.093
H-VTI	126×626	50	6-(2)	1.040
H-VTI	251×1251	25	6-(2)	0.999
H-VTI	501×2501	12.5	6-(2)	0.973
CG-VTI	51×251	125	5-(2)	1.012
CG-VTI	126×626	50	5-(2)	1.013
CG-VTI	251×1251	25	5-(2)	0.998
CG-VTI	501×2501	12.5	5-(2)	0.98

Table 2.4 – Grid step size effect. Number of fixed-point iterations and the maximum number of performed FSM iterations (FPI-(FSM)), and convergence rate α for homogeneous VTI (H-VTI), constant gradient velocity VTI (CG-VTI) models with four different grid step sizes 125, 50, 25 and 12.5 m. For all the examples α is close to one and the number of fixed-point iterations is not changed for different discretization step lengths. For all the examples the threshold for fixed-point and FSM convergence is 10^{-4}.

MODEL	Mesh size	Grid step size (m)	FPI-(FSM) iteration#	α
H-TTI	51×251	125	7-(2)	0.70
H-TTI	126×626	50	10-(4)	0.80
H-TTI	251×1251	25	5-(3)*	0.80*
H-TTI	501×2501	12.5	-	-
CG-TTI	51×251	125	8-(3)	0.56
CG-TTI	126×626	50	10-(3)	1.21
CG-TTI	251×1251	25	19-(2)*	9.3*
CG-TTI	501×2501	12.5	-	-
BP-salt	51×251	125	5-(4)	0.818
BP-salt	126×626	50	5-(4)	0.981
BP-salt	251×1251	25	6-(4)	1.14
BP-salt	501×2501	12.5	8-(4)	1.036

Table 2.5 – Grid interval effect. Number of fixed-point iterations and the maximum number of performed FSM iterations (FPI-(FSM)), and convergence rate α for homogeneous TTI (H-TTI), constant gradient velocity TTI (CG-TTI) and BP salt models with four different grid step sizes 125, 50, 25 and 12.5 m. For all the examples the fixed-point and FSM convergence threshold is 10^{-4}, but for those that are indicated by "*" the final satisfied threshold for the fixed-point convergence is 10^{-3}. Note that for H-TTI and CG-TTI model when the grid step size is 12.5 m the algorithm can not satisfy even the convergence criterion 10^{-3} for fixed-point.

For the VTI models (H-VTI and CG-VTI), all the grid intervals lead to a convergence rate close to one, and the number of the fixed-point iterations is the same for all the models (Table 2.4). However, I show a different impact of the discretization when considering the TTI models instead of the VTI counterparts (Table 2.5). For instance, for H-TTI and CG-TTI models with a grid interval of $h = 12.5$ m, the algorithm never satisfied the convergence criterion of the fixed-point iteration, even for $\|T_{n+1} - T_n\|_\infty < 10^{-3}$, while different grid intervals for the BP salt model do not impact upon the convergence of the algorithm with a convergence rate is close to one for grid step length 12.5 m.

The reasons behind these observations can be attributed to two issues: 1) factorization performance for long-offset configuration for homogeneous models, and 2) the inescapable accumulation of numerical error at far offset due to round-off error. In the following I discuss briefly these issues.

1) In order to clarify the first issue, I refer to the factorization definition where $T = T_0\tau$ and the traveltime gradients are defined through equation system (2.42). After estimation of τ gradients with the finite difference stencil in Fig. 2.5 (eq. (2.47)), I can rewrite the

traveltime gradients in eq.(2.42) as

$$T_x = T_{0_x}(C)\tau(C) + \frac{\tau(C) - \tau(W)}{h} T_0(C),$$
$$T_z = T_{0_z}(C)\tau(C) + \frac{\tau(C) - \tau(S)}{h} T_0(C). \tag{2.56}$$

In these expressions when the given model is homogeneous or the τ field tends to be very smooth, I can state that $\tau(C) \approx \tau(W)$ or $\tau(C) \approx \tau(S)$, and in result the nominators of the fractions in eq. (2.56) are a small number while the denominators (h) are relatively a big number. This makes these division operations prone to be contaminated with round-off error. [8]

In order to verify this statement I can perturb the smoothness of τ and compare the convergence of the algorithm before and after this perturbation. This perturbation increases the magnitude of the nominators in eq. (2.56) and consequently should decrease the round-off error. Accordingly, I insert a 100×100 m square vertical velocity anomaly in the H-TTI model which is discretized with $h = 12.5$ m (Fig. 2.18). With this perturbation the algorithm satisfies the threshold $\|T_{n+1} - T_n\|_\infty < 10^{-3}$ in five fixed-point iterations, while the algorithm did not satisfy this convergence criterion for the original unperturbed model (Table 2.5).

2) The next issue can be related to the fixed-point strategy. The coefficients of the elliptic eikonal equation which are updated by the fixed-point iterations, eq. (2.39), are influenced by round-off error. This is significant when the tilted angle is not zero and the calculation of the anelliptic term, eq. (2.37), and the coefficients of the elliptic eikonal equation, eq. (2.36), involve arithmetic operations of $\cos\theta$ and $\sin\theta$. These errors have a significant role in the convergence and accuracy of our algorithm to such an extent that I was never able to satisfy the threshold criterion $\|T_{n+1} - T_n\|_\infty < 10^{-3}$ for the VTI and TTI models with the single precision version of the code.

Although in real applications we never face with such a large homogeneous models like the H-TTI and CG-TTI model, but it is important for us to be aware of these errors when we are defining initial models for forward solver in the slope tomography.

2.5 Conclusions

I have presented an iterative factorized eikonal solver for heterogeneous TTI media. The strategy is based on a fixed-point iterative solver: the simpler elliptic eikonal equation is solved by a fast sweeping method with updated right-hand-side terms for converging toward the TTI solution. Adding the factorization procedure removes the singularity at the source and improves dramatically the precision of the solution. Examples of complex geological TTI models

8. As a remedy to this problem one can propose to restrict the factorization scheme to the grid points around the source. But this affect the final results because of the footprint of different definitions for gradients around the source.

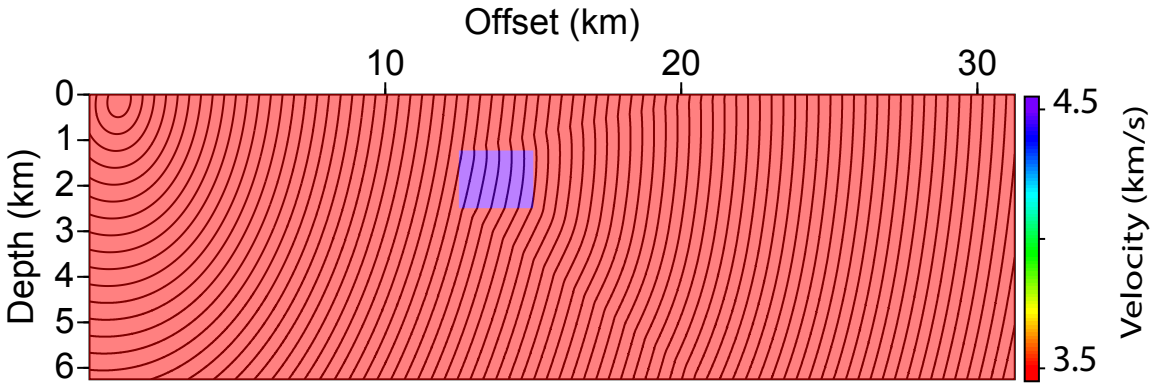

Figure 2.18 – Round-off error effect. First arrival traveltimes contours calculated with our eikonal solver, for homogeneous TTI example in Fig 2.11 in presence of a small velocity anomaly. The background colors designates the vertical velocity. Without this anomaly the eikonal solver can not satisfy the convergence criterion 10^{-3} for the fixed-point while in presence of this anomaly it converges. This velocity anomaly reduce the round-off error in the traveltime gradient calculations.

point out the ability of this method for a precise estimation of first-arrival times for near and far-offsets. I have identified the convergence difficulties according to different discretization scheme and probable sources of error. Nevertheless, the method is enough accurate to consider it as a forward engine for traveltime tomography and anisotropic slope tomography.

Chapter 3

Adjoint slope tomography: Inverse problem

Chapter overview: In this chapter I present the principles of the adjoint slope tomography in the framework of inverse problem theory and its application to various synthetic examples. First I review the basics of inverse problem theory with a focus on the main methods which are applied in this study such as the quasi-Newton L-BFGS optimization algorithm based on the gradient of the misfit function and the adjoint-state method for computing this gradient. In the second section, which is the duplicated version of our paper published in the *Geophysical Journal International* (Tavakoli F. et al., 2017b), I introduce the isotropic formulation of slope tomography that is followed, in the last section, by its extension to tilted transversely isotropic media. Each of these two sections includes a review of the formulation of classic ray-based stereotomography, our definition of the model and data spaces in the adjoint slope tomography, and finally an assessment of the method with synthetic examples of increasing complexity. First, some toy examples based upon simple subsurface models are used to assess the resolution power of the adjoint slope tomography as well as the relative role of slopes and traveltime in the inversion. Then, more complex subsurface models such as the Marmousi and BP TTI salt models are used to assess the potentialities of this method in more realistic settings. Application on real data is presented in chapter 4.

3.1 Review on inverse problem theory

The forward modelling applies a physical theory on a set of parameters or quantities (model parameters) to predict the observables (data) of an experiment. Classic physics states that, as long as the model parameters are not changed, the observables are the same. The inverse problems arise when we do not know, or we partially know, the model parameters but we have some observables and we wish to find the related model parameters (Lee et al., 2002). In the framework of exploration seismology, observables consist of the seismograms, and the physical

properties of the Earth interior that govern the seismic wave propagation are the model parameters we seek to reconstruct. Accordingly, if I consider that the Earth interior properties are gathered in the model vector $\mathbf{m} = [m_1, m_2, m_3, \cdots, m_M]^t$, where M denotes the number of model parameters, (the superscript t designates the transpose operator) and the observables from the recorded seismograms define the elements of the data vector $\mathbf{d}_{obs} = [d_1, d_2, d_3, \cdots, d_N]^t$, as N observables, the observed data and model parameters can be related by one or more equations

$$\begin{aligned} g_1(\mathbf{d}_{obs}, \mathbf{m}) &= 0, \\ g_2(\mathbf{d}_{obs}, \mathbf{m}) &= 0, \\ &\vdots \\ g_L(\mathbf{d}_{obs}, \mathbf{m}) &= 0, \end{aligned} \qquad \text{or} \qquad \mathbf{g}(\mathbf{d}_{obs}, \mathbf{m}) = \mathbf{0}. \tag{3.1}$$

where L is the number of equations and operator g_i governs the physics of seismic wave propagation within the Earth. Therefore, the inverse problem consists in solving the equation $\mathbf{g}(\mathbf{d}_{obs}, \mathbf{m}) = \mathbf{0}$ for the unknown set of model parameters $\mathbf{m}$. Depending whether these equations are linear or not, the inverse problems can be mainly categorized as linear or non-linear problems, the resolution of which are discussed in the following sections.

The goal of this section is to review some key features of inverse problem theory which are of concern in this study rather than providing a detailed review. For the sake of simplicity, I keep the formulation in the framework of unconstrained optimization problems.

3.1.1 Explicit linear inverse problems

If operator $\mathbf{g}$ in eq. (3.1) describes a linear relationship [1] between data and model parameters, I can rewrite eq.(3.1) as

$$\mathbf{d}_{obs} = \mathbf{F}\mathbf{m}, \tag{3.2}$$

where matrix $\mathbf{F}_{N\times M}$ leads to N equations ($N = L$). Solution of these linear equations can be achieved with

$$\mathbf{m} = \mathbf{F}^{-1}\mathbf{d}_{obs}, \tag{3.3}$$

where the superscript -1 designates the inverse operator. This exact solution is achievable only and only if $\mathbf{F}^{-1}$ exists. This leads us toward classifying the inverse problems by answering the question of whether the equation $\mathbf{d}_{obs} = \mathbf{F}\mathbf{m}$ provides enough information to determine uniquely the model parameters.

The problem is called under-determined if the equation $\mathbf{d}_{obs} = \mathbf{F}\mathbf{m}$ does not provide enough information to specify uniquely the model parameters. This typically happens when the number of unknowns is more than known (i.e. $M > N$). If the equation $\mathbf{d}_{obs} = \mathbf{F}\mathbf{m}$ contains too much information such that determining an exact solution is not possible, the problem is over-determined. In such a problem the number of observables is higher than the number of model parameters; generally this occurs when $N > M$. The problem is even-determined if there is enough information to specify the model parameters.

1. Operator g is linear if $g(\beta_1\mathbf{m}_1 + \beta_2\mathbf{m}_2) = \beta_1 g(\mathbf{m}_1) + \beta_2 g(\mathbf{m}_2)$.

However, in many applications, the inverse problem is neither completely over-determined or under-determined, and it can be considered as a mixed-determined problem (Menke, 2012, chapter 3). This makes the direct calculation of the exact model parameters with $\mathbf{m} = \mathbf{F}^{-1}\mathbf{d}_{obs}$ impossible, and requires to use numerical optimization methods (e.g., the least-squares method) to solve the inverse problem.

Least-squares method

I consider a more general linear inverse problem in which the observables are contaminated with noise. This modifies the relation (3.2) to

$$\mathbf{d}_{obs} = \mathbf{F}\mathbf{m} + \mathbf{e}, \tag{3.4}$$

where "$\mathbf{e}$" designates an unknown error. A widely used method for resolution of this equation is the least-squares method which aims to minimize the misfit between the observables and the predicted data $\mathbf{F}\mathbf{m}$. This leads to defining a misfit function as

$$C(\mathbf{m}) = \mathbf{e}^t\mathbf{e} = (\mathbf{d}_{obs} - \mathbf{F}\mathbf{m})^t(\mathbf{d}_{obs} - \mathbf{F}\mathbf{m}) = \|(\mathbf{d}_{obs} - \mathbf{F}\mathbf{m})\|_2^2, \tag{3.5}$$

where $\|\cdot\|_2$ designates the l_2-norm operator. The estimated solution of eq.(3.4) is $\tilde{\mathbf{m}} = \underset{\mathbf{m}}{\operatorname{argmin}}\ C(\mathbf{m})$, i.e. $\tilde{\mathbf{m}}$ is the minimizer of the misfit function $C(\mathbf{m})$. Therefore, zeroing the derivative of misfit function (3.5) with respect to $\mathbf{m}$ ($\nabla C(\mathbf{m}) = 0$) results in $\tilde{\mathbf{m}}$

$$\nabla C(\mathbf{m}) = 0 \quad \rightarrow \quad \mathbf{F}^t\mathbf{F}\mathbf{m} - \mathbf{F}^t\mathbf{d}_{obs} = 0, \tag{3.6}$$

and

$$\tilde{\mathbf{m}} = (\mathbf{F}^t\mathbf{F})^{-1}\mathbf{F}^t\mathbf{d}_{obs}. \tag{3.7}$$

If all the observables do not have the same uncertainty, we need to bias the inversion toward the observables which are more reliable. Therefore, for such a problem it is necessary to modify the misfit function (3.5) as follow

$$C(\mathbf{m}) = (\mathbf{d}_{obs} - \mathbf{F}\mathbf{m})^t\mathbf{C}_d^{-1}(\mathbf{d}_{obs} - \mathbf{F}\mathbf{m}), \tag{3.8}$$

where $\mathbf{C}_d$ is the data covariance matrix and includes each observable uncertainty. This $N \times N$ diagonal matrix weights inversely the residual term $\mathbf{d}_{obs} - \mathbf{F}\mathbf{m}$ such that, the observables with large uncertainty have less contribution to the inversion, and those with small uncertainty have more contribution and steer the inversion.

Moreover, if different classes of observables were recorded with different order of magnitudes, it is desirable to define a weighting matrix ($\mathbf{W}_d$) in the misfit function to balance the role of each class of observables during the inversion. For this case, I can rewrite the misfit function (3.5) as

$$C(\mathbf{m}) = (\mathbf{d}_{obs} - \mathbf{F}\mathbf{m})^t\mathbf{W}_d(\mathbf{d}_{obs} - \mathbf{F}\mathbf{m}), \tag{3.9}$$

where $\mathbf{W}_d$ balances the residuals from different observable classes.

For these modified misfit functions (eqs (3.8)-(3.9)) the estimated model parameters can be calculated in a similar way as in eq. (3.7). Estimated solutions with these misfit functions respectively are

$$\tilde{\mathbf{m}} = (\mathbf{F}^t \mathbf{C}_d^{-1} \mathbf{F})^{-1} \mathbf{F}^t \mathbf{C}_d^{-1} \mathbf{d}_{obs}, \tag{3.10}$$

$$\tilde{\mathbf{m}} = (\mathbf{F}^t \mathbf{W}_d \mathbf{F})^{-1} \mathbf{F}^t \mathbf{W}_d \mathbf{d}_{obs}. \tag{3.11}$$

The estimated model $\tilde{\mathbf{m}}$ depends directly on the observables and generalized inverse matrix $(\mathbf{F}^t\mathbf{F})^{-1}\mathbf{F}^t$. This arises following important debates about inverse problem solutions (Sen and Stoffa, 2013): existence and uniqueness of solution, robustness (insensitivity of the solutions to outliers in the observables) and stability (insensitivity of the solutions to small changes in the observables).

Non-uniqueness of solutions means there are at least two different models which satisfy the observables. Some of the reasons for non-unique solutions are: 1) the model discretization (e.g., seismic tomographic methods discretize the subsurface model to finite elements while the Earth interior is a continuous function of coordinates), 2) limit in model identification by data or modelling operator (e.g. in seismic ray based tomography the shadow zones are not identified by modelling operator. Different values can be considered in the shadow zones while the final predicted data are same), 3) cross-talk between model parameters (e.g., in the framework of anisotropic seismic tomography, the leakage between vertical velocity and ϵ and δ can generate different combinations of these parameter classes which result in the same predicted data), 4) singularity of the square matrix $\mathbf{F}^t\mathbf{F}$. A large condition number [2] for this matrix can be a reason for instability of the estimated solutions.

Lack of the uniqueness, robustness and/or stability properties lead to an ill-posed inverse problem. The ill-posedness of inverse problems, as well as the costly calculation of the generalized inverse matrix for large data sets, make the iterative methods more desirable where the inversion starts with an initial model and updates the model parameters during the iterations. As a well known iterative approach, I can mention the steepest descent (for a review refer to Lines and Treitel (1984)). I shall discuss more about the iterative approaches in the sections concerning the resolution of non-linear inverse problems. In the following I briefly address the regularization approach which can reduce the ill-posedness of inverse problems (both linear and non-linear).

Regularization

The objective of the regularization theory is to penalise the unknown model parameters by adding a penalty term (or so-called side-constraint (Sen and Stoffa, 2013)) to the minimization of misfit function (3.5) as

$$\min \ \|(\mathbf{d}_{obs} - \mathbf{F}\mathbf{m})\|_2^2 \qquad \text{subject to} \qquad \mathbf{L}\mathbf{m} < d, \tag{3.12}$$

where $\mathbf{L}$ is the regularization operator and d is an unknown threshold. Here, operator $\mathbf{L}$ imposes some specific properties to the estimated model parameters. For instance, a Gaussian filter

2. Condition number measures how much the output of operators are sensitive to the changes in inputs.

imposes the smoothness, and l_2-norm operator controls the energy of the model parameters. It is important to note that, since the desirable d is unknown, one can not call this problem a constrained minimization problem, and in fact the term "side-constraint" is not a proper name for $\mathbf{Lm} < d$.

In the framework of Tikhonov regularization (Tikhonov, 1963) I can penalise the solutions by modifying the misfit function (3.8) as

$$C(\mathbf{m}) = (\mathbf{d}_{obs} - \mathbf{Fm})^t \mathbf{C}_d^{-1} (\mathbf{d}_{obs} - \mathbf{Fm}) + \beta(\mathbf{m}^t \mathbf{Lm}), \tag{3.13}$$

where $\beta \in IR$ is the regularization parameter. In the framework of the least-squares method, the estimated model from regularized misfit function (3.13) is

$$\tilde{\mathbf{m}} = (\mathbf{F}^t \mathbf{C}_d^{-1} \mathbf{F} + \beta \mathbf{L})^{-1} \mathbf{F}^t \mathbf{C}_d^{-1} \mathbf{d}_{obs}. \tag{3.14}$$

The regularization parameter controls the relative contribution of the misfit term and penalty term in the model parameter estimation. A relative large β causes the calculated data ($\mathbf{d}_{clc} = \mathbf{F}\tilde{\mathbf{m}}$) do not fit the observables completely, and a relative small β results in well data fitting but poor regularized model parameters. On the other hand, term $\beta\mathbf{L}$ can remove the singularity of generalized inverse matrix and control the condition number of the inverse operator $(\mathbf{F}^t \mathbf{C}_d^{-1} \mathbf{F} + \beta \mathbf{L})^{-1}$. There are different approaches to set β which among them are L-curve, U-curve and Generalized Cross Validation (CGV) methods (Aster et al., 2004).

3.1.2 Non-linear inverse problems

In many inverse problems the relation between the data and model parameters is not linear. This leads to the definition of non-linear inverse problems. Considering the relation $\mathbf{g}(\mathbf{d}, \mathbf{m}) = \mathbf{0}$ of a general inverse problem, this non-linearity can be in data and/or model parameters. Here, I suppose the inverse problem is only non-linear in data space, in another word,

$$\mathbf{d} = \mathbf{f}(\mathbf{m}). \tag{3.15}$$

I define the l_2-norm misfit function $C(\mathbf{m})$ as

$$C(\mathbf{m}) = \frac{1}{2}(\mathbf{d}_{obs} - \mathbf{f}(\mathbf{m}))^t (\mathbf{d}_{obs} - \mathbf{f}(\mathbf{m})), \tag{3.16}$$

the minimizer of which is the estimated solution of eq.(3.15). This leads to an unconstrained optimization problem as

$$\min_{\mathbf{m}} C(\mathbf{m}), \tag{3.17}$$

where the optimization parameters are gathered in the vector $\mathbf{m}$. There are different algorithms to solve such an optimization problem and Nocedal and Wright (2006) listed the following properties for a proper algorithm:

— Robust: applicable on a wide range of problems for all reasonable initial models.

— Efficient: not expensive from computational cost and storage requirement point of views.

— Accurate: able to find a solution with precision and not sensitive to error in data and arithmetic rounding error.

However, there are some trade-offs between these properties. For example, there are some robust algorithms which are slow and there are some fast methods which require a huge memory (refer to Castellanos et al. (2015) for an illustration). Therefore, in order to choose a suitable optimization approach, one needs to forge a compromise between these properties.

Since for many applications the misfit functions (for both linear and non-linear problems) are not strictly convex, there are the concepts of local and global minimizers (Fig. 3.1), and accordingly the optimization algorithms are divided in two groups: global optimization and local optimization methods.

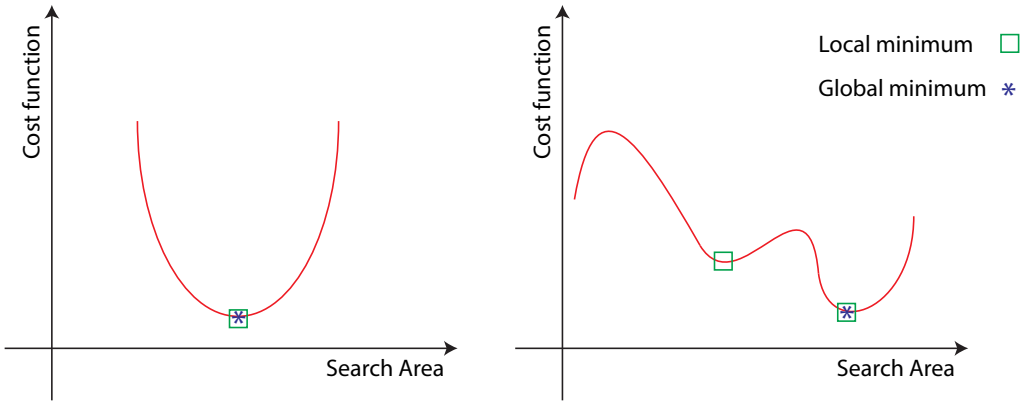

Figure 3.1 – Schematic presentation of local and global minimums. a) For a strictly convex misfit function there is only one local minimum which it is also the global. b) For many applications the misfit function can have both global and local minimum.

Global optimization methods

In some applications the global minimizer of misfit function is required and this entails applying a global optimization methods (Dixon and Szego, 1978; Floudas, 2013; Grossmann, 2013). These methods can be classified in two branches: deterministic and stochastic methods.

Deterministic methods perform an exhaustive search over the search area of model parameters and find the global minimizer. For example, at the expense of computational cost, in the framework of "finite exact methods" one can ensure the success of optimization in finite iteration number (Arora et al., 1995).

The stochastic methods are based on random search. These ensure, in a probabilistic sense, that the algorithm will find the global minimum. The pure random search is the simplest method in this category in which through an inefficient approach for k random samples of search area one calculates the misfit function and considers the minimum one as the global minimum.

Among stochastic methods, perhaps, the most well-known is the Genetic algorithm in which the random selection procedure is replaced by the process of natural selection (Goldberg and Kuo, 1987; Hajela, 1990). Here, the algorithm starts with a population of candidate solutions which is called a generation. A set of better populations, which has smaller misfit function value, is stochastically selected from the given generation while they are allowed to reproduce and cross among themselves with the aim of producing a better population (Arora et al., 1995). In result, there is a better generation, like a natural population of biological creatures, which it is more likely to be the solution of optimization problem. This iterative procedure continues to satisfy a convergence criterion like maximum number of produced generations (Arora et al., 1995).

Although the global optimization algorithms from computational cost point of view are expensive, but recent achievements in computer science make some of them applicable. Sen and Stoffa (2013) review some of the global optimization approaches in geophysical problem. Also, in the context of velocity macro-model building, global optimization methods are applied (Datta and Sen, 2016; Sajeva et al., 2016).

Local optimization

Iterative local optimization approaches start with an initial guess for the model parameters and try to update them toward the global minimizer. These algorithms use some information about misfit function (such as the value and first or/and second order derivatives with respect to the model parameters) in the current iteration or, depending on the method, earlier iterations to define the update direction. If the initial guess be close to the global minimizer, converging to the global minimizer is more probable.

If I consider the updated model $\mathbf{m}$ as the perturbed version of a known reference model like $\mathbf{m}_0$, then I can write

$$\mathbf{m} = \mathbf{m}_0 + \mathbf{\Delta m}, \tag{3.18}$$

where $\mathbf{\Delta m}$ is the perturbation model and $\mathbf{\Delta m} \ll \mathbf{m}_0$. Therefore, the minimum of the misfit function $C(\mathbf{m})$ in (3.16) is sought in the vicinity of the reference model $\mathbf{m}_0$. A second-order Taylor's expansion of the misfit function reads

$$C(\mathbf{m}_0 + \mathbf{\Delta m}) = C(\mathbf{m}_0) + \mathbf{\Delta m}^t \nabla C(\mathbf{m}_0) + \frac{1}{2}\mathbf{\Delta m}^t \nabla^2 C(\mathbf{m}_0)\mathbf{\Delta m} + \mathcal{O}(\mathbf{m}^3). \tag{3.19}$$

By taking derivative with respect to the model parameters I can write

$$\nabla C(\mathbf{m}) = \nabla C(\mathbf{m}_0) + \nabla^2 C(\mathbf{m}_0)\mathbf{\Delta m}. \tag{3.20}$$

At the minimum of the misfit function, the first-order derivative of the misfit function is zero. This leads to the perturbation model as

$$\mathbf{\Delta m} = -[\nabla^2 C(\mathbf{m}_0)]^{-1} \nabla C(\mathbf{m}_0). \tag{3.21}$$

This relation (3.21) defines the descent direction as the product of the inverse Hessian and

gradient vector. Here the first-order derivative of cost function and the Hessian are calculated for $\mathbf{m} = \mathbf{m}_0$.

Considering the definition of misfit function (3.16), the derivative of the misfit function with respect to the model parameters reads

$$\begin{aligned} \nabla C(\mathbf{m}) &= -\Big(\frac{\partial \mathbf{f}(\mathbf{m})}{\partial \mathbf{m}}\Big)^t (\mathbf{d}_{obs} - \mathbf{f}(\mathbf{m})) \\ &= -\Big(\frac{\partial \mathbf{d}_{clc}}{\partial \mathbf{m}}\Big)^t (\mathbf{d}_{obs} - \mathbf{d}_{clc}) = -\mathbf{J}^t \boldsymbol{\Delta}\mathbf{d}, \end{aligned} \tag{3.22}$$

where $\mathbf{J}$ is the Fréchet derivative or sensitivity matrix and $\mathbf{d}_{clc}$ stands for the calculated (predicted) data. Accordingly, the Hessian can be derived with differentiation of the expression (3.22)

$$\begin{aligned} \nabla^2 C(\mathbf{m}) &= \mathbf{J}^t\mathbf{J} + \Big[\frac{\partial \mathbf{J}^t}{\partial m_1}\boldsymbol{\Delta}\mathbf{d} + \frac{\partial \mathbf{J}^t}{\partial m_2}\boldsymbol{\Delta}\mathbf{d} + \cdots + \frac{\partial \mathbf{J}^t}{\partial m_M}\boldsymbol{\Delta}\mathbf{d}\Big] \\ &= \mathbf{J}^t\mathbf{J} + \Big[\frac{\partial \mathbf{J}^t}{\partial \mathbf{m}^t}(\boldsymbol{\Delta}\mathbf{d}\cdots\boldsymbol{\Delta}\mathbf{d})\Big]. \end{aligned} \tag{3.23}$$

For misfit functions which are quadratic function of $\mathbf{m}$ (this is the case for the linear problem $\mathbf{d} = \mathbf{F}\mathbf{m}$), the error term $\mathcal{O}(\mathbf{m}^3)$ and the second term in the Hessian are zero and the expression (3.21) gives the desired perturbation (the one that makes $\mathbf{m} = \mathbf{m_0} + \boldsymbol{\Delta}\mathbf{m}$ the minimizer of the misfit function) in one iteration.

For non-linear problems several iterations are needed to take into account the non zero error term $\mathcal{O}(\mathbf{m}^3)$ in the updates of $\mathbf{m}$. Here, the first step is adopting a strategy for the model updating procedure and then calculation of the required information such as the gradient of misfit function, sensitivity matrix and Hessian. In the following sections, I first address the line search method (Nocedal and Wright, 2006) as the updating strategy and then I shall review a approach for the gradient calculation.

The above discussed approach based on the assumption of $\mathbf{m} = \mathbf{m}_0 + \boldsymbol{\Delta}\mathbf{m}$, is called the linearization of the inverse problem and this raises the discussion about possibility of linearization of the forward modelling. Therefore, before following the discussion about resolution of non-linear inverse problems, I briefly address linearization of forward modelling.

Linearization of forward modelling

Considering $\mathbf{m} = \mathbf{m}_0 + \boldsymbol{\Delta}\mathbf{m}$ and the expression $\mathbf{d} = \mathbf{f}(\mathbf{m})$, the first-order Taylor's expansion of the forward modelling around the reference model $\mathbf{m}_0$ reads

$$\mathbf{f}(\mathbf{m}_0 + \boldsymbol{\Delta}\mathbf{m}) = \mathbf{f}(\mathbf{m}_0) + \frac{\partial \mathbf{f}(\mathbf{m}_0)}{\partial \mathbf{m}}\boldsymbol{\Delta}\mathbf{m} + \mathcal{O}(\mathbf{m}^2), \tag{3.24}$$

where for $|\boldsymbol{\Delta}\mathbf{m}| \ll \mathbf{m}_0$, for some forward modellings I can ignore the error term $\mathcal{O}(\mathbf{m}^2)$ and write

$$\mathbf{d} = \mathbf{d}_0 + \frac{\partial \mathbf{f}(\mathbf{m}_0)}{\partial \mathbf{m}}\boldsymbol{\Delta}\mathbf{m}, \tag{3.25}$$

where $\mathbf{d}_0 = \mathbf{f}(\mathbf{m}_0)$. In a compact form I rewrite this expression as

$$\mathbf{d} - \mathbf{d}_0 = \mathbf{\Delta d} = \mathbf{F}_0 \mathbf{\Delta m}. \tag{3.26}$$

Here $\mathbf{F}_0$ designates the Fréchet derivative matrix. This expression linearly links the model parameter perturbations to the data perturbations, and it is similar to the linear inverse problem (3.2) while here the inversion objective is estimation of $\mathbf{\Delta m}$. Therefore, one can use a linear inverses problem solver to estimated the solution of eq. (3.26). Note that, for the non-linear forward modelling the error term $\mathcal{O}(\mathbf{m}^2)$ is not zero and estimation of the perturbation $\mathbf{\Delta m}$ requires several iterations. As an example for the forward modelling linearization, Hole (1992) linearized the forward modelling in first arrival traveltime tomography.

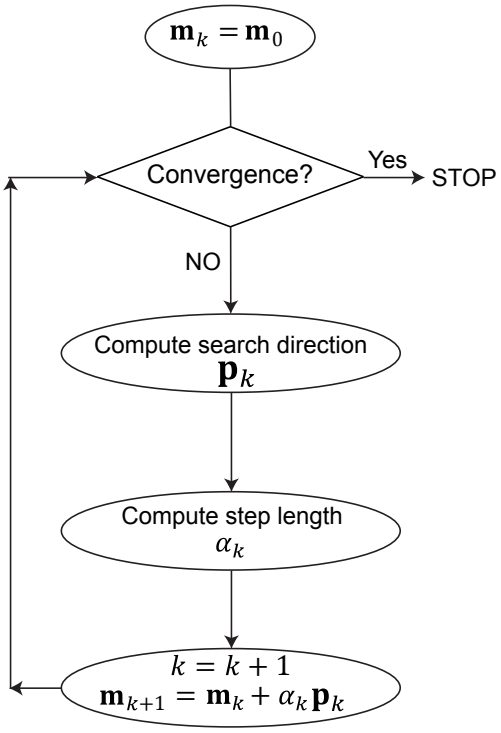

Figure 3.2 – A general flowchart for local optimization approaches based on the line search strategy. Convergence is achieved when there is no more update in model or the updated model is approximated with certain accuracy. $\mathbf{m}_0$ is the initial guess for model (figure is adopted from Sen and Stoffa (2013))

3.1.3 Line search strategy

In iterative local optimization schemes, there are different strategies for moving from one model update to the next one, and one of the widely used strategies is the line search.

During each iteration of the line search (Fig. 3.2), an update along the defined search direction $\mathbf{p}_k$ (index k designates the iteration number for the model updates) is sought such that the updated model

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha \mathbf{p}_k \tag{3.27}$$

reduces the misfit function. In other words, this strategy can be summarized in the following one dimensional optimization problem

$$\min_{\alpha_k \in IR^+} C(\mathbf{m}_k + \alpha_k \mathbf{p}_k), \tag{3.28}$$

where $\alpha \in IR^+$ is the step length. But, in practice the exact solution of this minimization problem is not required and with trial and error one can find the approximated minimizer. However, defining a suitable update for the iterate $\mathbf{m}_k$ depends on both search direction and the step length.

Search direction

The search direction can be defined in a general form as (Nocedal and Wright, 2006)

$$\mathbf{p}_k = -\mathbf{B}_k^{-1} \nabla C(\mathbf{m}_k), \tag{3.29}$$

where B_k is a symmetric and non-singular matrix. Definition of this matrix varies depending on the local optimization methods.

For the steepest descent method the search direction is the steepest descent where $\mathbf{p}_k = -\nabla C(\mathbf{m}_k)$ (i.e. $\mathbf{B}$ is identity matrix). All the search directions with less than $\pi/2$ radian deviation from $-\nabla C(\mathbf{m}_k)$ guarantee the misfit function reduction. This can be understood by Taylor's series. For small α_k I can write

$$C(\mathbf{m}_k + \alpha_k \mathbf{p}_k) \approx C(\mathbf{m}_k) + \alpha_k \mathbf{p}_k^t \nabla C(\mathbf{m}_k) + \mathcal{O}(\alpha_k^2), \tag{3.30}$$

where

$$\mathbf{p}_k^t \nabla C(\mathbf{m}_k) = \|\mathbf{p}_k\| \|\nabla C(\mathbf{m}_k)\| \cos\theta \leq 0. \tag{3.31}$$

Here θ is the angle between $\nabla C(\mathbf{m}_k)$ and search direction $\mathbf{p}_k$ and $\pi/2 < \theta < \pi$.

For the Gauss-Newton methods $\mathbf{B}_k = \mathbf{J}^t\mathbf{J}$, i.e. the approximate Hessian in which one ignores the second term of the full Hessian definition in eq. (3.23). In the Newton methods $\mathbf{B}_k = \nabla^2 C(\mathbf{m})$ and the updates are based on the full Hessian information.

The quasi-Newton methods define $\mathbf{B}_k$ as an approximate expression of the Hessian, and updates this approximation during the iterations by using the misfit function information in current and earlier iterations.

The truncated Newton method also uses the approximation of the Hessian to define the search direction while this approximation only depends on the current iteration. This requires to solve a linear equation through iterative solvers. This strategy accounts for the higher eigenvalues of the inverse Hessian operator (Métivier et al., 2013).

The Newton approaches generally are fast with quadratic local convergence rate but they require calculation of the Hessian which can be expensive and difficult to achieve for some optimization problems. Also, for the Gauss-Newton methods the calculation of sensitivity matrix $\mathbf{J}$ can be expensive for large data sets. These drawbacks of the Newton and Gauss-Newton methods make the quasi-Newton approaches attractive alternatives where explicit calculation of the sensitivity matrix and the Hessian can be avoided. In this study I apply the L-BFGS method (Byrd et al., 1995) (a quasi-Newton method) to handle the inverse problem in slope tomography. In the next sections I shall explain the basis of this method.

Step length calculation

As mentioned before, the ideal value for α_k is the minimizer of minimization problem (3.28) which I rewrite it for a given search direction $\mathbf{p}_k$ as

$$\min_{\alpha_k \in IR^+} \Phi(\alpha_k), \tag{3.32}$$

where $\Phi(\alpha_k) = C(\mathbf{m}_k + \alpha_k \mathbf{p}_k)$. This calculation is costly and even performing a local optimization can be expensive due to repetitive misfit function calculation and probably its gradient. Therefore, we perform an inexact line search to find a minimizer which reduce adequately the misfit function. This procedure consists of two step: bracketing to find an interval containing a suitable step length, interpolation to extract a step length within the interval.

A necessary but not sufficient condition in the step length selection is the misfit function reduction. Here, I apply the Wolfe conditions to satisfy this criterion and ensure a suitable step length calculation. The Wolf conditions consist of "sufficient decrease" and "curvature" conditions (Fig. 3.3). The sufficient decrease condition reads

$$C(\mathbf{m}_k + \alpha_k \mathbf{p}_k) \leq C(\mathbf{m}_k) + c_1 \alpha_k \mathbf{p}_k^t \nabla C(\mathbf{m}_k), \tag{3.33}$$

where $c_1 \in (0, 1)$. This inequality states that the reduction in the misfit function should be proportional with α_k and the directional gradient $\mathbf{p}_k^t \nabla C(\mathbf{m}_k)$. Here the right hand side is a linear function of α_k and it is indicated by $l(\alpha)$ in Fig. 3.3. Since for many sufficient small values of α this inequality can be satisfied, this condition is not enough to ensure a reasonable progress in the search for α. In order to avoid these unacceptably small steps one also need to consider the curvature condition.

The curvature condition reads

$$\mathbf{p}_k^t \nabla C(\mathbf{m}_k + \alpha_k \mathbf{p}_k) \geq c_2 \mathbf{p}_k^t \nabla C(\mathbf{m}_k), \tag{3.34}$$

where $c_2 \in (c_1, 1)$ [3]. According to this inequality, the slope $\Phi'(\alpha_k) = \mathbf{p}_k^t \nabla C(\mathbf{m}_k + \alpha_k \mathbf{p}_k)$ should be greater than c_2 times the initial slope $\Phi'(0) = \mathbf{p}_k^t \nabla C(\mathbf{m}_k)$. Considering $\mathbf{p}_k$ as the descent direction, if the slope $\Phi'(\alpha_k)$ be strongly negative we ensure the misfit function decreases in this direction (Fig. 3.3), while if the slope $\Phi'(\alpha_k)$ be slightly negative or positive the misfit

3. The typical value of c_2 for the quasi-Newton approaches is 0.9.

function does not decrease more in this direction and we can stop the line search (Nocedal and Wright, 2006).

Therefore, in each iteration of the model updating, there is an inner loop for α_k^j. Here, the index k refers to the model updates while j designates the iterations in the loop over α_k. For each iteration k I initialize the inner loop with an initial guess as α_k^0 from a predefined range. In each iteration of loop over α I narrow the search area by bracketing. Candidate α_k^j in the current iteration j is derived by the interpolation of search area boundary values. The candidate which satisfies the Wolfe conditions, is considered the step length for iteration k. With this α_k and the given search direction $\mathbf{p}_k$ one can update the model parameters through $\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \mathbf{p}_k$.

3.1.4 L-BFGS method

The Limited-memory BFGS (L-BFGS) is one of the quasi-Newton approaches which I shall implement in the slope tomography inversion. This method is the modified version of the BFGS method which, compared to the BFGS method, reduces the memory requirement for in the resolution of inverse problems.

At each iteration of the BFGS method, the updated model is

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \mathbf{H}_k \nabla C(\mathbf{m}_k). \tag{3.35}$$

Here "$\mathbf{H}_k$" stands for the Hessian approximation at iteration k and during the iterations is updated through

$$\mathbf{H}_{k+1} = \mathbf{V}_k^t \mathbf{H}_k \mathbf{V}_k + \rho_k \mathbf{s}_k \mathbf{s}^t, \tag{3.36}$$

where

$$\rho_k = \frac{1}{\mathbf{y}_k^t \mathbf{s}_k}, \qquad \mathbf{V}_k = \mathbf{I} - \rho_k \mathbf{y}_k \mathbf{s}_k^t, \tag{3.37}$$

($\mathbf{I}$ is identity matrix) and

$$\mathbf{s}_k = \mathbf{m}_{k+1} - \mathbf{m}_k, \qquad \mathbf{y}_k = \nabla C(\mathbf{m}_{k+1}) - \nabla C(\mathbf{m}_k). \tag{3.38}$$

This approximation for the Hessian circumvents the cumbersome calculation of the full Hessian but still storing and manipulation of this matrix for large model size is expensive. To avoid this problem, the L-BFGS approach proposes a modified version of Hessian in eq. (3.36).

According to the L-BFGS method, one stores n recent vector pairs $\{\mathbf{s}_i, \mathbf{y}_i\}$ and calculate the approximate Hessian by

$$\begin{aligned}
\mathbf{H}_k =& (\mathbf{V}_{k-1}^t \cdots \mathbf{V}_{k-m}^t) \mathbf{H}_k^0 (\mathbf{V}_{k-m} \cdots \mathbf{V}_{k-1}) \\
&+ \rho_{k-m} (\mathbf{V}_{k-1}^t \cdots \mathbf{V}_{k-m+1}^t) \mathbf{s}_{k-m} \mathbf{s}_{k-m}^t (\mathbf{V}_{k-m} \cdots \mathbf{V}_{k-1}) \\
&+ \rho_{k-m+1} (\mathbf{V}_{k-1}^t \cdots \mathbf{V}_{k-m+2}^t) \mathbf{s}_{k-m+1} \mathbf{s}_{k-m+1}^t (\mathbf{V}_{k-m+1} \cdots \mathbf{V}_{k-1}) \\
&\cdots \\
&+ \rho_{k-1} \mathbf{s}_{k-1} \mathbf{s}_{k-1}^t,
\end{aligned} \tag{3.39}$$

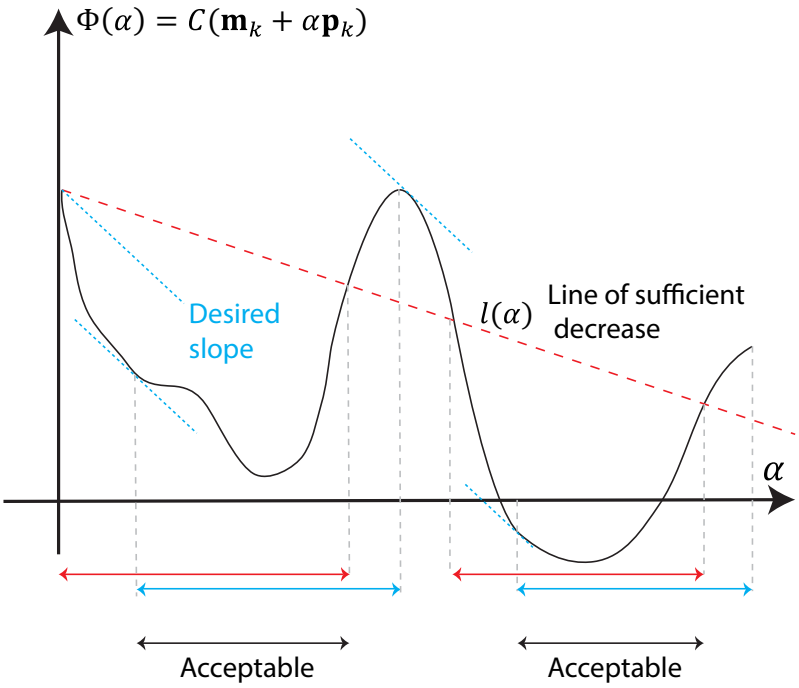

Figure 3.3 – The Wolfe conditions. The sufficient decrease condition restricts the search area for α to the red arrows where $\Phi(\alpha) \leq l(\alpha)$. The curvature condition avoids unacceptably small α steps and ensures a relatively big reductions in the $\Phi(\alpha)$ value by considering those α's (indicated by the blue arrows) which make $\Phi'(\alpha)$ greater than desired slope $c_2\Phi'(0)$ (figure is adopted from Nocedal and Wright (2006)).

which includes only some inner products and vector summations. Here $\mathbf{H}_k^0$ is the initial guess for the Hessian in each iteration and simply can be the identity matrix. After each iteration I replace the oldest pair of stored $\{\mathbf{s}_i, \mathbf{y}_i\}$ with the newer version. Through this approach curvature information of n recent iterations is considered for the model updates (more detail in Nocedal and Wright (2006, section 7.2)). In practice $3 \leq n \leq 20$ is sufficient. This method provides in a cheap way an approximate Hessian and it is a reliable candidate for large scale problems like full waveform inversion (Métivier and Brossier, 2016). The above discussion on the L-BFGS implementation leads us toward the calculation of misfit function gradient. In the following I discuss how to perform this calculation.

3.1.5 Gradient calculation

One of the main ingredients of the local optimization techniques, such as quasi-Newton methods, is the calculation of misfit function gradient with respect to the model parameters. This gradient can be calculated directly by applying the finite difference method but this is really expensive and not always precise (Chavent, 2009, section 2.1). According to eq. (3.22), this gradient can be achieved through $\nabla C(\mathbf{m}) = -\mathbf{J}^t \boldsymbol{\Delta}\mathbf{d}$ where $\mathbf{J} = \partial \mathbf{d}_{clc}/\partial \mathbf{m}$ is the Fréchet

derivative matrix.

For a general physical inverse problem, one can consider that the data depends on the so-called state variables which these variables are a function of the model parameters. Therefore, the Fréchet derivative matrix includes the derivative of the state variables with respect to the model parameters (Plessix, 2006). The relation between the state variables and model parameters are set through the state equations (also can be called forward modelling equations). In the context of the reflection tomography, the state equations are the ray tracing equations and the state variables are the spatial coordinates and the slowness vector of the rays. This defines the Fréchet derivative matrix as the derivative of the spatial coordinates and the slowness vector with respect to to the background slowness and the reflector positions (Plessix, 2006).

If I define a non-linear inverse problem as $\mathbf{d} = \mathbf{R}(\mathbf{u})$ (where $\mathbf{u}$ is the vector which gathers the state variables, and R is the observation operator) and consider the state equation as $\mathbf{h}(\mathbf{m}, \mathbf{u}) = 0$, derivation of the state equation with respect to the model parameters reads

$$\frac{\partial \mathbf{h}}{\partial m_i}(\mathbf{m}, \mathbf{u}) + \frac{\partial \mathbf{h}}{\partial \mathbf{u}}(\mathbf{m}, \mathbf{u}) \frac{\partial \mathbf{u}}{\partial m_i} = 0 \qquad i = 1 \cdots M. \tag{3.40}$$

By solving the M linearized equations in expression (3.40) one can derive $\dfrac{\partial \mathbf{u}}{\partial m_i} = 0$, and calculate each column of Fréchet derivative matrix $\mathbf{J}$ through

$$J_i = \mathbf{R}'(\mathbf{u}) \frac{\partial \mathbf{u}}{\partial m_i}, \tag{3.41}$$

where $\mathbf{R}'(\mathbf{u})$ is derivative of the observation operator and $\mathbf{J} = [J_1, J_2, \cdots, J_M]$. Then, the misfit function gradient can be achieved by $\nabla C(\mathbf{m}) = -\mathbf{J}^t(\mathbf{d_{obs}} - \mathbf{R}(\mathbf{u}))$.

On the other hand, one can take advantage of the forward modelling linearization, eq. (3.26), to perform M modellings and calculate the perturbation in data due to the perturbation in the model parameters. However, these approaches can be expensive and memory demanding for large scale problems and this promotes taking advantage of the adjoint state method for calculation of the misfit function gradient.

Adjoint state method

The adjoint state method (Chavent, 1974; Akcelik et al., 2002; Plessix, 2006) provides a framework for calculation of a misfit function gradients with respect to its model parameters without explicitly building the sensitivity matrix. Here, I use the method of Lagrangian multipliers to establish this method.

By considering the non-linear inverse problem introduced above,

$$\begin{cases} \mathbf{d} = \mathbf{R}(\mathbf{u}), \\ \mathbf{h}(\mathbf{m}, \mathbf{u}) = \mathbf{0}, \end{cases} \tag{3.42}$$

I can rewrite the unconstrained optimization problem

$$\min_{\mathbf{m}} \|\mathbf{d}_{obs} - \mathbf{R}(\mathbf{u})\|_2^2, \tag{3.43}$$

as the following constrained optimization problem

$$\min_{\mathbf{m}} \|\mathbf{d}_{obs} - \mathbf{R}(\mathbf{u})\|_2^2 \qquad \text{subject to} \qquad \mathbf{h}(\mathbf{m}, \mathbf{u}) = \mathbf{0}. \tag{3.44}$$

I introduce a Lagrangian function $\mathcal{L}$ and through its stationary points I derive solution of the constrained problem (3.44) according to the method of Lagrangian multiplier. This Lagrangian function reads

$$\mathcal{L}(\mathbf{u}, \boldsymbol{\lambda}, \mathbf{m}) = \frac{1}{2}(\mathbf{d}_{obs} - \mathbf{R}(\mathbf{u}))^t(\mathbf{d}_{obs} - \mathbf{R}(\mathbf{u})) + \boldsymbol{\lambda}^t\mathbf{h}(\mathbf{m}, \mathbf{u}), \tag{3.45}$$

where $\boldsymbol{\lambda}$ is the Lagrangian multiplier or adjoint state variable. Note that, in the Lagrangian definition the state variable does not depend on the model parameters. According to Karush-Kuhn-Tucker (KKT) conditions (also known as first order optimality conditions) the gradient of the Lagrangian with respect to $\mathbf{d}, \boldsymbol{\lambda}$ and $\mathbf{m}$ should be zero at optimum

$$\begin{Bmatrix} \nabla_{\mathbf{u}}\mathcal{L} \\ \nabla_{\boldsymbol{\lambda}}\mathcal{L} \\ \nabla_{\mathbf{m}}\mathcal{L} \end{Bmatrix} (\mathbf{u}, \boldsymbol{\lambda}, \mathbf{m}) = \mathbf{0}. \tag{3.46}$$

By inserting the Lagrangian (3.45) into this system of equations I conclude

$$\begin{Bmatrix} \nabla_{\mathbf{u}}\mathcal{L} \\ \nabla_{\boldsymbol{\lambda}}\mathcal{L} \\ \nabla_{\mathbf{m}}\mathcal{L} \end{Bmatrix} (\mathbf{u}, \boldsymbol{\lambda}, \mathbf{m}) = \begin{Bmatrix} -\mathbf{R}^t(\mathbf{d}_{obs} - \mathbf{R}(\mathbf{u})) + \nabla_{\mathbf{u}}\mathbf{h}(\mathbf{m}, \mathbf{u})^t\boldsymbol{\lambda} \\ \mathbf{h}(\mathbf{m}, \mathbf{u}) \\ \nabla_{\mathbf{m}}\mathbf{h}(\mathbf{m}, \mathbf{u})^t\boldsymbol{\lambda} \end{Bmatrix} = \mathbf{0}. \tag{3.47}$$

In order to solve the optimization problem (3.44) one needs to satisfy the KKT conditions. In other words, one should solve the optimization problem (3.47) where the optimization space includes $\mathbf{u}, \boldsymbol{\lambda}$ and $\mathbf{m}$ (full space method). Resolution of this non-linear system requires an iterative approach where

$$x_k = x_k + \alpha_k \Delta x_k. \tag{3.48}$$

Here x stands for the optimization space and Δx_k is the search direction. Similar to the line search method in section 3.1.2, one can choose different search directions (e.g. steepest descent and Newton search directions). But, from optimization point of view resolution of this system for large-scale problems is not feasible (Akcelik et al., 2002; van Leeuwen and Herrmann, 2015).

As an alternative approach, there is the so-called reduced space method (Akçelik, 2002) which is based on a block elimination of the constraints, and consequently leads to an unconstrained optimization problem (Akçelik, 2002; van Leeuwen and Herrmann, 2015). In reduced space approach the optimization space x is equivalent to $\mathbf{m}$. The procedure for block elimination is as follow

— Solve the block

$$\nabla_{\boldsymbol{\lambda}}\mathcal{L}(\mathbf{u},\lambda,\mathbf{m}) = \mathbf{h}(\mathbf{u},\mathbf{m}) = 0, \tag{3.49}$$

which requires the forward modelling and results in state variables.

— Solve the adjoint state equation

$$\nabla_{\mathbf{u}}\mathcal{L}(\mathbf{u},\lambda,\mathbf{m}) = -\mathbf{R}^t(\mathbf{d}_{obs} - \mathbf{R}(\mathbf{u})) + \nabla_{\mathbf{u}}\mathbf{h}(\mathbf{m},\mathbf{u})^t\boldsymbol{\lambda} = 0, \tag{3.50}$$

which results in the adjoint state variable $\boldsymbol{\lambda}$.

— Use the calculated adjoint state variable and derive the reduced gradient with

$$\nabla_{\mathbf{m}}\mathcal{L} = -\nabla_{\mathbf{m}}\mathbf{h}(\mathbf{m},\mathbf{u})^t\boldsymbol{\lambda}, \tag{3.51}$$

which is equivalent to gradient of the misfit function (3.43) with respect to $\mathbf{m}$.

Here the reduced gradient calculation is matrix-free and explicit building of the sensitivity matrix is not needed, but, in each iteration, it requires a forward modelling and resolution of adjoint state equation. The second order optimality condition results in calculation of the Hessian (van Leeuwen and Herrmann, 2015) which I do not discuss it here.

Analytical resolution of the KKT system for many problems is not possible and one needs to discretize the problem and use numerical methods. Here there are two approaches: discretiziation of the continuous KKT system, or defining the discrete Lagrangian and deriving the discrete KKT system (Akçelik, 2002). However, if forming the continuous system is possible, it can be used as a guide line for designing the discrete formulation (Chavent, 2009, section 2.7).

Note that, if the forward modelling operator be continuous, one can solve the continuous system to derive the exact gradient expressions and then discretize the gradient, or, discretize the forward modelling and form the discrete Lagrangian and calculate the gradient through a discrete system (discretized adjoint equations). In the later option, the gradients are the approximate derivative of discrete misfit function. But for problems with discrete modelling engine, the resulted derivatives of the discrete KKT system are the exact derivative of the discrete cost function (Chavent, 2009, chapter 2).

Plessix (2006) reviewed the adjoint state method for some geophysical applications such as the full waveform inversion, stereotomography and least-square migration. He obtained the reduced gradients through perturbation theory and augmented functional.

3.1.6 Scaling of model parameters

In non-linear multi-parameter inverse problems, the existence of different parameters with different order of magnitudes makes non-Newton based optimization algorithms poorly scaled, where the dominant parameter steers the inversion (Nocedal and Wright, 2006, p. 27). A common practice to make these inverse problems better-conditioned and make the model parameters dimensionless is to scale/adimensionalize parameters. If I define $\mathbf{M} = \{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\}$ as the model space for an inverse problem, the scaled/adimensionalized parameter $\tilde{\mathbf{m}}_1$ simply can be

defined as

$$\tilde{\mathbf{m}}_1 = \frac{\mathbf{m}_1}{\mathtt{m}_0}, \tag{3.52}$$

where $\mathtt{m}_0 \in \mathbb{R}_+$ is the scaling factor. The direct consequence of this scaling emerges in the misfit function gradient calculation and Hessian operator where

$$\frac{\partial C}{\partial \tilde{\mathbf{m}}_1} = \mathtt{m}_0 \frac{\partial C}{\partial \mathbf{m}_1}, \tag{3.53}$$

$$\frac{\partial^2 C}{\partial \tilde{\mathbf{m}}_1^2} = \mathtt{m}_0^2 \frac{\partial^2 C}{\partial \mathbf{m}_1^2}. \tag{3.54}$$

This scaling can influence partially the way the Hessian corrects for the leakage between parameter classes through the modification of the relative amplitudes of the off-diagonal blocks of the Hessian. This is in particular true when data are affected by noise and damping regularization is added to the Hessian (for an example in FWI refer to Operto et al. (2013)). As mentioned before, the influence of scaling depends also on the optimization scheme. The algorithms which take advantage of Hessian approximation or the exact Hessian, tune partially by themselves the cooperation of different parameter classes with different order of magnitudes. But, methods like steepest descent, which relies only on the gradient information have no control on the update steps, are more likely to fail without scaling of parameters.

3.2 Inverse problem in isotropic adjoint slope tomography

This section includes our paper, isotropic adjoint slope tomography (Tavakoli F. et al., 2017b).

Slope tomography based on eikonal solvers and the adjoint-state method

B. Tavakoli F., S. Operto, A. Ribodetti, J. Virieux

Published in *Geophysical Journal International* (2017) 209 (3): 1629-1647

3.2.1 Summary

Velocity macro-model building is a crucial step in the seismic imaging work-flow as it provides the necessary background model for migration or full waveform inversion. In this study, we present a new formulation of stereotomography that can handle more efficiently long-offset acquisition, complex geological structures and large-scale datasets. Stereotomography is a slope tomographic method based upon a semi-automatic picking of local coherent events. Each local coherent event, characterised by its two-way traveltime and two slopes in common-shot and common-receiver gathers, is tied to a scatterer or a reflector segment in the subsurface. Ray tracing provides a natural forward engine to compute traveltime and slopes but can suffer from non-uniform ray sampling in presence of complex media and long-offset acquisitions. Moreover, most implementations of stereotomography explicitly build a sensitivity matrix, leading to the resolution of large systems of linear equations, which can be cumbersome when large-scale datasets are considered. Overcoming these issues comes with a new matrix-free formulation of stereotomography: a factored eikonal solver based on the fast sweeping method to compute first-arrival traveltimes and an adjoint-state formulation to compute the gradient of the misfit function. By solving eikonal equation from sources and receivers, we make the computational cost proportional to the number of sources and receivers while it is independent of picked events density in each shot and receiver gather. The model space involves the subsurface velocities and the scatterer coordinates, while the dip of the reflector segments are implicitly represented by the spatial support of the adjoint sources and are updated through the joint localization of nearby scatterers. We present an application on the complex Marmousi model for a towed-streamer acquisition and a realistic distribution of local events. We show that the estimated model, built without any prior knowledge of the velocities, provides a reliable initial model for frequency-domain FWI of long-offset data for a starting frequency of 4Hz, although some artefacts at the reservoir level result from a deficit of illumination. This formulation of slope tomography provides a computationally efficient alternative to waveform inversion method such as reflection waveform inversion or differential-semblance optimization to build an initial model for prestack depth migration and conventional FWI.

3.2.2 Introduction

Building a velocity macro-model from reflection data remains one of the most crucial and challenging issues in seismic imaging. The difficulty arises from the nonlinearity of the inverse problem associated with the long-wavelength reconstruction of the subsurface; by contrast the imaging of reflectivity by migration is a far more linear problem (Claerbout, 1985). Velocity model building provides the necessary background or starting model to perform prestack-depth migration (Etgen et al., 2009) or full waveform inversion (FWI) (Tarantola, 1984a).

Among the most popular methods for velocity macro-model building, reflection traveltime tomography (Bishop et al., 1985; Farra and Madariaga, 1988) and migration-based velocity analysis (MVA) (Gardner et al., 1974; Al-Yahya, 1989) were specifically designed for seismic reflection data. Reflection traveltime tomography, which builds a velocity model through the minimization of the traveltime residuals, is computationally efficient but relies on tedious picking of continuous horizons. MVA methods rely on an explicit scale separation between the background velocities and the reflectivity to iteratively alternate the velocity update and the migration. The velocity update is driven by flattening the reflectors in the common image gathers or by minimizing reflection data residuals in the time domain after demigration. Among the MVA methods, some rely on picking in the prestack migration volume (Al-Yahya, 1989) or on more automatic waveform-based misfit criteria as in differential semblance optimization (DSO) (Symes, 1998; Chauris and Noble, 2001) or migration-based traveltime inversion (MBTT) (Clément et al., 2001). Recently, the governing ideas of the MBTT method have been recast in the framework of FWI, leading to the so-called reflection waveform inversion (RWI) (Xu et al., 2012; Brossier et al., 2015; Wu and Alkhalifah, 2015; Wang et al., 2016) with an extension to the joint inversion of diving waves and reflections proposed by Zhou et al. (2015). One drawback of MVA and RWI is the computational cost resulting from the migration performed at each cycle of the alternating optimization.

The computational cost of MVA and the above-mentioned variants of the FWI makes traveltime tomography an attractive technique for velocity model building, where we do not need repeated migrations. However, picking either in the data domain or in the migrated domain is a burden in traveltime tomography. As a remedy for the issue of horizon-based picking in reflection tomography, use of local coherent events in prestack datasets has been a breakthrough, first introduced by Rieber (1936) and followed by Riabinkin (1957) and Sword (1987) in the framework of the controlled directional reception (CDR) method. These local events are interpreted as arrivals reflected from small reflecting facets. Main motivation of CDR has been to add, in the data domain, the slopes of the local coherent events to the traveltimes to better constrain the velocity model building process.

Billette and Lambaré (1998) extended the CDR method to stereotomography, which relies on a semi-automatic picking of traveltimes and slopes of local coherent events in both shot and receiver gathers. Each of these local events is tied to a reflecting/diffracting facet in the subsurface. The misfit between recorded and modelled two-way traveltimes and slopes at the source and receiver positions is minimised to update the subsurface velocities and the coordinates of the facets in depth. Moreover, take-off angles and the one-way traveltimes of the rays connecting the facet to the source and the receiver can be added as optimization parameters with the aim

of making the forward problem more efficient by avoiding two-point ray tracing or making the inverse problem less prone to local minima by expending the search space. For this, the source and receiver positions need also to be fitted since there is no guarantee than the rays associated with the initial guess of take-off angles and one-way traveltimes will reach the shot and receiver positions.

Through different applications of stereotomography (Billette et al., 2003; Lambaré et al., 2004; Nag et al., 2006; Alerini et al., 2007; Prieux et al., 2013), some practical workflows have been designed to mitigate the non-linearity of the inversion. These strategies rely on a suitable scaling of the different optimization parameter classes, suitable initial localization of the facets and a multi-scale approach during which the grid on which the velocity model is parametrized is progressively refined. In spite of all these efforts, stereotomography still suffers from the limitations of ray-based techniques to handle complex media (Lambaré, 2008). A second limitation of computational nature is related to the fact that the inversion algorithm relies on the explicit building from paraxial quantities of the sensitivity matrix which has a (data size) $\times$ (model size) complexity. Although this matrix is a sparse matrix and its implementation provides suitable information for a sensitivity analysis of the multi-parameter inversion, it can represent a considerable computational burden when huge 3D datasets are tackled. A third limitation is related to the significant number of optimization parameters of different nature in ray-based stereotomography (e.g. velocity and ray parameters) that can make the inversion poorly conditioned and hence difficult to scale.

In order to overcome these difficulties, we present a new formulation of stereotomography. The first key ingredient is the computation of traveltimes by solving the factored eikonal solver with the fast sweeping method (FSM) (Zhao, 2005; Fomel et al., 2009). Our main motivation behind the use of eikonal solver at the expense of ray tracing is to avoid the issue of non uniform ray sampling that can arise in presence of low velocity zone or area of rapidly-varying wavespeeds. Moreover, a factored form of the eikonal equation improves the accuracy of the traveltime computation nearby the sources. Exploiting source-receiver reciprocity, we compute the synthetic slopes at the source and receiver positions by finite differences from traveltime maps computed from neighbour source and receiver positions. By doing so, we limit the computational burden of our approach by solving the eikonal equations from the source and receiver positions rather than from the scatterers. The second key ingredient is the computation of the gradient of the stereotomography misfit function with the matrix-free adjoint-state method. Plessix (2006) developed the gradient of the stereotomography misfit function from the ray equations and showed how solving the adjoint-state system is less computationally intensive than solving the propagator system required to compute the sensitivity matrix with paraxial ray equations. In our formulation, the state equations resulting from the use of a finite-difference eikonal solver as forward modelling engine (eikonal equation and slope estimation from finite differences of traveltime maps) drive us toward a model space formed by the subsurface velocities and the coordinates of the scatterers. This model space may be easier to manage than those used in ray-based stereotomography (Billette and Lambaré, 1998) because it involves a smaller number of parameter classes. Therefore, this choice may mitigate issues of parameter cross-talk and of ill-posed inversion. In this study, we support the legitimacy of this statement with an application to the complex Marmousi model. Other applications of classic stereotomography on the Marmousi model with different model parametrizations are presented by Chauris et al.

(2002b) and Billette et al. (2003).

In the following, we first review the principles of the classical ray-based stereotomography. Then we introduce the forward-problem (state) equations that are used to compute traveltimes and slopes in our formulation referred to as adjoint slope tomography. From these equations, we introduce the data and model spaces, and compute the gradient with a formulation of a semi-discretized adjoint-state method. We then discuss some practical aspects of the implementation of the method before showing its relevance with three synthetic examples. The third example shows an application to the Marmousi model performed with a realistic synthetic dataset inferred from a sparse picking of reflectors in the Marmousi model. The accuracy of the estimated model is assessed as a starting model for frequency-domain FWI with a realistic starting frequency of 4Hz. We manage to build a reliable initial velocity model for FWI without taking advantage of any prior knowledge of the velocity structure. However, our results also show the sensitivity of the method to sparse distribution of picks in deep complex zone or near the ends of the acquisition layout.

3.2.3 2D Classical stereotomography

Slope tomography methods allow for the estimation of subsurface velocity models from slopes and traveltimes of local coherent events on unmigrated data (Fig. 3.4). These local coherent events correspond most of the time to a reflection from a small reflector segment (Billette et al., 2003). However, they can also represent diffractions or diving waves (Plessix, 2006; Prieux et al., 2013). Since the structures that generate local coherent events are parametrized by point diffractor in slope tomography, we will refer to them as scatterers in the following. For two dimensional acquisitions, a stereotomographic dataset consists of N locally-coherent events that can be parametrized by

$$\mathbf{d} = [(\mathbf{s}, \mathbf{r}, T_{s,r}, p_s, p_r)_n]_{n=1}^{N}, \tag{3.55}$$

where $\mathbf{s}$ and $\mathbf{r}$ denote the position of the source and receiver, p_s and p_r are the slopes picked in the common-receiver and common-shot gathers for source s and receiver r, respectively (equivalently, the horizontal component of slowness vector at the position of source s and receiver r), and $T_{s,r}$ is the two-way traveltime (Fig. 3.5).

The scatterer related to each pair of local events is located at a position denoted by $\mathbf{x}$ from which a pair of rays are propagated in a background velocity model toward the source and receiver. The take-off angles of these two rays define the dip component of the scatterer that is mapped by the source-receiver pair. The scatterer position and the take-off angles can be represented by dip bars as shown by Billette et al. (2003, their figs 4 and 9) the assemblage of which can be viewed as a skeleton of a migrated image. This ray-based description leads to the following definition of the stereotomographic parameters

$$\mathbf{m} = \{(\mathbf{x}, \theta_s, \theta_r, T_s, T_r)_n|_{n=1}^{N}, [c_m]_{m=1}^{M}\}, \tag{3.56}$$

where fields T_s and T_r stand for one-way traveltimes of the rays shot with take-off angles θ_s and θ_r from the scatterer toward the source and receiver (Fig. 3.5). Quantities T_s, T_r, θ_s and θ_r are

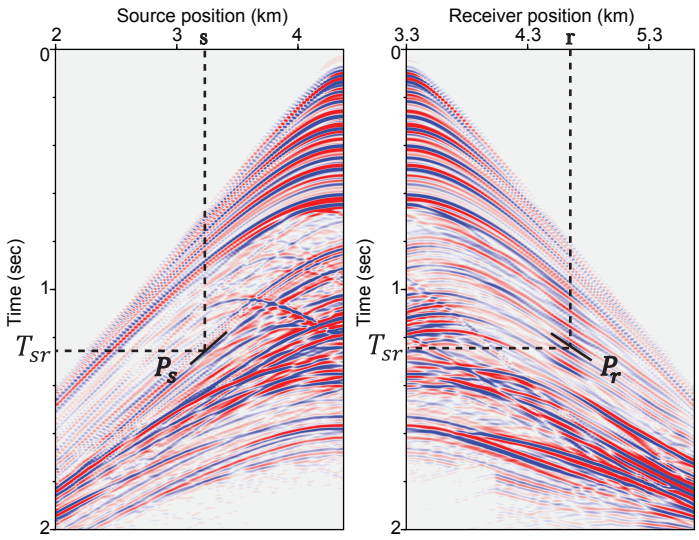

Figure 3.4 – Local coherent events in stereotomography. Each local coherent event is described by the source and receiver positions $(\mathbf{s}, \mathbf{r})$, the slopes (p_s, p_r) picked in common-shot and common-receiver gathers and two-way traveltime T_{sr}.

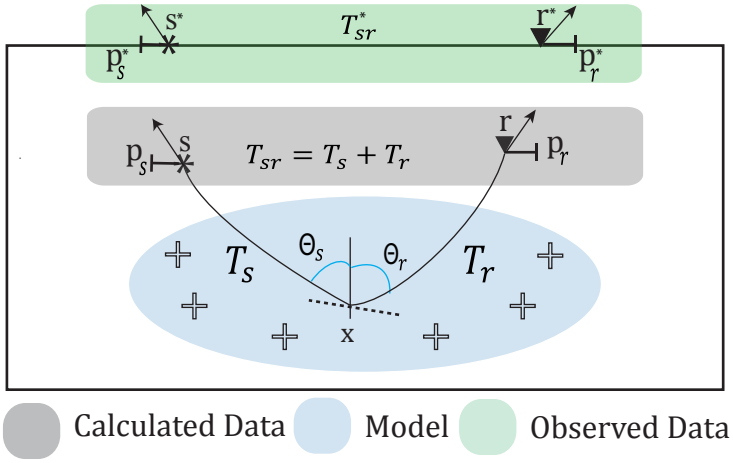

Figure 3.5 – Data space and model space in classical stereotomography. The symbols + denote the velocity nodes. Two rays are shot toward the source "s" and receiver "r" from the scatterer $\mathbf{x}$ with take-off angles Θ_s and Θ_r. Corresponding one-way traveltimes are T_s and T_r. The velocities, the scatterer coordinates $\mathbf{x}$ and the ray attributes (Θ_s, Θ_r, T_s and T_r) form the model space of classical stereotomography. The scatterer position and the two take-off angles (Θ_s, Θ_r) define a dip bar (or migration facet) shown by the segment running through the scatterer. The horizontal component of the slowness vectors at the source and receiver positions, p_s and p_r, the two-way traveltime T_{sr} and the source and receiver positions form the data space of the classical stereotomography. The observed data are designated with the symbol " $*$ " (figure adapted from Billette and Lambaré (1998)).

introduced as optimization parameters primarily for sake of efficient ray tracing with end points near the actual source and receiver locations. The fact that the end points do not match the true shot and receiver positions requires involving these later as observables in the data space (eq. 3.55). However, these parametrizations for the data space and model space provide the most general definition for the inverse problem where different ray tracing methods can use a subset of these parametrization as the data and model space. For example as an alternative, two-point ray tracing can be performed at the expense of computational efficiency. This option allows one to remove T_s, T_r, θ_s and θ_r from the model space and $\mathbf{s}$ and $\mathbf{r}$ from the data space. The velocity macro-model can be parametrized by the coefficients c_m of cardinal cubic B-spline functions, regardless of the ray tracing method, which ensure the second-order continuity of velocity model required by paraxial estimations (Billette et al., 2003).

A subsurface model is iteratively updated by minimization of the least-squares misfit between picked and modelled data. The inverse problem is solved through the explicit building of the sensitivity matrix, which can be inferred from paraxial ray quantities (Billette and Lambaré, 1998). The resulting tomographic system augmented with smoothing constraints is solved with a linear conjugate gradient such as LSQR (Paige and Saunders, 1982). In practice, different strategies have been implemented in stereotomography to cope with potential non-linearities and ill-posedness resulting from the multi-parameter (i.e. velocity, ray parameters and scatterer positions) nature of the reconstruction. Billette (1998) proposed a workflow subdivided in three steps: initialization of ray parameters in a homogeneous background velocity model, estimation of a constant-gradient velocity model that best fits the stereotomographic observed data, and joint inversion of the ray parameters, scatterer positions and spline coefficients modifying the background constant-gradient velocity model. This procedure faced difficulties in case of complex geological structures where the second step cannot converge to a sufficiently-accurate velocity model (Billette et al., 2003). Time stripping (Alerini et al., 2007) is another technique to incorporate the stereotomographic picks in a progressive manner, according to their traveltimes. For very smooth and layering velocity models, this approach seems efficient but, since the position of the scatterers and the one-way traveltimes are updated independently, this method may fail to converge in presence of strong lateral velocity variations. Target-oriented strategies are also possible to reduce the non-linearity of the inversion (Billette, 1998).

To establish a powerful, automatic and comprehensive approach, Billette et al. (2003) introduced a multi-scale approach. They proposed an initialization of the model parameters by considering a sparse distribution of cubic cardinal B-spline nodes in each direction and a simple geometrical consideration for the ray parameters in a homogeneous background velocity model. The next step is the localization, that includes updating independently the ray parameters and positions of scatterers while the velocity is fixed at its initial value. In the third step, they perform a joint inversion of all model parameters through a multi-scale approach where the grid of spline nodes is progressively refined over the different multi-scale steps. In this study, we design a workflow along these lines, with different model parametrization and inversion implementation.

3.2.4 Adjoint slope tomography

Compared to classical stereotomography, our implementation of slope tomography relies, on the one hand, on solving the forward problem by numerical resolution of the eikonal equation and, on the other hand, on using the adjoint-state method to compute the gradient of the slope tomography misfit function.

Although eikonal solvers do not provide explicit information on rays attributes such as slopes, they avoid the issue of non-uniform ray sampling caused by low or rapidly varying velocities in particular near grazing angles. It is worth reminding that the aim of the slope tomography is to build smooth velocity models. Therefore, our motivation behind using eikonal solvers is not to use diffractions or head waves generated by low or high velocity contrasts, these arrivals being challenging to pick in the data, but rather to take advantage of robust and accurate computation of uniformly-sampled traveltime maps in smooth media.

Current implementations of eikonal solvers do not identify multiple arrivals generated by multi-pathing which remains a computational challenge (Qian and Leung, 2004). Therefore, we consider only the fastest solution in the present formulation. Operto et al. (2000, their figs 10 and 11) showed that ray+Born migration/inversion using first-arrival traveltimes provide acceptable quantitative migrated images of the Marmousi model as long as the background model is sufficiently smooth. In this case, the amplitudes of the velocity perturbations are reasonably estimated where true velocity structure is complex and generates ray folding. This suggests that the information carried out by multiple arrivals in the data is accounted for in an average sense. Indeed, this amplitude recovery from first-traveltime arrivals is achieved at the expense of the quality of the focusing, that is hampered by the smoothness of the background model (compare figures 7 and 9 in Operto et al. (2000)). We can transpose this reasoning from the migration task to the velocity model building counterpart by assuming that the velocity models estimated by slope tomography are smooth enough such that first-arrival eikonal solver predicts in average sense traveltimes of multiple arrivals that would have been picked in the data. In summary, our choice of eikonal solver rather than ray tracing is pragmatical. It provides a robust tool to compute uniformly-sampled traveltime maps that accurately predict picked traveltimes as long as the velocity model is forced to be smooth enough.

For solving the inverse problem, we compute explicitly the gradient of the misfit function with respect to the model parameters using an adjoint-state formulation. This allows us to avoid building the sensitivity matrix and performing large-scale matrix resolution. The matrix-free nature of this approach makes the inversion scheme more capable of handling large data sets. In the following section, we review the model and data spaces used in our forward modelling engine and the adjoint formulation of slope tomography. All the mathematical symbols used in our formulation are presented in Table 3.1.

Data and model space definition

In adjoint slope tomography, we consider descending wave propagation from the source and receiver positions toward the scatterers (Fig. 3.6), unlike ray-based approaches where rays are shot from the scatterer toward the source and receiver positions with prescribed shooting

Table 3.1 – Mathematical symbols.

Symbol	Description
N	Total number of scatterers.
M	Total number of cubic B-spline nodes.
$N_{\mathbf{s}}$	Number of distinct shot gathers.
$N_{\mathbf{r}}$	Number of distinct receiver gathers.
$N_{\mathbf{r}}^{s}$	Number of distinct receivers in the shot-gather s.
$N_{\mathbf{s}}^{r}$	Number of distinct sources in the receiver-gather r.
$N_{\mathbf{n}}^{s,r}$	Number of scatterers for source-receiver pair (s, r).
$n_{s,r}$	nth scatterer associated with source and receiver pair (s, r).
$\mathbf{x}_{n_{s,r}}$	Position of nth scatterer associated with source-receiver pair (s, r) .
$\mathbf{x}_s$, $\mathbf{x}_r$	Position of source s and receiver r .
$T^*_{s,r,n_{s,r}}$	Observed two-way traveltime for scatterer $n_{s,r}$.
$T_{s,r,n_{s,r}}$	Calculated two-way traveltime for scatterer $n_{s,r}$.
$p^*_{s,n_{s,r}}$	Observed local slope for scatterer $n_{s,r}$ at the position of sth source of receiver-gather r.
$p_{s,n_{s,r}}$	Calculated local slope for scatterer $n_{s,r}$ at the position of sth source of receiver-gather r.
$p^*_{r,n_{s,r}}$	Observed local slope for scatterer $n_{s,r}$ at the position of rth receiver of shot-gather s.
$p_{r,n_{s,r}}$	Calculated local slope for scatterer $n_{s,r}$ at the position of rth receiver of shot-gather s.
$(\dots)^t$	Transpose operator.
$\mathbf{v}$	Velocity model on the Cartesian grid.
$\mathbf{c}$	B-spline velocity coefficients: $\mathbf{v} = \mathbf{B}\mathbf{c}$
$\mathbf{m} = \{\mathbf{c}, \mathbf{x}_{n_{s,r}}\}$	Adjoint slope tomography model parameters.
$\mathbf{d}$	Adjoint slope tomography observables.
$C(\mathbf{m})$	Adjoint slope tomography misfit function

angles (Billette, 1998). These shooting angles explicitly define the dip of the scatterer (i.e., reflector segment if the scatterer lies on a reflector) sampled by the source-receiver pair. In our approach, we shall show that the dip component of a scatterer is implicitly embedded in the adjoint sources. This implicit representation results from the way we estimate the slopes at the source and receiver positions by finite differences.

Since one eikonal resolution provides a traveltime map in the whole target and each shot and receiver gather can be associated with several picked coherent events, descending propagations from the surface to the scatterers mitigates the number of forward modellings as the number of shots and receivers is expected to be one to two order of magnitude smaller than the number of scatterers. The local-coherent events, picked on common-shot and common-receiver gathers, are parametrized by their two-way traveltimes and the two slopes (in 2D) at the source and receiver positions. These quantities define the data space of the adjoint slope tomography. To each picked event in the data space is associated a scatterer in the subsurface. Accordingly, we label each scatterer with the subscripts of the source and receiver to which it is related. This prompts the following notations for the data space of the adjoint slope tomography:

$$\mathbf{d} = \left(T_{s,r,n_{s,r}}, p_{s,n_{s,r}}, p_{r,n_{s,r}}\right) \big|_{s=1}^{N_{\mathbf{s}}} \big|_{r=1}^{N_{\mathbf{r}}^{\mathbf{s}}} \big|_{n_{s,r}=1}^{N_{\mathbf{n}}^{s,r}}. \tag{3.57}$$

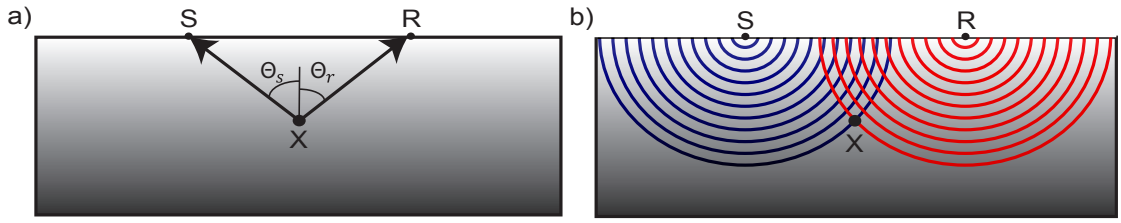

Figure 3.6 – Layout of forward modelling in a) classic stereotomography: two rays are back propagated from scatterer toward the associated source and receiver, and in b) adjoint slope tomography: two traveltime maps are generated from associated source and receiver toward the scatterer.

where s and r denote a source and receiver, $n_{s,r}$ a scatterer associated with the pair (s,r), $T_{s,r,n_{s,r}}$ the two-way traveltime along the paths connecting the source s and the receiver r to the scatterer $n_{s,r}$, $p_{s,n_{s,r}}$ and $p_{r,n_{s,r}}$ the horizontal component of slowness vector at the source and receiver position, respectively. Unlike in classical stereotomography, the source and receiver positions are not part of the data space since we perform wave propagation from these exact positions. In classical stereotomography, one needs to minimize distances between source or receiver position and the end points of the rays emitted with the prescribed shooting angle from the scatterer: some specificities of the ray-tracing engine has led to a particular extension of both data space (source/receiver positions) and model space (shooting angles and one-way traveltimes).

Consequently, the model space involves two categories of parameter classes: coordinates of the scatterers $\mathbf{x}_{n_{s,r}} = (x_{n_{s,r}}, z_{n_{s,r}})$ and subsurface velocities.

$$\mathbf{m} = \left([c_m]_{m=1}^{M} , \{ \mathbf{x}_{n_{s,r}} |_{s=1}^{N_\mathbf{s}} |_{r=1}^{N_\mathbf{r}^s} |_{n_{s,r}=1}^{N_\mathbf{n}^{s,r}} \} \right)^t . \tag{3.58}$$

Here, c_m denote the coefficients of the cubic cardinal B-splines, that are used to parametrize the subsurface velocity model (Billette, 1998). This sampling of the velocity model operates as an implicit regularization of the stereotomography. While the cubic cardinal B-spline parametrization is generally a natural choice for ordinary differential equations such as (paraxial) ray tracing, eikonal solvers as partial differential equations are often formulated for efficiency on a Cartesian grid. Nevertheless, we consider that B-splines will describe in a compact way the smooth velocity structure we are looking for. We may consider this sampling choice as a hard constraint on the model space. We shall also promote a soft constraint through the Gaussian filtering of the misfit gradient.

Forward problem

According to the data space defined in the eq. (3.57), the forward-modelling engine requires calculation of the slopes at the source and receiver and the two-way traveltimes for each triplet $(s,r,n_{s,r})$. We first compute the traveltime maps $t_s(\mathbf{x})$ and $t_r(\mathbf{x})$ initiated from each source

and each receiver, respectively, in the velocity model using zero traveltimes at the source and receiver positions as Dirichlet boundary conditions in the eikonal equation

$$H(\mathbf{x}, \nabla t_s(\mathbf{x})) = |\nabla t_s(\mathbf{x})|^2 - \frac{1}{v^2(\mathbf{x})} = 0, \tag{3.59}$$

$$t_s(\mathbf{x}_s) = 0 \tag{3.60}$$

and

$$H(\mathbf{x}, \nabla t_r(\mathbf{x})) = |\nabla t_r(\mathbf{x})|^2 - \frac{1}{v^2(\mathbf{x})} = 0, \tag{3.61}$$

$$t_r(\mathbf{x}_r) = 0 \tag{3.62}$$

where the velocity model $v(\mathbf{x})$ is parametrized on the Cartesian grid that is used for building the discrete field solution.

The two-way traveltimes $T_{s,r,n_{s,r}}$ for the scatterer $n_{s,r}$ are given by

$$T_{s,r,n_{s,r}} = Q_{n_{s,r}}\mathbf{t}_s + Q_{n_{s,r}}\mathbf{t}_r, \tag{3.63}$$

where $Q_{n_{s,r}}$ is a sampling operator, which extracts the traveltime at the scatterers position, $\mathbf{x}_{n_{s,r}}$. In this study, we implement this sampling operator with a windowed sinc function (Hicks, 2002) although other options can be considered.

We compute the traveltime maps with a factored eikonal solver based on FSM (Fomel et al., 2009). Although the local finite-difference stencil considers only outward directions when computing traveltimes for external nodes of the grid, the sweeping scheme guarantees the coverage of all the wave propagation directions. Moreover, the factorization technique allows one to remove the singularity at the point source position. The extension of this method to TTI anisotropic media as proposed by Waheed et al. (2015) and Tavakoli F. et al. (2015) may provide the necessary forward modelling engine for adjoint slope tomography in anisotropic media.

As previously mentioned, two slopes that are picked in a shot and receiver gathers and tied to a scatterer in the subsurface correspond to the horizontal component of the slowness vectors at the receiver and source positions. Exploiting the reciprocity of the wave propagation between source and scatterer and between receiver and scatterer, and assuming a dense sampling of shots and receivers, we estimate in a finite-difference sense these two slopes from the traveltime maps generated by the left and right neighbour receivers/sources (Fig. 3.7). This leads to the following expression of the slopes

$$p_{s,n_{s,r}} = \frac{\partial t_s(\mathbf{x}_s)}{\partial x_s} \approx (Q_{n_{s,r}}\mathbf{t}_{s+1} - Q_{n_{s,r}}\mathbf{t}_{s-1})/2\Delta s, \tag{3.64}$$

$$p_{r,n_{s,r}} = \frac{\partial t_r(\mathbf{x}_r)}{\partial x_r} \approx (Q_{n_{s,r}}\mathbf{t}_{r+1} - Q_{n_{s,r}}\mathbf{t}_{r-1})/2\Delta r, \tag{3.65}$$

where Δs and Δr denote the source and receiver intervals, respectively.

With this reciprocity-based strategy, slopes at the source and receiver positions are inferred

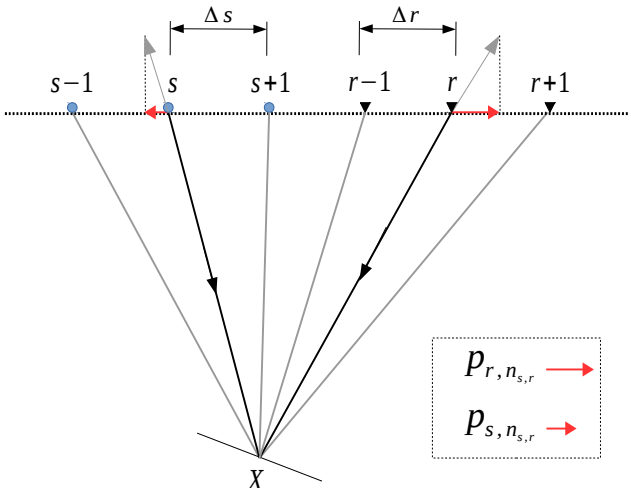

Figure 3.7 – Horizontal component of slowness vectors at the source/receiver positions, $(p_{s,n_{s,r}}, p_{r,n_{s,r}})$, are inferred from the traveltime fields emitted from neighbour sources/receivers. Δs and Δr are the source and receivers interval, respectively.

from traveltimes sampled at the scatterer positions, namely far away from the sources or the receivers where traveltimes have singular values. Therefore, we expect these traveltimes to be accurate enough for a reliable finite-difference estimation of the slopes at the sources or at the receivers, thanks to the smoothness of the velocity model. Alternatively, slopes can be estimated through a partial differential equation by solving the angle equation $\nabla\theta.\nabla\mathbf{t} = \mathbf{0}$ in isotropic media. This means that the gradient of the take-off angle θ is constant along the ray (Noble et al., 2012; Belayouni, 2013) which can be recast into the more general expression $\partial H/\partial\mathbf{p}\cdot d\mathbf{p}$ for anisotropic media with point source condition. We prefer the semi-discretized formulation because, in this case, synthetic fields to be found are only the two traveltimes maps. We are confident that the finite-difference estimation of slopes is accurate enough in smooth media we are looking for, thanks to the use of the reciprocity.

Eqs 3.59-3.65 define state equations that will be used to estimate the gradient of the adjoint slope tomography misfit function. With our formulation, the number of forward problems scales to the number of sources and receivers.

Multi-parameter inverse problem

From the definition of the data space (eq. 3.57), the misfit function of the adjoint slope tomography is given by

$$
\begin{aligned}
C(\mathbf{m}) &= \frac{1}{2\sigma^2_{T_{s,r}}} \sum_{s=1}^{N_\mathbf{s}} \sum_{r=1}^{N^s_\mathbf{r}} \sum_{n_{s,r}=1}^{N^{s,r}_\mathbf{n}} \left(T_{s,r,n_{s,r}}(\mathbf{m}) - T^*_{s,r,n_{s,r}}\right)^2 \\
&+ \frac{1}{2\sigma^2_{p_s}} \sum_{s=1}^{N_\mathbf{s}} \sum_{r=1}^{N^s_\mathbf{r}} \sum_{n_{s,r}=1}^{N^{s,r}_\mathbf{n}} \left(p_{s,n_{s,r}}(\mathbf{m}) - p^*_{s,n_{s,r}}\right)^2 \\
&+ \frac{1}{2\sigma^2_{p_r}} \sum_{r=1}^{N_\mathbf{r}} \sum_{s=1}^{N^r_\mathbf{s}} \sum_{n_{s,r}=1}^{N^{s,r}_\mathbf{n}} \left(p_{r,n_{s,r}}(\mathbf{m}) - p^*_{r,n_{s,r}}\right)^2,
\end{aligned} \tag{3.66}
$$

where the symbol * denotes the picked data and quantities $\sigma^2_{T_{s,r}}$, $\sigma^2_{p_s}$, and $\sigma^2_{p_r}$, are elements of a diagonal covariance matrix containing the variance of each data class. The inverse of this matrix plays the role of a weighting operator which balances the relative contribution of each residual class during the inversion (Tarantola, 1987).

Minimizing the misfit function $C(\mathbf{m})$, eq. 3.66, is a non-linear problem which can be tackled either with global or local optimization techniques. Although global optimization can be used for 2D tomographic problems (Datta and Sen, 2016; Sajeva et al., 2016), we focus this study on Newton-based local optimization scheme that updates iteratively the model parameters $\mathbf{m}$ as

$$
\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \left(\frac{\partial^2 C}{\partial \mathbf{m}^2}(\mathbf{m}_k) \right)^{-1} \frac{\partial C}{\partial \mathbf{m}}(\mathbf{m}_k). \tag{3.67}
$$

Here the positive real step length α_k is estimated by line search while satisfying the Wolfe conditions (Nocedal and Wright, 2006). We use the L-BFGS method (Byrd et al., 1995) implemented in the SEISCOPE optimization toolbox (Métivier and Brossier, 2016) to account for the Hessian operator in the inversion. Considering the contribution of the Hessian operator is crucial to manage the potential leakage between the two different parameter classes (velocity and positions of scatterers). Because this is often not enough, a common practice makes the different parameter classes dimensionless in multi-parameter inversion to better balance the relative contribution of each parameter class. This will improve the condition number of the inversion. In our case, with the units for velocity (meter/second) and for scattered positions (meter), we have not found the need to do so.

Gradient computation with the adjoint-state method

We compute the gradient of the misfit function (eq. 3.66) with the adjoint-state method implemented through the Lagrangian formalism (Chavent, 1974; Plessix, 2006).

We build the Lagrangian misfit function by augmenting the misfit function C with equality constraints requiring that the state variables are realizations of the state equations (eqs 3.59-

3.65):

$$
\begin{aligned}
\mathcal{L}(\mathbf{m},\mathbf{u},\bar{\mathbf{u}}) &= \mathcal{H}(\mathbf{u},\mathbf{m}) \\
&- \sum_{s=1}^{N_\mathbf{s}} \sum_{r=1}^{N_\mathbf{r}^s} \sum_{n_{s,r}=1}^{N_\mathbf{n}^{s,r}} \mu_{s,r,n_{s,r}} \left(T_{s,r,n_{s,r}} - Q_{n_{s,r}} \mathbf{t}_s - Q_{n_{s,r}} \mathbf{t}_r \right) \\
&- \sum_{s=1}^{N_\mathbf{s}} \sum_{r=1}^{N_\mathbf{r}^s} \sum_{n_{s,r}=1}^{N_\mathbf{n}^{s,r}} \xi_{s,n_{s,r}} \left(p_{s,n_{s,r}} - (Q_{n_{s,r}} \mathbf{t}_{s+1} - Q_{n_{s,r}} \mathbf{t}_{s-1})/2\Delta s \right) \\
&- \sum_{r=1}^{N_\mathbf{r}} \sum_{s=1}^{N_\mathbf{s}^r} \sum_{n_{s,r}=1}^{N_\mathbf{n}^{s,r}} \xi_{r,n_{s,r}} \left(p_{r,n_{s,r}} - (Q_{n_{s,r}} \mathbf{t}_{r+1} - Q_{n_{s,r}} \mathbf{t}_{r-1})/2\Delta r \right) \\
&- \frac{1}{2} \sum_{s=1}^{N_\mathbf{s}} \Big\langle \lambda_s(\mathbf{x}) \mid H(\mathbf{x}, \nabla t_s(\mathbf{x}) \Big\rangle_\Omega - \frac{1}{2} \sum_{r=1}^{N_\mathbf{r}} \Big\langle \lambda_r(\mathbf{x}) \mid H(\mathbf{x}, \nabla t_r(\mathbf{x}) \Big\rangle_\Omega \\
&- \sum_{s=1}^{N_\mathbf{s}} \psi_s \, t_s(\mathbf{x}_s) - \sum_{r=1}^{N_\mathbf{s}} \psi_r \, t_r(\mathbf{x}_r), \qquad (3.68)
\end{aligned}
$$

where scalar products, $\langle \cdot, \cdot \rangle$, are defined on the targeted subsurface domain Ω. The arguments of the Lagrangian $\mathbf{u} = (T_{s,r,n_{s,r}}, p_{s,n_{s,r}}, p_{r,n_{s,r}}, \mathbf{t}_s, \mathbf{t}_r, t_s(\mathbf{x}_s), t_r(\mathbf{x}_r))$ and $\bar{\mathbf{u}} = (\mu_{s,r,n_{s,r}}, \xi_{s,n_{s,r}}, \xi_{r,n_{s,r}}, \lambda_s, \lambda_r, \psi_s, \psi_r)$ gather the state and adjoint-state variables, respectively. Since the state variables, the adjoint-state variables and the model parameters are processed as independent quantities in the eq. (3.68), we rewrite $C(\mathbf{m})$ as $\mathcal{H}(\mathbf{u},\mathbf{m})$. Finally, we substitute in the state equation satisfied by $p_{s,n_{s,r}}$ and $p_{r,n_{s,r}}$ the formal expression of the slopes, namely $\partial \mathbf{t}_s / \partial x_s$ and $\partial \mathbf{t}_r / \partial x_r$, by their finite-difference approximation as provided in eqs (3.64) and (3.65), eliminating the need of considering slope maps in the derivations.

The adjoint-state equations are obtained by zeroing the partial derivative of the Lagrangian with respect to the state variables. It is straightforward to derive the expression of the adjoint-state variables $\mu_{s,r,n_{s,r}}$, $\xi_{s,n_{s,r}}$ and $\xi_{r,n_{s,r}}$, which correspond to the traveltime and slope residuals weighted by the coefficients of the covariance matrices.

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial T_{s,r,n_{s,r}}} = 0 \quad &\rightarrow \quad \mu_{s,r,n_{s,r}} = \frac{1}{\sigma^2_{T_{s,r}}} (T_{s,r,n_{s,r}} - T^*_{s,r,n_{s,r}}) = \frac{\Delta T_{s,r,n_{s,r}}}{\sigma^2_{T_{s,r}}}, \\
\frac{\partial \mathcal{L}}{\partial p_{s,n_{s,r}}} = 0 \quad &\rightarrow \quad \xi_{s,n_{s,r}} = \frac{1}{\sigma^2_{p_s}} (p_{s,n_{s,r}} - p^*_{s,n_{s,r}}) = \frac{\Delta p_{s,n_{s,r}}}{\sigma^2_{p_s}}, \\
\frac{\partial \mathcal{L}}{\partial p_{r,n_{s,r}}} = 0 \quad &\rightarrow \quad \xi_{r,n_{s,r}} = \frac{1}{\sigma^2_{p_r}} (p_{r,n_{s,r}} - p^*_{r,n_{s,r}}) = \frac{\Delta p_{r,n_{s,r}}}{\sigma^2_{p_r}}.
\end{aligned}
$$

Zeroing the partial derivative of the Lagrangian with respect to $t_s(\mathbf{x}_s)$ and $t_r(\mathbf{x}_r)$ indicates the adjoint-state variables ψ_s and ψ_r are zero, therefore, the zero traveltime boundary conditions on the sources and receivers do not insert any information to our formalism.

To develop the adjoint-state equations satisfied by $\lambda_s(\mathbf{x})$ (namely, $\partial \mathcal{L} / \partial \mathbf{t}_s = 0$), we have to identify all of the terms containing $\mathbf{t}_s$ in the summation over sources of the state equations satisfied by $p_{s,n_{s,r}}$ (third line in eq. (3.68)). We find two contributions coming from the neighbour

source $s+1$ and $s-1$. We finally obtain for $\partial\mathcal{L}/\partial\mathbf{t}_s$

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\mathbf{t}_s} &= \sum_{r=1}^{N_{\mathbf{r}}^s}\sum_{n_{s,r}=1}^{N_{\mathbf{n}}^{s,r}} Q^t_{n_{s,r}}\mu_{s,r,n_{s,r}} \\ &+ \frac{1}{2\Delta s}\sum_{r=1}^{N_{\mathbf{r}}^{s-1}}\sum_{n_{s-1,r}=1}^{N_{\mathbf{n}}^{s-1,r}} Q^t_{n_{s-1,r}}\xi_{s-1,n_{s-1,r}} \\ &- \frac{1}{2\Delta s}\sum_{r=1}^{N_{\mathbf{r}}^{s+1}}\sum_{n_{s+1,r}=1}^{N_{\mathbf{n}}^{s+1,r}} Q^t_{n_{s+1,r}}\xi_{s+1,n_{s+1,r}} \\ &+ \sum_{s=1}^{N_{\mathbf{s}}}\Big(\Big(\nabla\cdot(\lambda_s(\mathbf{x})\nabla t_s(\mathbf{x}))\Big)_\Omega - \Big(\lambda_s(\mathbf{x})\nabla t_s(\mathbf{x})\cdot\mathbf{n}\Big)_\Gamma\Big),\end{aligned} \tag{3.69}$$

where the last line has been obtained by integration by parts (Taillandier et al., 2009). Here, Γ denotes the boundaries of Ω and $\mathbf{n}$ denotes the outward unit normal vectors to them. Without loss of generality, we shall impose Dirichlet boundary conditions for the adjoint variables, so that $\lambda_s(\mathbf{x})\nabla t_s(\mathbf{x})\cdot\mathbf{n}=0$ over Γ: this will be a non-restrictive hypothesis as long as all sources and receivers are inside the numerical domain (Taillandier et al., 2009). In a similar way we calculate $\partial\mathcal{L}/\partial\mathbf{t}_r$.

Zeroing eq. (3.69) gives the adjoint-state equation satisfied by $\lambda_s(\mathbf{x})$

$$\begin{aligned}\Big(\nabla\cdot(\lambda_s(\mathbf{x})\nabla t_s(\mathbf{x}))\Big)_\Omega &= -\sum_{r=1}^{N_{\mathbf{r}}^s}\sum_{n_{s,r}=1}^{N_{\mathbf{n}}^{s,r}} Q^t_{n_{s,r}}\mu_{s,r,n_{s,r}} \\ &+ \frac{1}{2\Delta s}\sum_{r=1}^{N_{\mathbf{r}}^{s+1}}\sum_{n_{s+1,r}=1}^{N_{\mathbf{n}}^{s+1,r}} Q^t_{n_{s+1,r}}\xi_{s+1,n_{s+1,r}} \\ &- \frac{1}{2\Delta s}\sum_{r=1}^{N_{\mathbf{r}}^{s-1}}\sum_{n_{s-1,r}=1}^{N_{\mathbf{n}}^{s-1,r}} Q^t_{n_{s-1,r}}\xi_{s-1,n_{s-1,r}}.\end{aligned} \tag{3.70}$$

In a similar way, the adjoint-state variables $\lambda_r(\mathbf{x})$ satisfy

$$\begin{aligned}\Big(\nabla\cdot(\lambda_r(\mathbf{x})\nabla t_r(\mathbf{x}))\Big)_\Omega &= -\sum_{s=1}^{N_{\mathbf{s}}^r}\sum_{n_{s,r}=1}^{N_{\mathbf{n}}^{s,r}} Q^t_{n_{s,r}}\mu_{s,r,n_{s,r}} \\ &+ \frac{1}{2\Delta r}\sum_{s=1}^{N_{\mathbf{s}}^{r+1}}\sum_{n_{s,r+1}=1}^{N_{\mathbf{n}}^{s,r+1}} Q^t_{n_{s,r+1}}\xi_{r+1,n_{s,r+1}} \\ &- \frac{1}{2\Delta r}\sum_{s=1}^{N_{\mathbf{s}}^{r-1}}\sum_{n_{s,r-1}=1}^{N_{\mathbf{n}}^{s,r-1}} Q^t_{n_{s,r-1}}\xi_{r-1,n_{s,r-1}}.\end{aligned} \tag{3.71}$$

We use the same approach as Taillandier et al. (2009) based on the FSM method (Zhao, 2005) to solve the linear eqs (3.70) and (3.71) for $\lambda_s(\mathbf{x})$ and $\lambda_r(\mathbf{x})$. However, we implement the traveltime and slope residuals as source terms at the scatterer positions in the right-hand sides of the adjoint equations rather than as a boundary condition in Taillandier et al. (2009, their equation 12). This results because we assume that the shots and receivers are inside the domain Ω. These adjoint sources back-propagate the traveltime and slopes residuals along two ray tubes connecting the scatterers to the shots and receivers. Traveltime residuals $\Delta T_{s,r,n_{s,r}}$ are back-propagated from the scatterer position $n_{s,r}$ toward the shot s and receiver r, as it would be in reflection traveltime tomography. Together with these traveltime residuals, the slope residuals, that are tied to the neighbour sources $s+1, s-1$ (eq. 3.70), and receivers $r+1, r-1$ (eq. 3.71), are back-propagated from the scatterers $n_{s-1,r}$, $n_{s+1,r}$, $n_{s,r-1}$ and $n_{s,r+1}$ toward the shot s and receiver r. Therefore, the spatial support spanned by these neighbour scatterers constrain the dip angle at the scatterer $n_{s,r}$ and, hence can be viewed as a discrete approximation of the two take-off angles that are conventionally estimated in ray-based stereotomography. In the example section, we illustrate the shape of the adjoint-state variables $\lambda_s(\mathbf{x})$ and $\lambda_r(\mathbf{x})$ and the role of residuals in scatterers repositioning. The contribution of all of the scatterers associated with one source or receiver and its neighbours are gathered in the right-hand side by summation. Therefore, the number of adjoint-state problems to be solved is equal to the number of sources and receivers and hence is independent of the number of scatterers, an important property which makes the computational cost independent of the number of picked events.

The gradient of the misfit function (3.66) with respect to the model parameters is obtained by taking the partial derivative of the Lagrangian (3.68) with respect to the model parameters.

The gradient of the misfit function with respect to the velocities $\mathbf{v}$ on the Cartesian grid is the sum of the adjoint-state variables $\lambda_s(\mathbf{x})$ and $\lambda_r(\mathbf{x})$ weighted by the velocity raised to the power 3.

$$\nabla_{v(\mathbf{x})} C = -\sum_{s=1}^{N_\mathbf{s}} \frac{\lambda_s(\mathbf{x})}{v(\mathbf{x})^3} - \sum_{r=1}^{N_\mathbf{r}} \frac{\lambda_r(\mathbf{x})}{v(\mathbf{x})^3}. \tag{3.72}$$

The gradient with respect to the cubic B-spline coefficients $\mathbf{c} = c_m|_{m=1}^{M}$ can be inferred from the eq. (3.72) by applying the chain rule of derivatives

$$\nabla_\mathbf{c} C = \mathbf{B}^t \nabla_\mathbf{v} C, \tag{3.73}$$

where $\mathbf{B}$ stands for the cubic cardinal B-spline operator ($\mathbf{v} = \mathbf{B}\mathbf{c}$) and ^{t} denotes the adjoint operator. The gradient with respect to the scatterer coordinates is given by

$$\begin{aligned}
\nabla_{\mathbf{x}_{n_{s,r}}} C =& \mu_{s,r,n_{s,r}} \left(\frac{\partial Q_{n_{s,r}}}{\partial \mathbf{x}_{n_{s,r}}} \mathbf{t}_s + \frac{\partial Q_{n_{s,r}}}{\partial \mathbf{x}_{n_{s,r}}} \mathbf{t}_r \right) \\
&+ \frac{\xi_{s,n_{s,r}}}{2\Delta s} \left(\frac{\partial Q_{n_{s,r}}}{\partial \mathbf{x}_{n_{s,r}}} \mathbf{t}_{s+1} - \frac{\partial Q_{n_{s,r}}}{\partial \mathbf{x}_{n_{s,r}}} \mathbf{t}_{s-1} \right) \\
&+ \frac{\xi_{r,n_{s,r}}}{2\Delta r} \left(\frac{\partial Q_{n_{s,r}}}{\partial \mathbf{x}_{n_{s,r}}} \mathbf{t}_{r+1} - \frac{\partial Q_{n_{s,r}}}{\partial \mathbf{x}_{n_{s,r}}} \mathbf{t}_{r-1} \right),
\end{aligned} \tag{3.74}$$

where the terms $\partial Q_{n_{s,r}} / \partial \mathbf{x}_{n_{s,r}}$ can be unambiguously obtained through the derivative of the

windowed sinc function (Hicks, 2002). The coordinates of the scatterers are updated from the traveltime maps computed from the sources and receivers to which they are related as well as from those computed from the neighbouring sources and receivers. Again, this highlights the additional constraints provided by slopes to locate the scatterers. The different steps required to compute the gradient are outlined in Algorithm 3.

Implementation of adjoint slope tomography in practice

We subdivide the adjoint slope tomography into three main steps following the strategy proposed by Billette et al. (2003). Three steps, (1) initialization of scatterer positions, (2) preliminary localization of the scatterers and (3) multi-scale inversion for joint velocity and scatterer updates, aim to mitigate the non linearity and ill-posedness of the inversion (Algorithm 4).

Initialization of scatterer positions: An initial guess of each scatterer position can be computed analytically from the observed traveltime and slopes by assuming straight rays (Billette et al., 2003, their appendix A). At this stage, no initial velocity model needs to be defined: each picked data related to a scatterer defines an effective homogeneous velocity. As underlined by Billette et al. (2003), closed-form solutions have always been found in our numerical examples and, the computational cost is negligible.

Localization of scatterer positions: In order to remove noisy picks and make the initial positions of the scatterers as close as possible to their true positions, we update the scatterer positions keeping a fixed background velocity model. One may use either a homogeneous velocity model or a constant-gradient velocity model to compute the traveltimes

Algorithm 3: Gradient algorithm

$\mathbf{v} = \mathbf{B}\,\mathbf{c}$
for s / $r = 1$ to $N_\mathbf{s}$ / $N_\mathbf{r}$ **do**
  Compute $\mathbf{t}_s$ / $\mathbf{t}_r$, eqs. (3.59)-(3.62)
end for
for s / $r = 1$ to $N_\mathbf{s}$ / $N_\mathbf{r}$ **do**
  for r / $s = 1$ to $N_\mathbf{r}^s$ / $N_\mathbf{s}^r$ **do**
    for $n_{s,r} = 1$ to $N_\mathbf{n}^{s,r}$ **do**
      Compute $T_{s,r,ns,r}$, $p_{s,n_{s,r}}$ / $p_{r,n_{s,r}}$, eqs. (3.63)-(3.65)
    end for
  end for
end for
Compute C, eq. (3.66)
for s / $r = 1$ to $N_\mathbf{s}$ / $N_\mathbf{r}$ **do**
  Compute λ_s / λ_r, eqs. (3.70)-(3.71)
end for
Compute $\nabla_\mathbf{v} C$, eq. (3.72)
Compute $\nabla_\mathbf{c} C = \mathbf{B}^t \nabla_\mathbf{v} C$, eq. (3.73)
Compute $\nabla_\mathbf{x} C$, eq. (3.74)

and slopes analytically and hence make this re-localization step computationally efficient. This background velocity model might be chosen to make the distribution of the scatterers as even as possible in the subsurface. For example, a too slow homogeneous model might tend to squeeze the scatterers in the shallow part of the subsurface medium. In contrast, a constant-gradient velocity model can help to equally distribute the scatterers in depth. If no *prior* information coming from well logs are available, finding an appropriate constant-gradient model by trial and error or grid search is not challenging due to the computational efficiency of the localization process. Since the velocity model is not updated during the localization step, the resolution of the eikonal equation needs to be performed only during the first iteration, where we store the traveltime maps.

Multi-scaling joint inversion: The last step consists of simultaneously updating all of the model parameters through a multi-scale approach. Multi-scale imaging is performed through a progressive refinement of the spline parametrization, that is implemented taking advantage of the subdivision property of the B-spline surfaces (see de Boor (1978) or Virieux and Farra (1991, their appendix)). We complement the regularization performed by the coarse B-spline parametrization by an additional Gaussian smoothing of the gradient computed on the Cartesian grid, similar to the one used by Taillandier et al. (2009) for first-arrival traveltime tomography. This smoothing acts as a soft constraint to reduce the null space, as opposed to the sampling strategy of the model space which is a hard constraint. With this smoothing regularization, the gradient of the misfit function with respect to the B-spline coefficients can be written as

$$\nabla_{\mathbf{c}} C = \mathbf{B}^t \mathbf{G} \nabla_{\mathbf{v}} C, \tag{3.75}$$

where $\mathbf{G}$ is the Gaussian smoothing operator and the 3D operator $\mathbf{B}^t\mathbf{G}$ can be written as the tensorial product of three 1D operators (Operto et al., 2003, appendix B). The main aim of the Gaussian smoothing is to filter out the footprint of the sources, receivers and scatterers in the velocity gradient computed in the Cartesian grid.

3.2.5 Synthetic examples

In this section, we assess the adjoint slope tomography with synthetic examples of increasing complexity. In the first and second examples, we design an ideal experimental set-up with a dense distribution of scatterers and a wide-aperture acquisition geometry to prevent velocity-versus-depth ambiguity (Bube et al., 2005) during the joint update of velocities and scatterers. The third example aims to reconstruct a smooth version of the Marmousi model from realistic synthetic picks that would be generated from a towed-streamer acquisition. We assess the accuracy of the slope tomography velocity model by using it as an initial model for frequency-domain FWI. For all of the following tests, we compute the observed data set, i.e. $\{T_{s,r,n_{s,r}}, p_{s,n_{s,r}}, p_{r,n_{s,r}}\}$, in the true velocity model and for the true scatterer positions with our forward-modelling engine. Assessment of adjoint slope tomography in more realistic settings, where two-way traveltimes and slopes are picked in the common shot/receiver gathers or in the common-image gathers, is left to future studies.

Algorithm 4: Adjoint slope tomography workflow. N_{ms}: number of multi-scale step; $\mathcal{S}$: B-spline subdivision operator. For sake of clarity, the model parameters $\mathbf{m}$, the B-spline velocity coefficients $\mathbf{c}$, the misfit function C and its gradient ∇C are indexed by the scale step i and the iteration number k.

Initialization of scatterer positions (Billette et al., 2003, Their appendix A)
Set initial B-spline velocity model $\mathbf{c}_0$
Preliminary re-localization of scatterers in $\mathbf{c}_0$
for $i = 1$ to N_{ms} **do**
  $\mathbf{c}_{0,i} = \mathcal{S}\,\mathbf{c}_{N_{it},i-1}$
  for $k = 1$ to N_{it} **do**
    call L-BFGS($\mathbf{m}_{k-1,i}$,$C_{k-1,i}$,$\nabla C_{k-1,i}$)
    $\mathbf{m}_{k,i} = \mathbf{m}_{k-1,i} + \Delta\mathbf{m}_{k-1,i}$
  end for
end for
$\mathbf{v}_{final} = \mathbf{B}_{N_{ms}}\mathbf{c}_{N_{ms},N_{it}}$

Example 1: circular anomaly

In this example, the true subsurface model consists of a constant-gradient velocity background model ($v(x,z) = v_v + a \times z$) inside a model of 20 km $\times$ 5 km, that contains a smooth circular velocity anomaly of radius 750 m (Fig. 3.8). We use $v_v = 1000$ m/s and $a = 0.9$ s^{-1} in the background model, while the velocity reaches 3820 m/s in the centre of the inclusion. A line of sources, spaced 200 m apart, is set at 500 m depth. Each shot is recorded by five receivers at 500 m depth with offsets of 0.8, 1.6, 2.4, 3.2 and 4 km. The scatterer layout is composed of 155 scatterers with a spacing of 200 m and 700 m in the horizontal and vertical directions, respectively. Each scatterer is located midway between a source and receiver positions. This setting provides a reasonable angular illumination of the target. The same Cartesian discretisation of the velocity model is used for the forward and inverse problems using a grid interval of 50 m. We use a low-pass Gaussian filtering of the misfit function gradient for regularization. The correlation length for this filtering is 200 m. Setting $(\sigma_{T_{s,r}}, \sigma_{p_s}, \sigma_{p_r})$=(1. ms, 0.01 ms/m, 0.01 ms/m) balances the three terms in the misfit function (eq. 3.66).

The initial velocity model is homogeneous with a wave-speed of 1000 m/s. The initial estimates for positions of the scatterers are randomly distributed around their exact positions with a maximum deviation of 400 m (Fig. 3.8a). Since the main goal of this test is to assess the capability of the method in finding the circular anomaly and the scatterer positions, we design the optimization process as follows. Firstly, we find value 0.4 for a that allows for the best data fit using the initial positions of the scatterers. Then, starting from the resulting approximated background velocity model, we jointly update the velocities and the scatterer positions. The final velocity model and scatterer positions after 100 iterations suggest no leakage between the two parameter classes, since the scatterers were moved to their correct positions (Fig. 3.8b). The decrease of the misfit function over iterations is shown in Fig. 3.9. The reconstruction of the inclusion shows some vertical smearing of the bottom end of the inclusion that may result from the narrowing of the angular illumination with depth (Fig. 3.8). In contrast to the

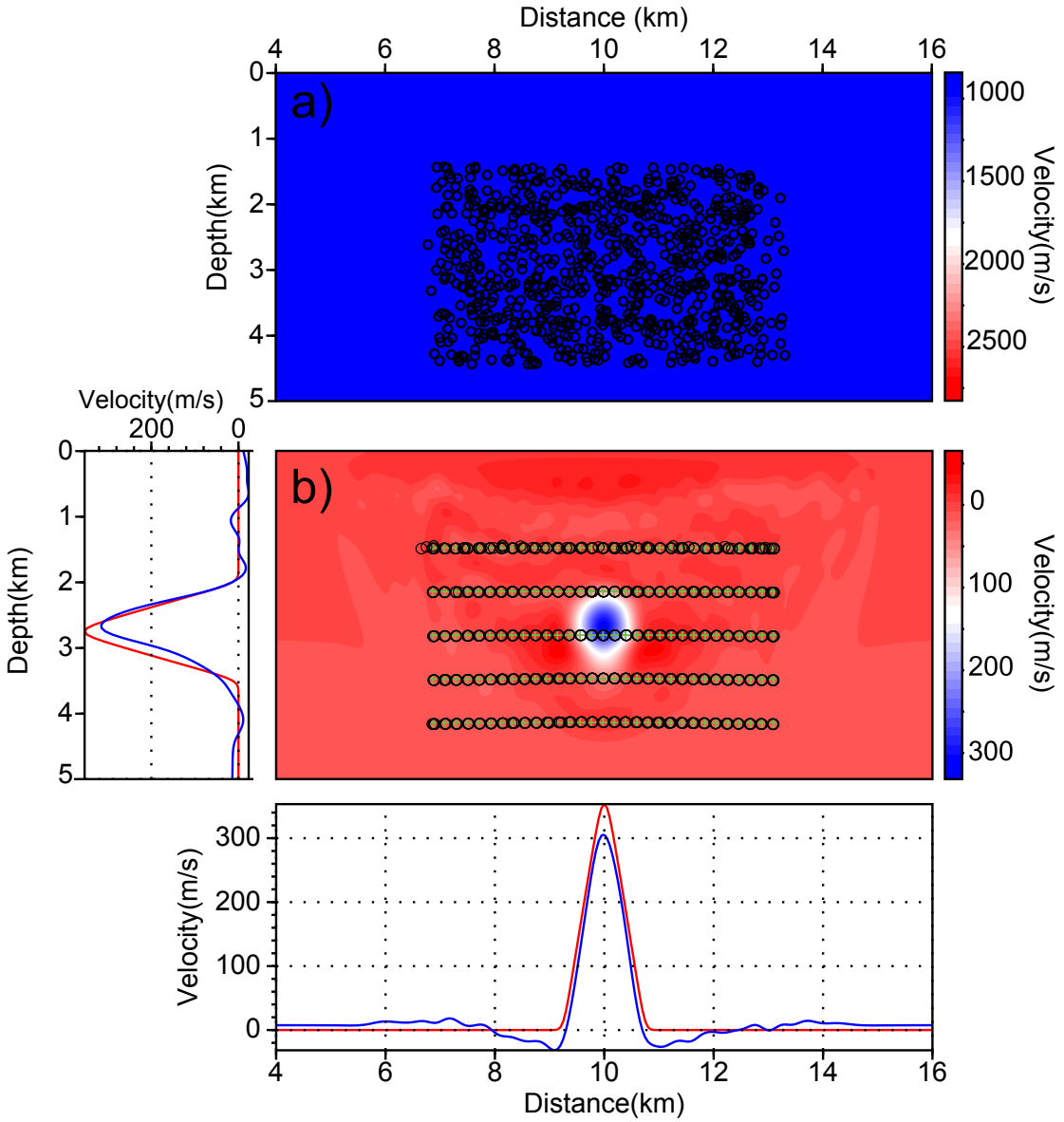

Figure 3.8 – Circular inclusion example. a) Initial position of scatterers, randomly distributed. b) Final adjoint slope tomography velocity perturbation model (inverted velocity model minus the true constant-gradient background velocity) with superimposed exact (cross) and calculated (circle) scatterers positions. Diagrams show the direct comparison between calculated (blue) and exact (red) velocity perturbations across horizontal and vertical profiles cross-cutting the centre of the anomaly.

reconstruction in the vertical direction affected by smearing, some low-velocity perturbations on both sides of the inclusion in the horizontal direction probably reflect a small deficit of low wavenumbers (Fig. 3.8, horizontal profiles). These limited bandwidth effects can be related to

the shape of the stereotomography sensitivity kernels, which connect the scatterers in depth to the sources and receivers at surface (Fig. 3.10). The elongated shape of these sensitivity kernels in sub-vertical directions favours a smooth reconstruction of the vertical wavenumbers as in transmission tomography, while the lateral deviation of the kernels generated by heterogeneities located between the surface and the scatterer may favour the update of intermediate horizontal wavenumbers at the expense of the low components as in reflection tomography.

For illustrative purposes, Fig. 3.10 shows the kernel of the misfit function gradient with respect to the velocity in a homogeneous background velocity model. Here we consider one

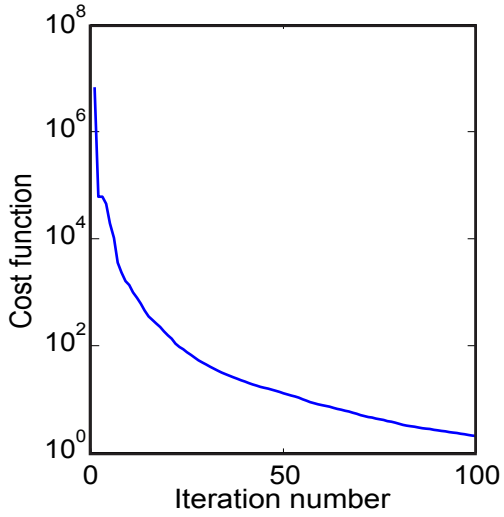

Figure 3.9 – Circular inclusion example. Convergence diagram in logarithmic scale for the test in Fig. 3.8.

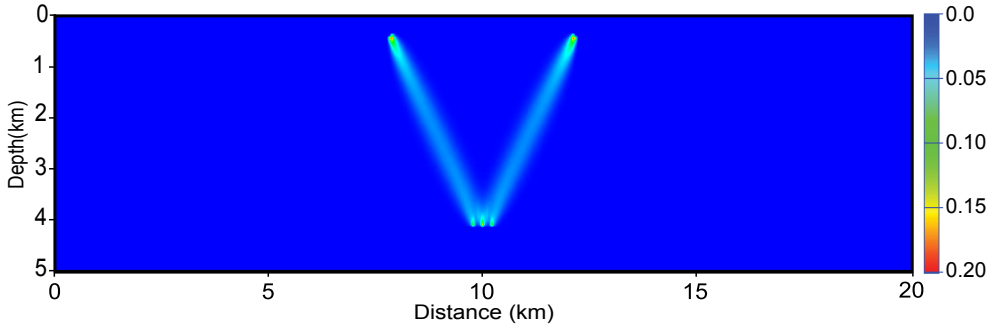

Figure 3.10 – Circular inclusion example. The kernel of misfit function gradient with respect to the velocity in a homogeneous background velocity model for one 4 km-offset source-receiver pair. The kernel consists of adjoint-state variables $\lambda_s(\mathbf{x})$ and $\lambda_r(\mathbf{x})$ which are weighted by the velocity raised power three. Here the source-receiver includes one scatterer at 4.3 km depth. Migration facets can be constructed by neighbour scatterers.

4 km-offset source-receiver pair and one scatterer located midway between the source and receiver positions at 4.3 km depth. This kernel is formed by the superposition of the adjoint-state variables $\lambda_s(\mathbf{x})$ and $\lambda_r(\mathbf{x})$ weighed by the velocity raised to a power three (eq. 3.72). Since the background model is homogeneous, the two neighbouring sources and the two neighbouring receivers that are required to build the right-hand side terms of the adjoint-state eqs (eqs 3.70 and 3.71) involve two scatterers located on both sides of the scatterer in question. The central and the two neighbour scatterers define the spatial support of the source of the adjoint-state equation, that back-propagate the time and slope residuals toward the source (left ray tube) and receiver (right ray tube) positions. The spatial support spanned by the two neighbouring scatterers define the so-called dip bar (here, a horizontal bar) (Billette et al., 2003), from which two rays are shot with appropriate take-off angles in ray-based stereotomography so that they honour the Snell's law at the scatterer. While these two take-off angles are explicitly part of the optimization parameters in ray-based stereotomography, they are implicitly estimated by adjoint slope tomography through the joint localization of the nearby scatterers, hence defining a model space with a more limited number of parameter classes.

Example 2: curved layer anomaly

The target of the second example is a layer with a synclinal geometry that is embedded in a 20 km $\times$ 5 km homogeneous velocity background model (Fig. 3.11a). The aim of this test is to validate the adjoint slope tomography in a classical reflection configuration and assess the contribution of the slope information to overcome potential velocity-depth cross-talk during the velocity estimation. The observables are generated from two series of scatterers that follow the top and the bottom of the layer and a horizontal line of scatterers located below the layer at a depth of 4200 m. This latter line of scatterers provides the necessary illumination in depth to constrain the bottom reflector of the layer.
The velocity in the homogeneous background model is 3000 m/s and the velocity within the smoothed layer anomaly reaches a maximum value of 4000 m/s. A 2400 m single-offset acquisition layout is designed with regular sources and receivers distribution on the surface meaning that each scatterer is reached by only one source-receiver pair. Here, each scatterer is observed with the source-receiver pair which results in minimum traveltime and this causes some scatterers have a common source-receiver pair. This configuration implies that, without slope information, the inversion may not cope with the velocity-depth ambiguity. In this example, we use the multi-scale approach during which the spline interval decreases throughout five steps from 1000 m to 60 m and from 2500 m to 150 m in the vertical and horizontal directions, respectively. The initial velocity model is homogeneous with an overestimated velocity of $v_0 = 3300$ m/s. The initial scatterers are located midway between the source and the receiver at a depth of $v_0 \times t/2$ where t denotes the true two-way traveltimes (Fig. 3.11a). Therefore, the initial dips at the scatterers as predicted by the initial slopes at the source and receiver positions are horizontal and do not embed any prior information on the geometry of the layer. Moreover, these horizontal dips do not match those that would be inferred from the relative positions of the initial scatterers since these latter are not horizontally distributed (Fig. 3.11a).
The velocity model and the position of scatterers that have been reconstructed after 210 iterations are shown in Fig. 3.11b. Here the correlation length of Gaussian filter is 200 m. Overall,

the shape of the syncline is well recovered as shown by the good agreement between the true and the reconstructed lines of scatterers aligned along the top and bottom reflectors with some small mismatches toward the ends of the layer (Fig. 3.11b). The bottom line of the scatterers shows more obvious footprint of velocity-depth ambiguity toward the ends of the line in the form of underestimated depths balanced by two circular blobs of low velocities. Overall, these mismatches remain acceptable owing the limited offset coverage considered in this example. The tomography reconstructs a smooth representation of the layer as shown by the horizontal and vertical profiles in Fig. 3.11. In the vertical direction, the smoothing of the layer generates overestimated velocities just above the top reflector that are balanced by underestimated background velocities in the first 1.5 km in depth. Below the bottom reflector, the velocities tend to be overestimated to balance the underestimated velocities in the smooth reconstructed layer. Toward the middle of the last scatterer line, a slight downward shift of the reconstructed line also contributes to partially balance these overestimated velocities below the layer. Fig. 3.12 shows the convergence diagram for multi-scaling optimization including different five scales.

In order to assess more precisely the contribution of the source and receiver slopes in the inversion, we perform three inversions that rely only on slope information (Fig. 3.13d), traveltime information (Fig. 3.13c) or both (Fig. 3.13b). The Fig. 3.13a shows a close-up of the initial model where the layer exhibits some dips as shown by the superimposed true and initial scatter positions. We initialize the scatterer positions midway between shot and receiver positions, which implies that pairs of scatterers share the almost same position (Fig. 3.13a, red asterisk symbol), while the true scatterers are uniformly distributed along the reflectors (Fig. 3.13a, green plus symbol). We design this configuration in order to assess the ability of the slope tomography to move the scatterers both horizontally and vertically.

Traveltime inversion fails to position correctly the scatterers (Fig. 3.13c). First, the depths of the scatterers on the top interface and bottom interfaces are overestimated and underestimated, respectively. This inaccurate vertical positioning is balanced by erroneous velocity update, manifested by overestimated velocities in the background model and underestimated velocities in the layer, in order to fit traveltimes. This highlights velocity-depth cross-talk during the multi-parameter reconstruction. Second, the traveltime inversion fails to move the scatterer horizontally, that highlights the missing slope constraints as suggested by the results of the slope inversion. In other word, in absence of the slopes information, there is not any constraint to locate the scatterer along the isochrone which is defined by observed traveltime (Chauris et al., 2002a).

Fig. 3.13d shows that the slope-only inversion fails to fit traveltimes, that is highlighted by the fact that the depth of the scatterers associated with the upper interface are underestimated and the velocities in the background model are overestimated. Unlike the traveltime inversion, the vertical mispositioning of the scatterers does not balance the erroneous velocities to honour traveltimes (these latter being not involved in the inversion). However, slope inversion manages to horizontally move the scatterers to their true horizontal coordinates, unlike traveltime inversion. This sensitivity to the horizontal positions can be intuitively understood from the gradient of the misfit function with respect to scatterer coordinates, which involves the difference between sampled neighbouring traveltime maps at the scatterer position, eq. (3.74).

Fig. 3.13b shows how the joint inversion of traveltimes and slopes allows to overcome these

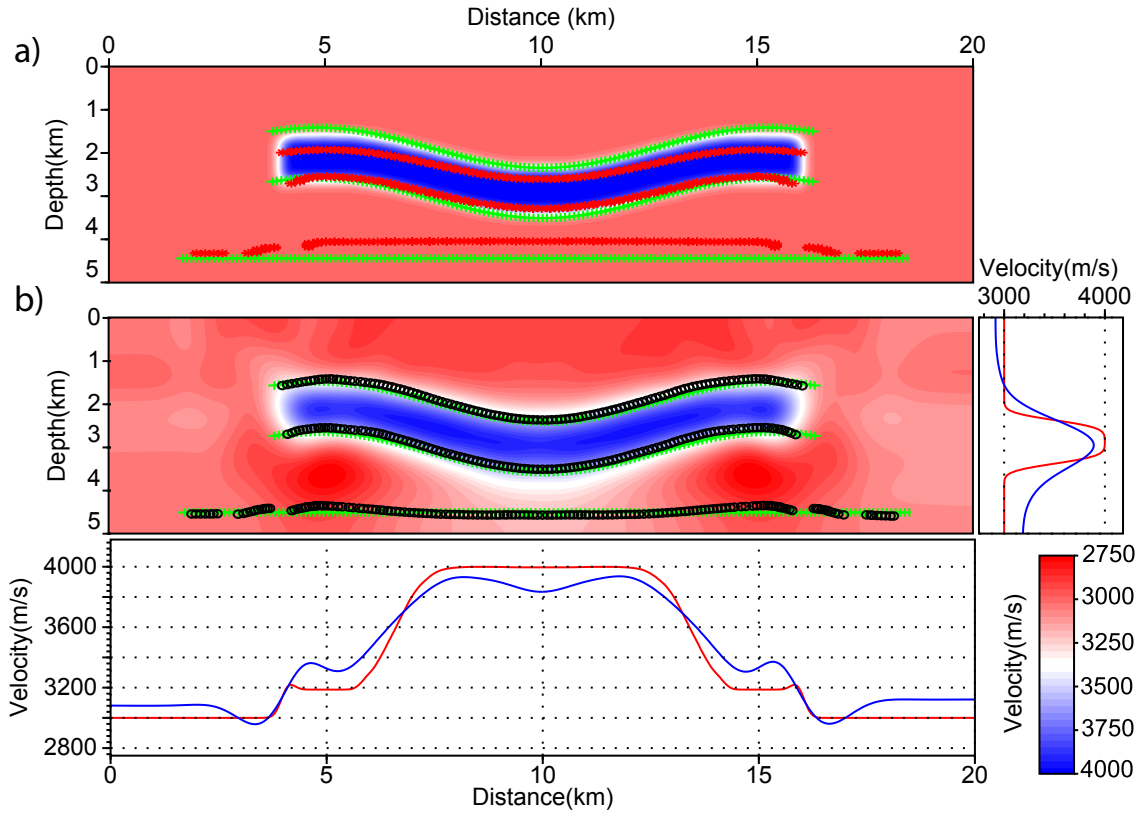

Figure 3.11 – Layer example. a) Exact velocity model with superimposed exact (green "+") and initial (red "*") scatterer positions. b) Reconstructed velocity model by adjoint slope tomography with superimposed exact (green "+"), initial (red "*") and calculated (black "o") scatterer positions. Diagrams show the direct comparison between calculated (blue) and exact velocities (red) across horizontal and vertical profiles cross-cutting the velocity model at 2.5km depth and 10km distance.

artefacts and retrieve both the correct velocities and the horizontal and vertical coordinates of the scatterers.

Example 3: Marmousi model

Building velocity macro-model

We now consider a more realistic example with the complex Marmousi model (Fig. 3.14a) (Bourgeois et al., 1991). Because of its structural complexity, the reconstruction of this velocity model by tomographic methods is challenging. As mentioned by Lambaré (2008) and shown by Billette et al. (2003), ray-based stereotomography has not fully succeeded in reconstructing the smooth components of this model, for three possible reasons: non-linearity of the inverse

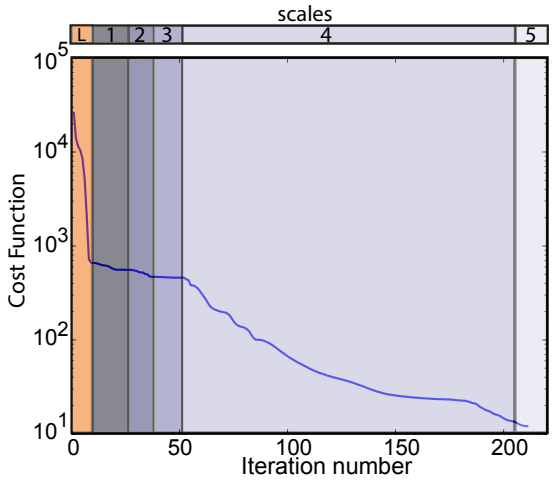

Figure 3.12 – Layer example. Convergence diagram in logarithmic scale for the example in Fig. 3.11. The colours show the localization step ('L') and scales number in multi-scaling approach scheme.

problem, intrinsic limitations of ray theory, and unreliability of stereotomographic picking. We generate the targeted velocity model of the stereotomographic inversion by Gaussian smoothing of the original Marmousi model with vertical and horizontal correlation lengths of 100 m (Fig. 3.14b). This smoothing is sufficiently mild to guarantee that the resulting model provides a good background model for prestack depth migration (e.g., Thierry et al., 1999) or a good initial model for FWI considering a realistic starting frequency of the order of 4 Hz. Note that we use a less aggressive smoothing (100 m correlation length instead of 240 m) than Billette et al. (2003) and Chauris et al. (2002b) to generate the true background velocity model to be reconstructed by stereotomography. In our application this can contribute to make the distribution of scatterers in depth more non-uniform and generate multi-pathing. The grid interval in the smoothed Marmousi model is 20 m.

To build the dataset for inversion, we pick manually the main reflectors in the Marmousi model (Fig. 3.14c) and compute the corresponding reflection traveltimes and slopes numerically with our forward engine. To generate a realistic dataset, we assign a source-receiver pair to each scatterer according to the Fermat's principle by searching the scatterer that has the minimum reflection traveltime along a reflector. Moreover, we check that the found scatterer corresponds to a specular reflection point; if not, we remove its associated traveltime and slopes from the dataset. This condition is fulfilled if the sum of the slowness vectors estimated at the scatterer position from the source and receiver traveltime-map gradients is aligned with the normal to the reflector. We consider a towed-streamer acquisition with offsets ranging between 100 m and 3425 m. Ninety-one shots every 100 m are recorded by 134 receivers spaced 25 m apart. The first shot is located at a distance of 7400 m in Fig. 3.14 and the acquisition layout is moving from right to left, implying a deficit of scatterers in the bottom-right part of the target. According to this source-receiver geometry, our reflector picking in the Marmousi model (Fig. 3.14c) results in more than 6000 observables in the data domain. Two shot gathers located at $x = 800$ m and

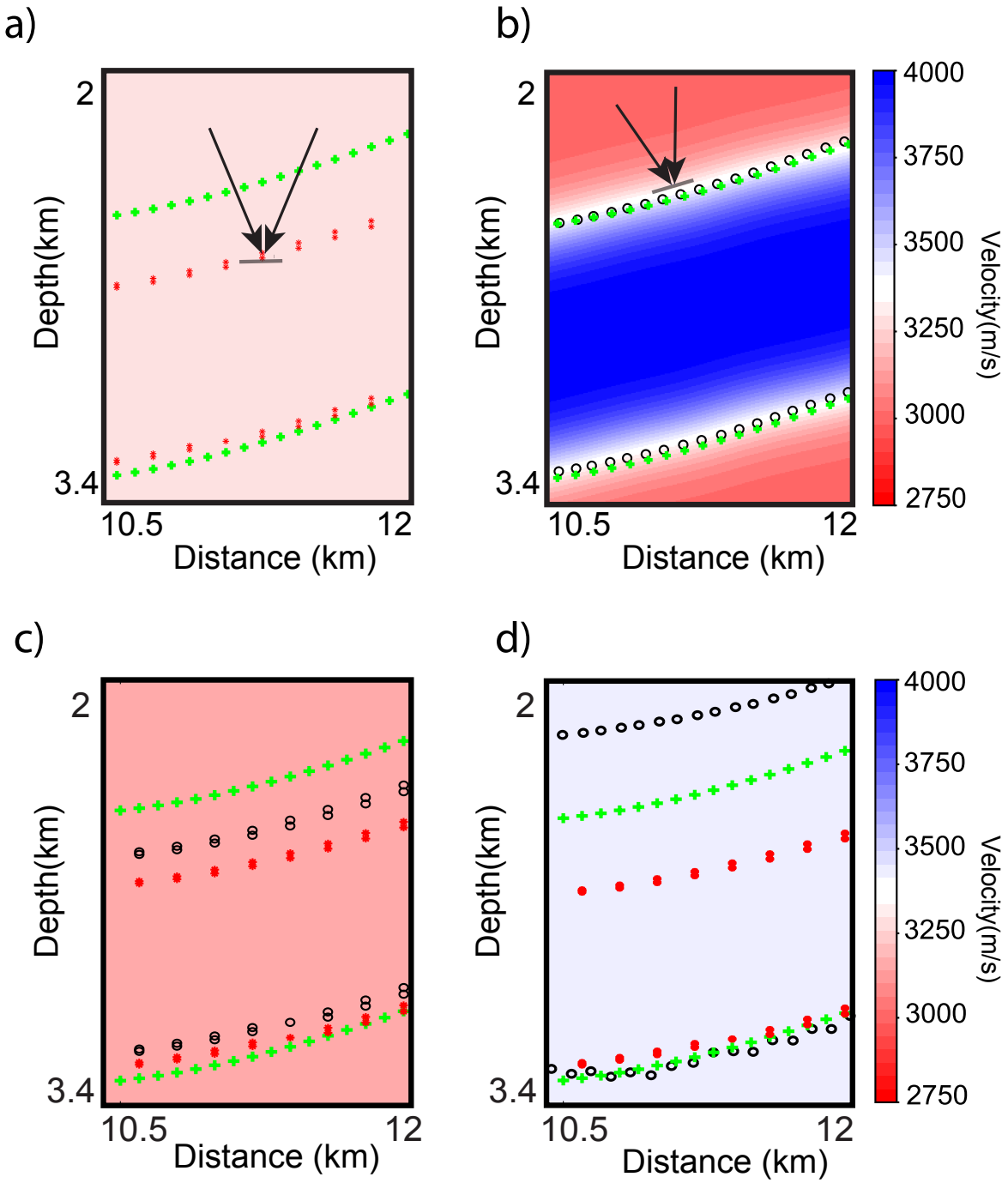

Figure 3.13 – Layer example. Comparison between joint traveltime+slope, traveltime only and slope only inversions. a) Close-up of the initial velocity model with superimposed true (green "+") and initial (red "*") scatterer positions. The initial positions of the scatterers are midway the source and receiver positions meaning that the local dip predicted by the shot and receiver slopes would be horizontal. Note that pairs of scatterers share the same initial position. (b-d) Close-up of the velocity model after (b) joint traveltime and slope inversion, (c) traveltime-only inversion, and (d) slope-only inversion (d). The retrieved scatterer positions are shown by black circles. See text for detailed interpretation.

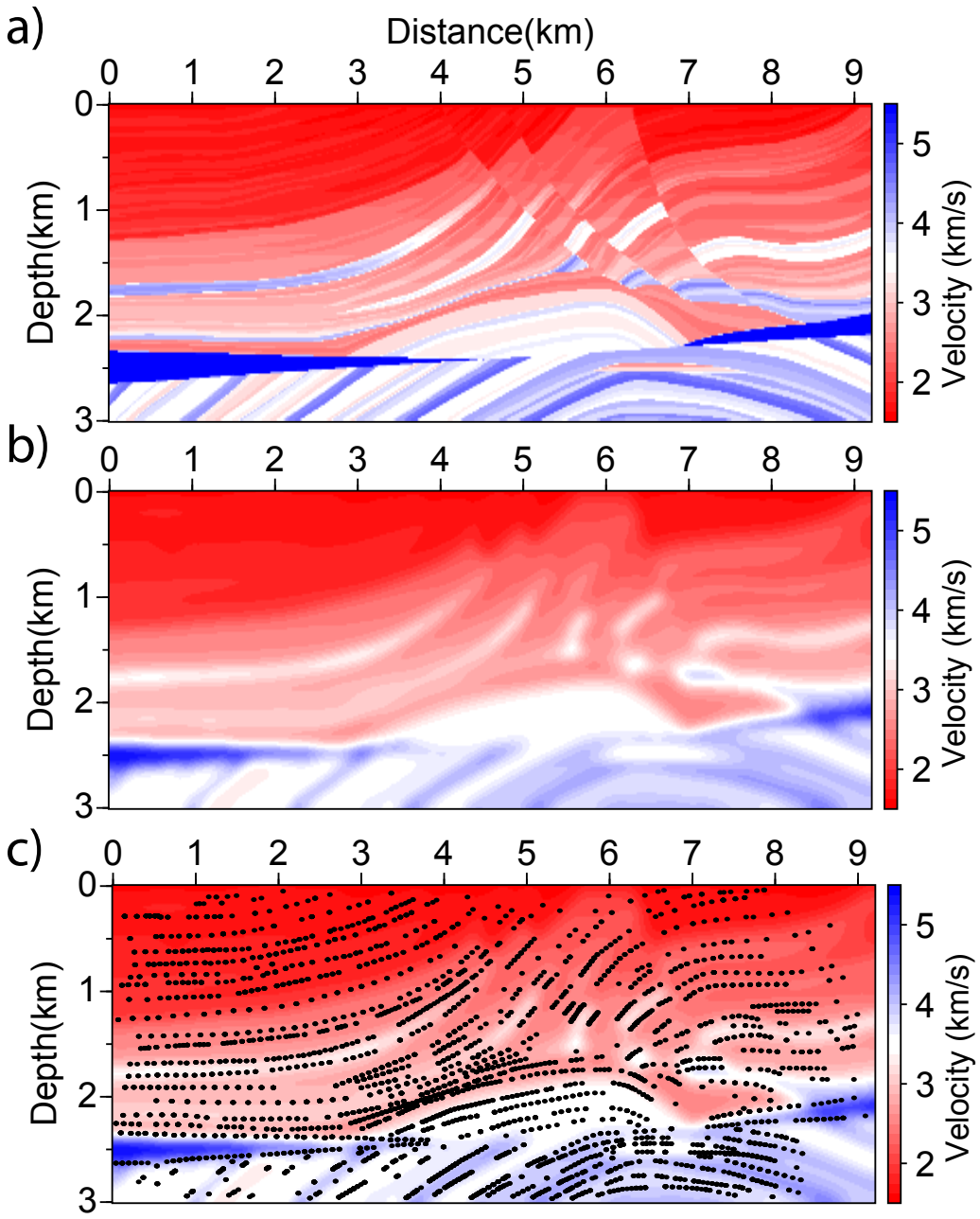

Figure 3.14 – Marmousi example. (a) True velocity model, (b) Smoothed Marmousi velocity model used as the targeted model of the adjoint slope tomography. (c) Scatterers that have been used to generate the traveltimes and slope dataset from the velocity model shown in (b).

$x = 4800$ m are shown in Fig. 3.15 with superimposed traveltimes and slopes generated from the reflector picks shown in Fig. 3.14c. Both of these gathers include the deepest scatterers; the shot gather at $x = 4800$ m, unlike the other one, includes the reflection from complex geological structures. Our picking of the Marmousi reflectors should be reasonably representative of the

one that would have been performed in the prestack migrated domain (namely, in common-image gathers). Indeed, traveltime and slope observables can be generated from image-domain picks by demigration (Chauris et al., 2002a; Guillaume et al., 2008), in a manner comparable to our ad-hoc approach except that we took advantage of our knowledge of the true model to measure the dips instead of picking them. Our reflector picking leads to a sparse distribution of locally-coherent events in the gathers as illustrated in Fig. 3.15. Note that some picks in the complex zone have not satisfied the specular-point condition. Therefore, our inversion can suffer from a deficit of illumination in this complex part of the model.

In order to reconstruct the smooth Marmousi velocity model, we follow the stereotomographic workflow introduced in the section 3.2.4. During the localization step and the following slope tomographic inversion, we use $(\sigma_{T_{s,r}}, \sigma_{p_s}, \sigma_{p_r})$=(1. ms, 0.01 ms/m, 0.01 ms/m). Also, we estimate an appropriate correlation length of 200 m for Gaussian smoothing by trial and error. After the initialization step (Fig. 3.16a), we performed a localization of the scatterers from a homogeneous velocity model with a 2000 m/s velocity (Fig. 3.16b). Then, we removed outliers associated with scatterers that fall outside the limits of the subsurface target. After localizing the scatterers, we performed four successive slope tomographic inversions by refining the velocity grid by a factor 2 in both directions at each step. During these four multi-resolution steps, the horizontal and vertical B-spline node spacings decrease from 1300 m to 160m and from 400 m to 50 m, respectively. During these four steps, the inversion has performed 3, 147, 50 and 165 iterations to generate the velocity models and the scatterer positions shown in Fig. 3.16(c-f). Overall, the positions of the scatterers tend to align with the reflectors of the Marmousi model as the resolution of the velocity model improves over the four multi-scale steps with however, a certain hierarchy driven by the local structural complexity of the velocity model. During the first-scale inversion, the scatterer positions and the background velocities were not significantly updated because of the B-spline grid intervals are too coarse (Fig. 3.16c). Large-scale variations of the background velocities are introduced in the whole model during the second step; however, the scatterers start being aligned with the reflectors only in the left part of the model where the structure is simpler (Fig. 3.16d). During the third step, the scatterer positions are moved to their correct positions in the upper-right part of the model but the inversion fails to significantly update their positions at the reservoir level (Fig. 3.16e). During the fourth step, the scatterers are aligned along the reflectors with improved accuracy as the resolution of the velocity model is improved with the most significant updates in the complex zone at the reservoir level (Fig. 3.16f). Through these iterative updates, the most obvious inaccuracies occur at the reservoir depths and near the bottom-right end of the model where the overburden exhibits significant dips and the acquisition geometry provides a limited illumination. The misfit function plotted as a function of the iteration number over the localization and the four multi-scale steps shows a regular decrease of the data misfit, suggesting a reasonable tuning of the inversion (Fig. 3.17). If we would have used a constant-gradient velocity model during the localization instead of a homogeneous model, we could have started the slope tomography inversions on a finer spline grid to converge in a smaller number of iterations toward a final model almost identical to the one shown here. In this test, by considering a homogeneous background velocity model, we intended to assess the capability of the model to image smooth components of complex structures almost from scratch.

As a quality control of the final results, we have superimposed the dip bars onto the original

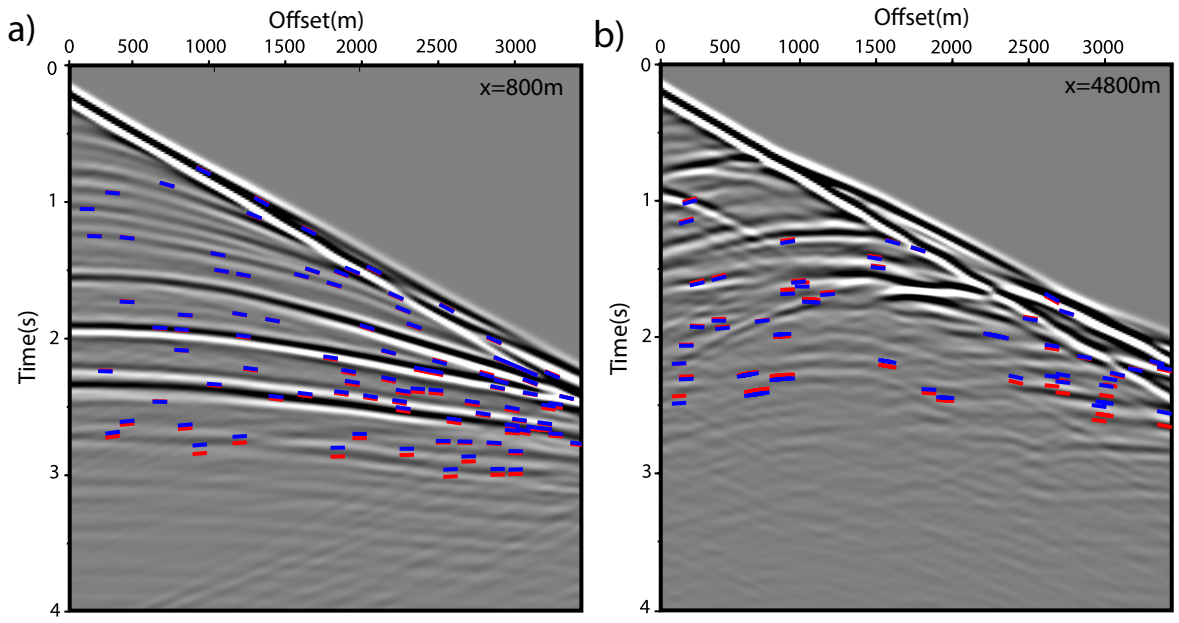

Figure 3.15 – Marmousi example. Examples of traveltime and slope observables (red segments) generated from the velocity model of Fig. 3.14b and the scatterers of Fig. 3.14c. The shot positions are x=800m(a) and x=4800m(b). Each red segment shows the traveltime and the slope of each picked event. The blue segments represent the traveltimes and slopes computed in the final adjoint slope tomographic model (Fig. 3.16f).

Marmousi velocity model (Fig. 3.18b). The dip associated to each scatterer is computed a posteriori from the traveltime gradient vectors $\nabla \mathbf{t}_s$ and $\nabla \mathbf{t}_r$ computed in the slope tomography velocity model (Fig. 3.18a). We also show the direct comparison between the dip bars (black segments) and the true position of the scatterers resulting from the picking of the Marmousi reflectors in Fig. 3.18c. Fig. 3.18a,c show a fairly good agreement at all depths in the left part of the Marmousi model where the dips are reasonably mild. In the more complex central and right parts, the alignment remains acceptable in the shallow part, while the adjoint slope tomography failed to provide a reliable reconstruction at depths greater than 2.5 km beyond 6 km of distance. This mispositioning of the dip bars in the complex deep part manifest as significant late traveltime misfit in the shot gathers of Fig. 3.15. An insufficient illumination aperture at the reservoir depths, resulting from the strong lateral variations in the overburden and the position of the target near the right end of the acquisition, might explain this failure.

The final estimated velocity model by adjoint slope tomography is complex enough to generate multi-valued ray fields. This is illustrated by calculating rays+wavefronts (Lambaré et al., 1996) for this velocity model (Fig. 3.19a). In Fig. 3.19a the rays+wavefronts are superimposed with the first-arrival wavefronts computed with our eikonal solver for a source at $x = 6$ km. The complexity of the slope tomographic velocity model generates caustics which are singular points for the first-arrival wavefronts (Fig. 3.19b). Our finite difference approach for the slope estimations near these caustic points can be erroneous; another possible reason for misposi-

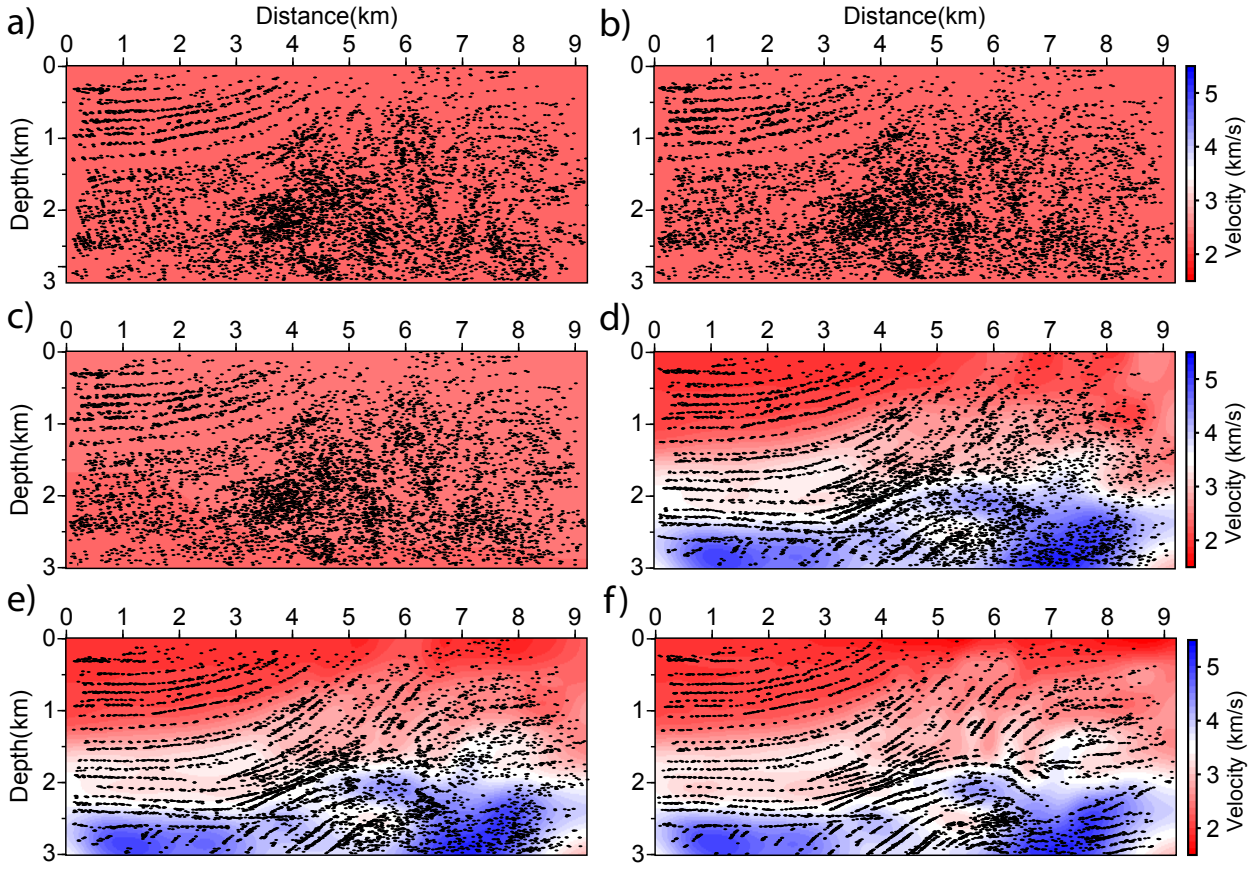

Figure 3.16 – Marmousi example. Adjoint slope tomography results. Velocity model and scatterer positions after (a) the initialization step, (b) the localization step and (c-f) four multi-scale adjoint slope tomographic inversions with spline-grid refinement. See text for details.

tioning of the dip bars in the complex zone (Fig. 3.18c). However, the calculated first-arrival wavefronts with the eikonal solver are completely matched with the solutions of wavefront construction method.

FWI with initial model from adjoint slope tomography

We now assess the quality of the adjoint slope tomographic velocity model (Fig. 3.20a) as a starting model for frequency-domain FWI of long-offset data. As such, we consider a stationary-receiver acquisition geometry with a 9 km maximum offset. The aim of this increased offset range is two fold. First, it increases the non-linearity of FWI associated with cycle skipping of both diving waves and wide-angle reflections and hence provides a suitable framework to assess the adjoint slope tomographic model as a good initial model for FWI. Second, this long-offset acquisition provides a sufficient wide scattering-angle illumination to prevent some

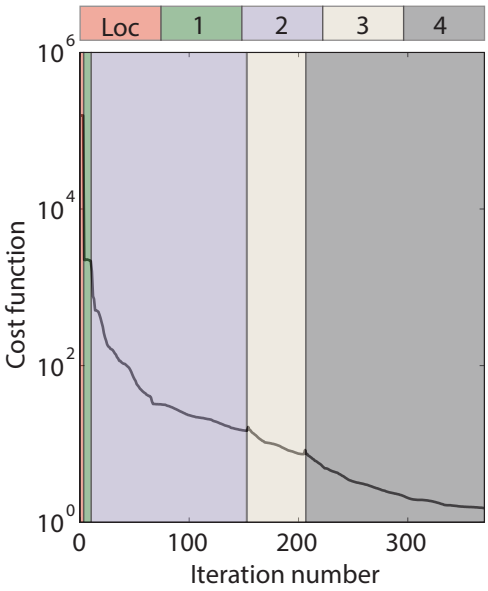

Figure 3.17 – Marmousi example. Misfit function versus iteration number plotted with a logarithmic scale. The colours delineate the localization step ('Loc') and the scales number the multi-scale inversion.

notches between the wavenumber content of the slope tomographic velocity model and that of the perturbation model generated by FWI when a starting frequency as high than 4 Hz is used. The acquisition consists of 41 sources every 200 m that are recorded by 231 receivers spaced 40 m apart. We invert sequentially five frequency components of the wavefields ($4, 5, 7, 9, 12$ and 16 Hz) with the L-BFGS algorithm. The final inverted velocity model by FWI (Fig. 3.20b) shows a good qualitative agreement with the true Marmousi velocity model except near the bottom-right of the model (Fig. 3.14b). Fig. 3.21 shows a more quantitative assessment of the accuracy of the FWI model by the direct comparison between a series of vertical profiles extracted from the final adjoint slope tomographic model, the final FWI model and the true Marmousi model. The agreement between the FWI and the true profiles is good down to 2 km depth both in terms of resolution content, positioning in depth and velocity estimation, except near the right-end of the model. The resolution degrades near the bottom of the model as the number of available events becomes more limited.

3.2.6 Discussion

We have presented a new formulation of stereotomography. The key ingredients are the computation of the traveltimes with an eikonal solver, the inference of the slopes at the source and receiver positions from these traveltimes by finite differences and the computation of the misfit function gradient with the matrix-free adjoint state method. Our formalism relies on a limited number of parameter classes (subsurface velocities and scatterer coordinates) compared to the ray-based stereotomography developed by Billette and Lambaré (1998). This directly

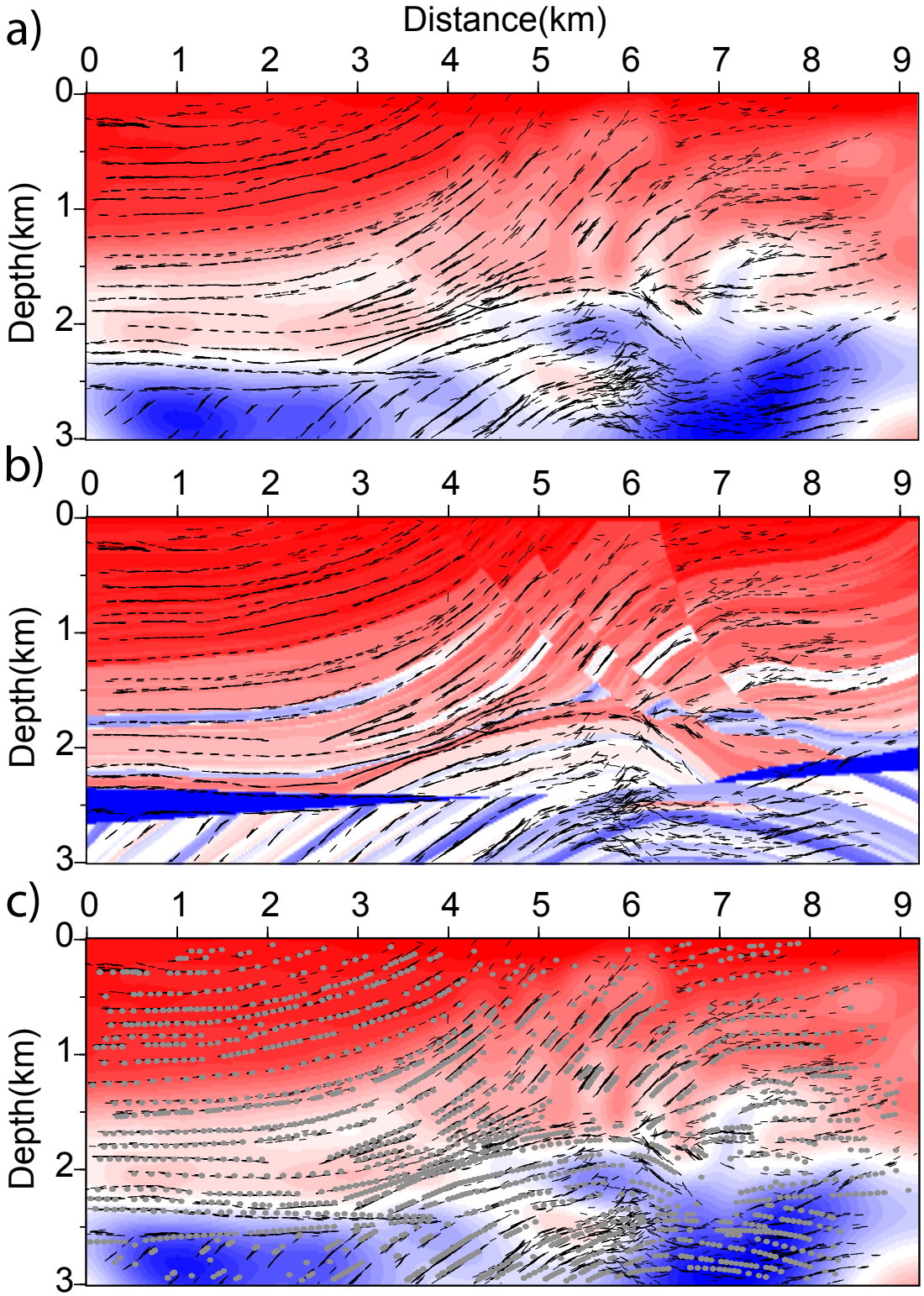

Figure 3.18 – Marmousi example. (a-b) Scatterer positions with dip bars estimated by adjoint slope tomography superimposed on the final inverted velocity model (a) and on the true Marmousi model (b). c) Direct comparison between the reconstructed dip bars (black segments) and the scatterers (grey circle) that have been used to generate the data set (Fig. 3.14c).

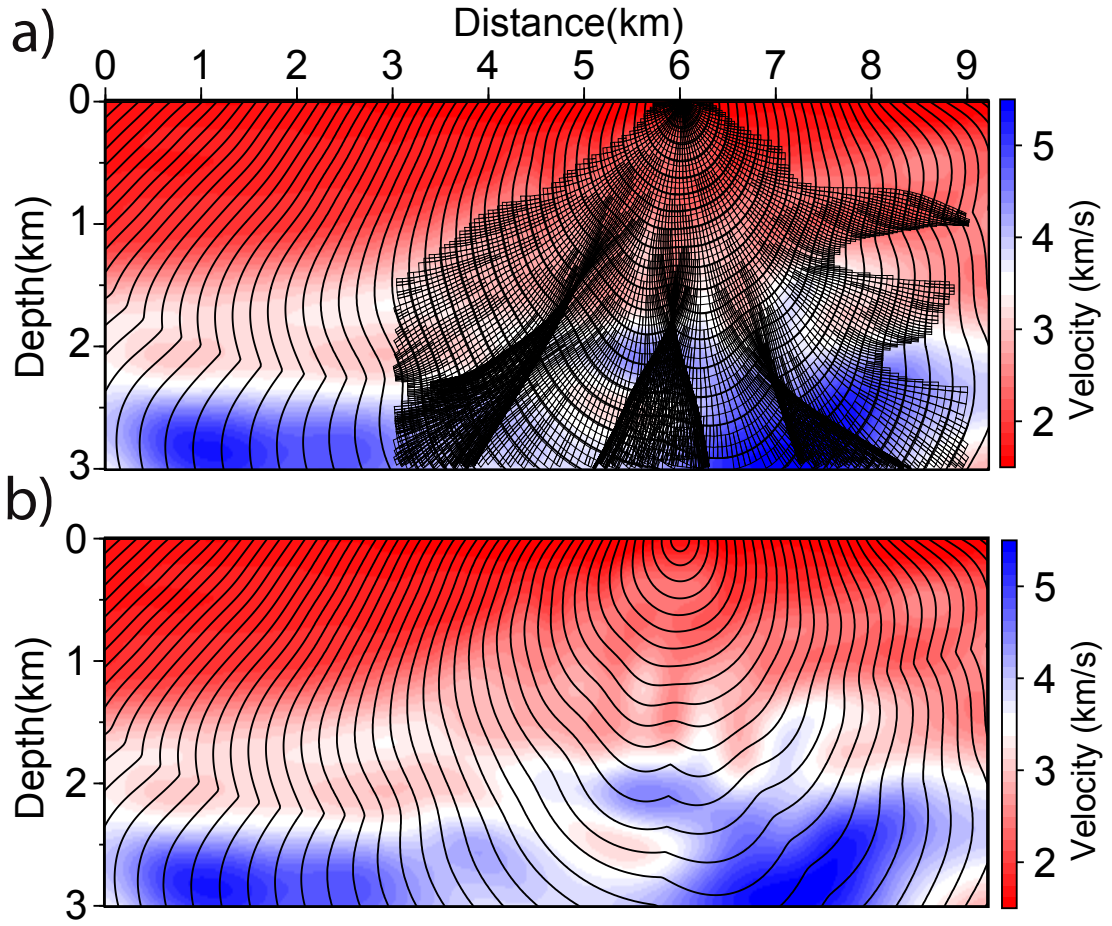

Figure 3.19 – Marmousi example. a) Smooth marmousi model built by slope tomography with superimposed ray+wavefronts, computed with the wavefront construction method of Lambaré et al. (1996), and wavefronts computed with the factored eikonal solver. (b) Same as (a) without the ray field.

results from the use of an eikonal solver instead of a ray-tracing algorithm to compute traveltimes and slopes. The eikonal equation using the shots and receivers as sources and the finite-difference estimation of the slopes from the traveltime solutions of the two eikonal resolution define the state equation of the adjoint problem. The adjoint-state variables associated with the source-side and receiver-side traveltimes are solution of a transport-like equation, which builds the sensitivity kernels of the velocity update along two ray tubes connecting the scatterer to the sources and receivers. The right-hand side of these adjoint-equations describe a source term whose spatial support, weighted by the two-way traveltime and slope residuals, defines a migration facet or dip bar at the scatterer position. The position of a scatterer is updated from the local gradients of the traveltime maps generated from the sources and receivers to which it is related and also from the gradient of the traveltimes generated by neighbouring sources and receivers. Again, the contribution of neighbour sources and receivers highlight the additional constraints provided by slopes to localize the scatterers. Our parsimonious parametrization may contribute

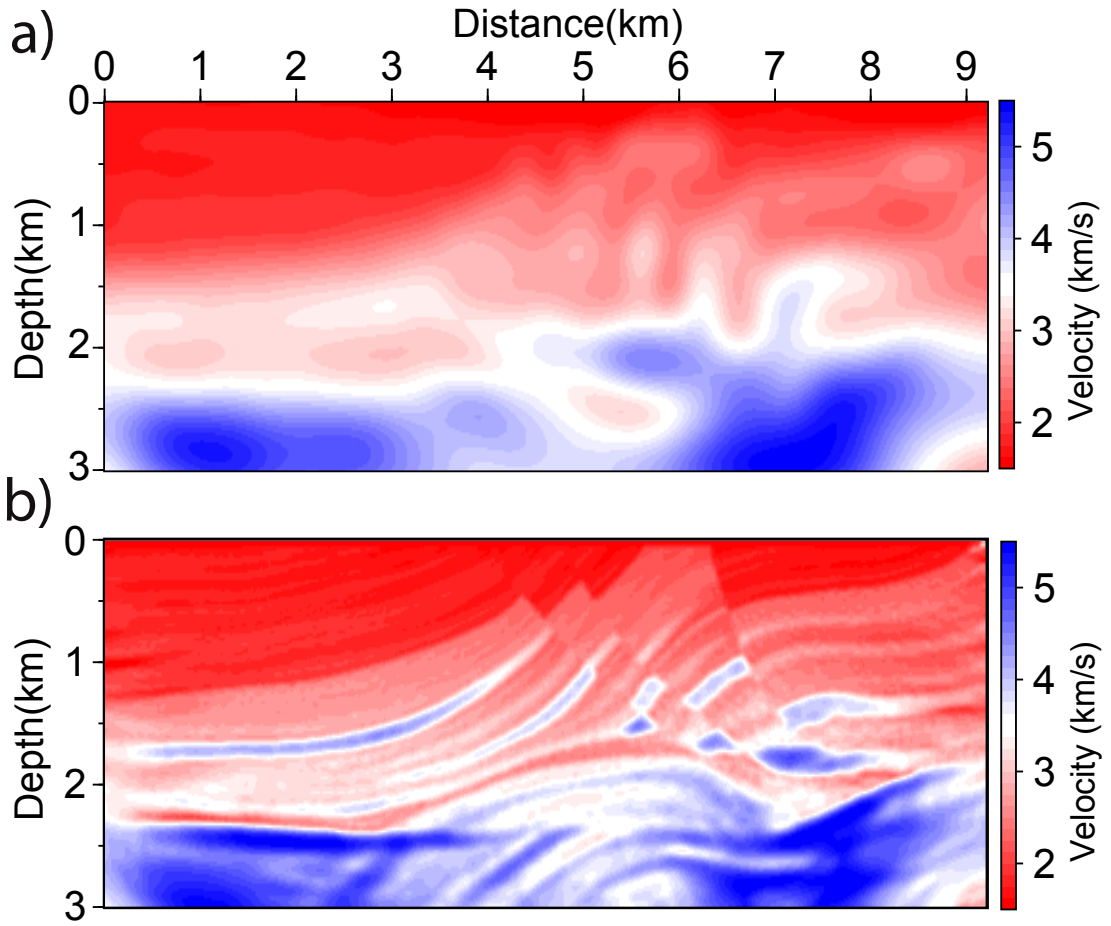

Figure 3.20 – Marmousi example. a) Final inverted velocity model by adjoint slope tomography. b) Final FWI velocity model using the adjoint slope tomography model (a) as initial model.

to make the inverse problem underlying adjoint slope tomography more robust and easier to tune than the ray-based counterpart by avoiding over-parametrization than can be lead to significant parameter cross-talk and ill-posed inversion (Billette and Lambaré, 1998). Moreover, the computational burden resulting from the use of an eikonal solver instead of ray tracing is efficiently balanced by the fact that traveltime computations are performed from the source and receiver positions rather than from the scatterers. The second advantage is related to our ability to process large-scale dataset without any matrix resolution. In return, we do not have access to the sensitivity matrix that is useful to perform local a posteriori resolution analysis. However, the computational efficiency of our approach can be used to build the resolution matrix numerically through spike tests. Alternatively, second-order adjoint-state method can be used to implement truncated Newton optimization as an alternative to the L-BFGS optimization. The efficient computation of Hessian-vector product with second-order adjoint state method (Mé-

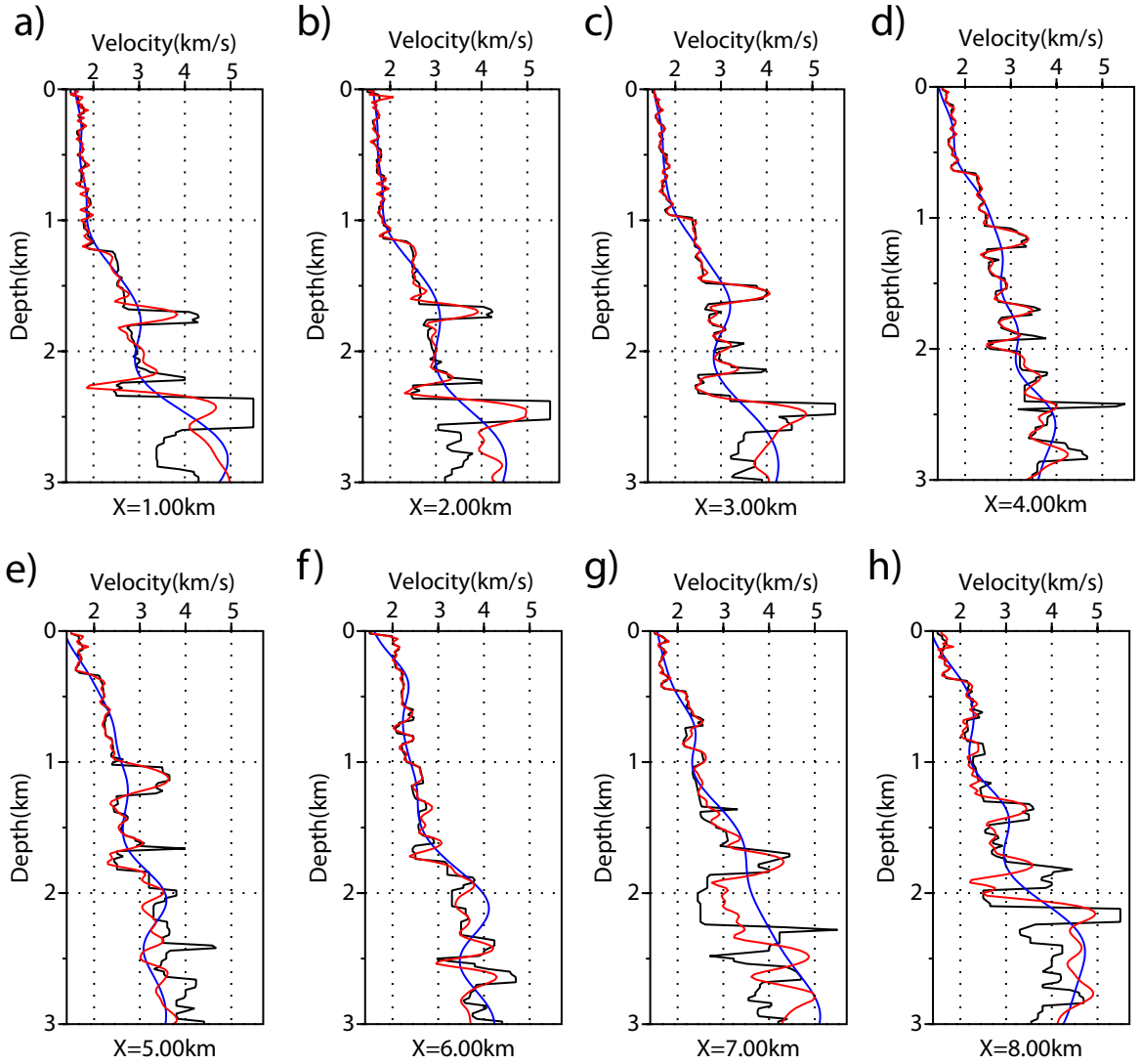

Figure 3.21 – Marmousi example. Direct comparison of logs between exact velocity model in Fig. 3.14(a), black line, and the estimated velocity by adjoint slope tomography in Fig 3.20(a), blue line, and inverted velocity model by frequency domain FWI in Fig 3.20(b), red line, while using adjoint slope tomography solution as initial model.

tivier et al., 2013, 2014a) can be also used to perform this resolution analysis (Fichtner and Trampert, 2011). Implementation of the truncated Newton method in our adjoint slope tomography is ongoing.

One of the main limitations of our method compared to more automatic waveform inversion techniques such as reflection waveform inversion (Xu et al., 2012; Brossier et al., 2015; Zhou et al., 2015) is related to their ability to handle complex structures (Lambaré, 2008). The main bottleneck remains the picking of local coherent events, which has not been investigated in this

study. Beyond the picking issue, any fast solver providing the traveltime of the most-energetic arrival (Kim, 2001) or of all the arrivals (Qian and Leung, 2004) instead of the first traveltime might improve the imaging in complex area where ray folding occurs, as has been shown in a prestack-depth migration context (Operto et al., 2000).

Chauris et al. (2002a) and Chauris et al. (2002b) have shown how the picking and stereotomographic inversion can be implemented in the prestack depth migration domain. Indeed, our approach can be also used when picking is performed in the migrated domain after demigration (modelling) of seismic invariants (Guillaume et al., 2008). The benefit of this strategy is that picking can be performed more easily in the migrated domain after suitable preprocessing of the common-image gathers, while the demigration allows one to avoid repeating migration at each iteration of the slope tomography. Chauris et al. (2002a) and Chauris et al. (2002b) also show how migration-based velocity analysis (or slope tomography after demigration) can be performed with a reduced data-space and model-space parametrization involving one slope in the data space and the velocity model in the model space. This reduced parametrization can be considered in the framework of our adjoint formulation, although its robustness should be assessed through numerical investigation. Do we converge to the true minimum, when considering implicitly scatterer positions? Do we mitigate non-linearities issues when adding these positions to the model space?

In slope tomographic methods, one scatterer is related to one shot-receiver pair and all of the scatterers are processed independently during the inversion. However, due to the intrinsic redundancy of seismic multi-offset acquisition, one position in the subsurface is sampled by many shot-receiver pairs. Any optimization constraints on the relative position (i.e., proximity) of scatterers and the alignment of facets that are related to neighbouring sources and receivers should make the inversion more robust.

In this study, we assess adjoint slope tomography for short-spread reflection acquisitions. The use of an eikonal solver provides a robust modelling engine to extend adjoint slope tomography to long-offset acquisitions where traveltimes and slopes of both diving waves and reflected waves can be involved in the inversion as shown by Prieux et al. (2013). Use of long-offset data raises the issue of anisotropy, which can be taken into account with a factored eikonal solver proposed by Waheed et al. (2015) and Tavakoli F. et al. (2015) or more sophisticated approaches based on the resolution of the Hamilton-Jacobi equation with level-set techniques (Qian et al., 2003). Extension of adjoint slope tomography to TTI anisotropy is part of our future work. In this long-offset framework, adjoint slope tomography can provide reliable initial model for FWI in the sense that both approaches share the idea of using the full information content in the data through a dense picking of high-frequency scattered waves in slope tomography and the use of the waveforms of all of the arrivals in FWI.

3.2.7 Conclusions

We develop an adjoint formulation of the slope tomography as a new formulation of stereotomography to avoid large-scale matrix resolution. Moreover, a finite-difference eikonal solver is proposed as an alternative to ray tracing to avoid the issue of non-uniform ray sampling that

can arise at long offsets and in complex models. Compared to the classical ray-based implementation of stereotomography, the model space involves subsurface velocities and scatterer positions, but not shooting angles and one-way traveltimes. The local subsurface dips, that are parametrized by the two shooting angles in ray-based stereotomography, are implicitly represented in the spatial support of the adjoint source and reconstructed through the update of close scatterers. The eikonal resolutions are performed from the shot and receiver positions to mitigate the number of the forward modelling required by the misfit function gradient: the number of state and adjoint problems to be computed scales to the number of sources and receivers but is independent of the number of scatterers, hence leading to an efficient computational framework. Exploiting the reciprocity principle of wave propagation, the slopes at the source and receiver positions are estimated in a finite-difference sense from traveltimes extracted in the subsurface far away from the source singularities. In this framework, source and receiver positions need not to be considered in the data space as traveltimes can be sampled accurately at arbitrary positions in the finite-difference grid with suitable interpolation schemes. The chosen smaller model space, compared to the model space in classic ray-based stereotomography, mitigates potential leakage between parameters and balances possible non-linear features of the optimization scheme. The L-BFGS optimization scheme, which accounts for the Hessian in a cheap way, mitigates the leakage between wave speeds and scatterer positions. Other second-order optimization algorithms, such as the Truncated Newton method, can be also implemented in the adjoint slope tomography. Our approach takes advantage of a multi-scale strategy during which the velocity grid is progressively refined. Through the estimation of velocity model for Marmousi model, we show that adjoint slope tomography is an appropriate candidate for building background velocity model for prestack depth migration and FWI. We plan to extend our formulation to 3D TTI media as well as to consider applications to real case studies.

Acknowledgements: We thank René-Edouard Plessix, Samuel Gray and an anonymous reviewer for useful comments which have improved the writing of this paper. This study is partially funded by the SEISCOPE consortium sponsored by CGG, CHEVRON, EXXON-MOBIL, JGI, SHELL, SINOPEC, STATOIL, TOTAL and WOODSIDE (see website *http://seiscope2.osug.fr*). We thank L. Métivier (LJK/CNRS) for providing us the SEISCOPE optimization toolbox, R. Brossier for the discussions on the inverse scheme and Gilles Lambaré and Thibaut Allemand for fruitful discussions on stereotomography.

3.3 Inverse problem in anisotropic adjoint slope tomography

In this section, I discuss the inverse problem in anisotropic adjoint slope tomography. The general framework is an extension of the isotropic formulation presented in section 3.2. This section is the first draft of a paper and its form follows a paper style.

3.3.1 Introduction

Different applications of seismic imaging techniques, such as migration and FWI, have proven that a precise and accurate subsurface image requires to consider anisotropy (Alkhalifah and Larner, 1994; Prieux et al., 2011; Pratt et al., 2001; Warner et al., 2013; Berkhout, 2014). Anisotropic seismic imaging, as an inverse problem, consists of two main ingredients: 1) a forward modelling engine to track the wave propagation in anisotropic media, and 2) a non-linear multi-parameter inverse scheme to manage the cross-talk between the different parameter classes. Because the inverse problem for general anisotropic elastic media is highly ill-posed, the Earth anisotropy is approximated by a simplified model in many applications. One of the most common approximation relies on the assumption that the subsurface is layered with a vertical or tilted symmetry axis. These subsurface models are generally referred to as VTI and TTI media, respectively. In the following, we review some of the methods which aim to estimate the anisotropic properties in VTI/TTI or more complicated models.

NMO-type velocity analysis is among the first approaches to update subsurface velocity in presence of anisotropy (Tsvankin and Thomsen, 1995; Alkhalifah and Tsvankin, 1995; Grechka and Tsvankin, 2000). This method analyses NMO moveout in transversely isotropic (TI) media by considering long-spread non-hyperbolic reflections, and try to estimate the quartic term of Taylor series expansion for P and vertical S waves t^2 - x^2 curves. Alkhalifah and Tsvankin (1995) simplified and stabilized this approach for acoustic VTI media by formulating the method with ray parameters. They showed that, P-wave NMO velocity for the dipping reflectors in a homogeneous transversely isotropic (TI) medium can be fully described by zero-dip NMO velocity and the effective η. They extended this formulation to vertically inhomogeneous media by generalizing the NMO equation of Tsvankin (1995) for anisotropic layering media (Alkhalifah and Tsvankin, 1995). They derived these parametrization for weak anisotropy, but in practice it is applicable for arbitrary strength of anisotropy. Here, the interval velocities are calculated by Dix formula and during inversion this calculation of velocity for layers leads to some instabilities due to error accumulation at depth (Zhou et al., 2004). Therefore, for complex structures, this approach requires an advanced expertise for avoiding instabilities.

Traveltime tomography (Aki et al., 1977; Nolet, 1987) is a well-known method for depth imaging in isotropic and anisotropic media. The forward-modelling engine in anisotropic traveltime tomography generally relies on ray tracing (Červený, 2001) while different approaches based on prior information and regularization have been developed to deal with the ill-posedness of the inverse problem. Using additional information, like VSP traveltimes, which puts more constraints on the reconstructed parameters is a common practice to resolve the non-uniqueness of the inversion solutions (Tsvankin, 2001b; Bakulin et al., 2010; Zhou and Greenhalgh, 2008;

Sexton and Williamson, 1998). Taking advantage of regularization techniques in inverse problem, is another approach where one can partially manage the ambiguity between parameters (Zhou et al., 2011; Wang and Tsvankin, 2013). Clapp et al. (2004); Fomel (2007) designed a shaping filter as regularization operator to update the subsurface models such that they follow the geological structures. The tomographic methods also differ in the way they manage the multi-parameter reconstruction. Zhou et al. (2011) proposed a joint inversion of all the subsurface parameters while Jiang and Zhou (2010), through an error analysis, showed that a hierarchical approach provides more reliable results; they first reconstruct the vertical velocity and ϵ before updating δ once the misfit function has been decreased by a sufficient amount. Woodward et al. (2008) applied a mono-parameter inversion for the vertical velocity and then estimate ϵ and δ by zeroing the RMO at well locations.

Migration velocity analysis (MVA) for anisotropic media relies on the same principle of isotropic MVA: flattening the reflector on common image gathers (CIGs). Sarkar and Tsvankin (2004) introduced a MVA method which takes advantage of factorized blocks for VTI in which anisotropic parameters are constant while velocity can vary. Behera and Tsvankin (2009) extended this method to TTI media where the tilted angle in each block respects the geological dip (Audebert et al., 2006). They use a Kirchhoff prestack depth migration, based on anisotropic ray tracing for heterogeneous TI media, to iteratively perform the migration. For complex subsurface models this method can be less efficient due to the limitations of ray-based modelling techniques in complex media and the constant-property assumption underlying the block factorization strategy (Wang and Tsvankin, 2013). Wave-equation MVA (WEMVA) (Sava and Biondi, 2004) avoids the ray-based methods by using a linearized wave equation operator to update image perturbations. Extension of this approach to 2D VTI media is introduced by Li and Biondi (2011); Li et al. (2012). Weibull and Arntsen (2014) formulated WEMVA for 2D TTI media where they update jointly v_v, ϵ and δ and utilize a regularization term in the misfit function. They also emphasized the necessity of additional information to overcome the non-uniqueness of solutions.

Full-waveform inversion (FWI) is a data-driven optimization problem which can be used to reconstruct the subsurface parameters, including anisotropic properties (Pratt et al., 2001, 2008; Prieux et al., 2011; Plessix and Cao, 2011; Gholami et al., 2013b,a; Alkhalifah and Plessix, 2014; Cheng et al., 2016; Rusmanugroho et al., 2017). In order to update anisotropic properties of a medium, FWI requires a suitable parametrization to resolve the ambiguity between parameters classes defining the anisotropy (Plessix and Cao, 2011; Gholami et al., 2013b,a). There are different applications of FWI in anisotropic media which only consider the anisotropy in waveform modelling and not inversion. As an example for multi-parameter anisotropic FWI, Gholami et al. (2013a) applied a 2D VTI acoustic FWI on synthetic and real wide-aperture data to invert jointly for vertical and horizontal velocities. When the observed data includes a wide range of azimuths and offsets, the footprint of anisotropy is more significant. Rusmanugroho et al. (2017) adapted a 2D TTI FWI where they use Voigt parametrization. A synthetic example shows that, without priory geological constraints, this approach recovers the tilt angle and anisotropic properties of the subsurface model.

Compared to the WEMVA and FWI methods, the lower computational cost and easier implementation of traveltime tomographic approaches make them still attractive for subsurface

parameter estimations. Among tomographic methods, slope tomography is a popular approach where the continuous horizon picking in conventional tomography is replaced with local coherent event picking (Rieber, 1936; Riabinkin, 1957; Sword, 1987; Billette, 1998). Extension of stereotomography (Billette, 1998), as a slope tomographic approach, to anisotropic media is introduced by Nag et al. (2006); Barbosa et al. (2006). These approaches rely on anisotropic ray tracing and the explicit building of Fréchet derivative matrix in forward and inverse problem, respectively. The main difference between these methods is their approach in mitigating the ill-posedness of the inverse problem. Nag et al. (2006, 2010) reduce the size of the model and data spaces by fixing the source and receiver depths (constant Z stereotomography), and Barbosa et al. (2006, 2008) parametrize the subsurface model by square of vertical, horizontal and diagonal velocity for an elliptic/anelliptic media. For these approaches, like other tomographic based methods, the joint inversion of all the subsurface parameters without extra information, like well logs or transmission arrivals, can not result in trustworthy solutions for the subsurface parameters. For example, Begat et al. (2008) proposed a joint slope tomographic approach where they use transmission arrivals, as well as picks from reflection data, to resolve the ambiguity between ϵ, δ and vertical velocity. There are few references in the literature about application of these methods on real data sets, although, oil industries has developed a strong practical expertise on anisotropic slope tomography.

Guillaume et al. (2008); Lambaré et al. (2008); Adler et al. (2008) introduced a non-linear slope tomographic method in which the picking procedure is performed in pre-stack depth migrated (PreSDM) or pre-stack time migrated (PreSTM) domains and the model updates are achieved by a non-linear tomographic inversion. This approach removes the problem of several migrations in MVA based methods, and compared to the classic slope tomography, provides the inversion with more reliable picks. Lambaré et al. (2009) developed this method to 3D velocity model building for time imaging where they use a general NMO equation (Castle, 1994) in time migration and through that update the RMS velocity and η in VTI media. Messud et al. (2015) extended this approach to 3D orthorhombic media and estimated the azimuthal velocity for time imaging. They parametrized the model space by RMS velocity and three parameters describing the ellipticity and anellipticity of the hyperbolas due to anisotropy. They built the Fréchet derivative matrix through the paraxial ray tracing and used a quasi-Newton approach as the optimization technique.

We reformulated the classic stereotomography by using a factored eikonal solver (Fomel et al., 2009) as a forward modelling engine and a matrix free formulation based upon the adjoint-state method (Plessix, 2006) for inversion. This formulation is the extension of our former study in isotropic adjoint slope tomography (Tavakoli F. et al., 2017b). The eikonal solver removes the well-known limitations of ray tracing (like non uniform sampling and shadow zones) and provides a uniform sampled traveltime map. In order to calculate the traveltimes, we solved the eikonal equation for the injection points at the source and receiver positions and sampled the resulted traveltime maps at the scatterer positions. We utilized source-scatterer and receiver-scatterer reciprocities to derive slopes at the source and receiver positions by finite difference of the traveltime maps associated with neighbouring sources and receivers. Therefore, the number of forward problem scales to the number of sources and receivers rather than scatterers. These forward and inverse problems result in a model space including only velocity and scatterer positions, and a data space consisting of two-way traveltimes and slopes at the source and

receiver positions. This smaller model space and data space may mitigate the ill-posedness of inverse problem. Here we extend this formulation to anisotropic media by applying a TTI factored eikonal solver (Tavakoli F. et al., 2015; Waheed et al., 2015) in forward modelling and solving the associated adjoint state equations for gradient calculations.

In the following, we first address the bases of classic anisotropic stereotomography and then we introduce the anisotropic formulation for the adjoint slope tomography. We evaluate our formulation with two synthetic examples, and in the final section we conclude with some important discussions such as the parameter cross-talks, sensitivity of the anisotropic inversion to the offset illumination and the subsurface parametrization.

3.3.2 Anisotropic extension of classic stereotomography

Nag et al. (2006); Barbosa et al. (2006) introduced extensions of the classic stereotomography to 2D VTI and general anisotropic media, respectively. They preserve the ascending ray based formulation of the classic approach and modify the forward modelling and inversion as we shall explain in the following. In these approaches the forward modelling is performed by anisotropic ray tracing and the inverse problem relies on a matrix-based formulation based upon the explicit building of the Fréchet derivative matrix through paraxial ray tracing. Introducing anisotropy to the classic formulation leads to a more ill-posed inverse problem. In the following, we briefly explain the strategies proposed in these methods for handling this ill-posedness.

Barbosa et al. (2006, 2008) considered the following data space in their approach

$$\mathbf{d} = (\mathbf{x}_s, \mathbf{x}_r, T_{s,r}, p_s, p_r)_i|_{i=1}^{N}. \tag{3.76}$$

which is the same as for the isotropic formulation. Here $\mathbf{x}_s$ and $\mathbf{x}_r$ stand for the source and receiver positions and $(T_{s,r}, p_s, p_r)$ are the two way traveltime, slope at the source and slope at the receiver position for the ith scatterer, respectively. They implement the anisotropic stereotomography for a medium with elliptic and anelliptic anisotropy where they can parametrize the subsurface model based on P-wave slowness surface. This leads to the following model space

$$\mathbf{m} = \{(\mathbf{x}, \theta_s, \theta_r, T_s, T_r), \mathbf{P}\}, \tag{3.77}$$

where $(\mathbf{x}, \theta_s, \theta_r, T_s, T_r)$ stand for the scatterer positions, take-off angles for the ascending rays and their one-way traveltimes. The end points of the ascending rays define the calculated position of the sources and receivers and their corresponding slopes. The parameters $\mathbf{P} = (P_1, P_2, P_3)$ describe the subsurface model where P_1 and P_2 are the square of the horizontal and vertical velocities and P_3 indicates square of velocity along a propagation direction defined by the ellipse orientation (Barbosa et al., 2008). They concluded that this subsurface parametrization makes the inversion less prone to parameter cross-talk, that are potentially generated by the narrow azimuth illumination provided by surface seismic experiments (Barbosa et al., 2008). Barbosa et al. (2006) assessed the method by a synthetic example where they use transmission data as the complementary information.

Nag et al. (2006, 2010) introduced the constant Z stereotomography to handle the ill-posedness of TTI/VTI stereotomography. With the aim to reduce the data space in anisotropic

stereotomography, they considered no uncertainties on the depth of sources (z_s) and receivers (z_r) and set these vertical positions to those of the observed data. This defines the data space as

$$\mathbf{d} = (x_s, x_r, T_{s,r}, p_s, p_r)_i|_{i=1}^{N}, \tag{3.78}$$

where x_s and x_r stand for the horizontal coordinates of the source and receiver positions for the ith scatterer. The other parameters are the same as those used in the classic formulation of stereotomography. As the depth of the sources and receivers are not treated as observables, the end points of the ascending rays are defined by these depths rather than from the one-way traveltimes. This means there is no need to consider one-way traveltimes among the model space parameters. Therefore, the model parameters can be defined as

$$\mathbf{m} = \{(\mathbf{x}, \theta_s, \theta_r), \mathbf{m}_b\}. \tag{3.79}$$

Here $\mathbf{m}_b$ denotes the subsurface parameters in 2D TTI/VTI anisotropic media (i.e. v_v, ϵ, δ and θ) on spline nodes. In practice, in forward modelling two ascending rays are shot from each scatterer with predefined take-off angles while their end points are at the intersection with the corresponding depth level z_s and z_r. This method is only assessed with a 2D synthetic example designed as a series of independent mono-parameter inversions for updating each of the subsurface parameters while other model parameters are fixed to their true values (Nag et al., 2010).

3.3.3 Anisotropic adjoint slope tomography for 2D acoustic TTI media

Anisotropic adjoint slope tomography inherits its main structure from our previous formulation for isotropic adjoint slope tomography (Tavakoli F. et al., 2017b). Compared to the isotropic formulation, two main extensions are necessary: adapting the forward modelling engine for calculation of traveltime map in TTI media (discussed in chapter 2) and solving the adjoint state equations for TTI media. Table 3.2 presents the mathematical symbols used in this section.

In the following, we first introduce the model and data spaces. Then, we discuss the forward modelling engine and the inverse problem. In the formulation of forward and inverse problems we use Thomsen parameters (Thomsen, 1986) (i.e. v_v, ϵ, δ and θ) to represent a 2D TTI medium. Other possible parametrizations are discussed in the section 3.4.2. Through two synthetic examples we assess our method potential for subsurface parameter estimation.

Table 3.2 – Mathematical symbols.

Symbol	Description
N	Total number of scatterers.
M	Total number of cubic B-spline nodes.
$N_{\mathbf{s}}$	Number of distinct shot gathers.
$N_{\mathbf{r}}$	Number of distinct receiver gathers.
$N_{\mathbf{r}}^{s}$	Number of distinct receivers in the shot-gather s.
$N_{\mathbf{s}}^{r}$	Number of distinct sources in the receiver-gather r.
$N_{\mathbf{n}}^{s,r}$	Number of scatterers for source-receiver pair (s, r).
$n_{s,r}$	nth scatterer associated with source and receiver pair (s, r).
$\mathbf{x}_{n_{s,r}}$	Position of nth scatterer associated with source-receiver pair (s, r) .
$\mathbf{x}_s$, $\mathbf{x}_r$	Position of source s and receiver r .
$T^*_{s,r,n_{s,r}}$	Observed two-way traveltime for scatterer $n_{s,r}$.
$T_{s,r,n_{s,r}}$	Calculated two-way traveltime for scatterer $n_{s,r}$.
$p^*_{s,n_{s,r}}$	Observed local slope for scatterer $n_{s,r}$ at the position of sth source of receiver-gather r.
$p_{s,n_{s,r}}$	Calculated local slope for scatterer $n_{s,r}$ at the position of sth source of receiver-gather r.
$p^*_{r,n_{s,r}}$	Observed local slope for scatterer $n_{s,r}$ at the position of rth receiver of shot-gather s.
$p_{r,n_{s,r}}$	Calculated local slope for scatterer $n_{s,r}$ at the position of rth receiver of shot-gather s.
$(\dots)^t$	Transpose operator.
$v_v(\mathbf{x})$	Vertical velocity model on Cartesian grid.
$\epsilon(\mathbf{x}), \delta(\mathbf{x})$	Anisotropy parameters on Cartesian grid.
$\theta(\mathbf{x})$	Tilted angle for TTI media on Cartesian grid.
$\mathbf{c}_i\|_{i=1}^4$	B-spline coefficients for subsurface parameters: $\mathbf{v}_v = \mathbf{Bc_1}$, $\boldsymbol{\epsilon} = \mathbf{Bc_2}$, $\boldsymbol{\delta} = \mathbf{Bc_3}$, $\boldsymbol{\theta} = \mathbf{Bc_4}$.
$\mathbf{m} = \{\mathbf{c}_i\|_{i=1}^4, \mathbf{x}_{n_{s,r}}\}$	Adjoint slope tomography model parameters.
$\mathbf{d}$	Adjoint slope tomography observables.
$C(\mathbf{m})$	Adjoint slope tomography misfit function

Data and model spaces definition

The definition of the data space in anisotropic adjoint slope tomography is the same as for the isotropic formulation (chapter 3.2)

$$\mathbf{d} = \left(T_{s,r,n_{s,r}}, p_{s,n_{s,r}}, p_{r,n_{s,r}}\right) |_{s=1}^{N_{\mathbf{s}}} |_{r=1}^{N_{\mathbf{r}}^{\mathbf{s}}} |_{n_{s,r}=1}^{N_{\mathbf{n}}^{s,r}}, \tag{3.80}$$

where $T_{s,r,n_{s,r}}, p_{s,n_{s,r}}$ and $p_{r,n_{s,r}}$ respectively stand for the two-way traveltime and slopes at the source s and receiver r for the nth scatterer associated with pair (s, r).

We define the model space $\mathbf{m}$ as

$$\mathbf{m} = \left(\{c_{i,m}\}|_{i=1}^{4}|_{m=1}^{M}, \{\mathbf{x}_{n_{s,r}}|_{s=1}^{N_\mathbf{s}}|_{r=1}^{N_\mathbf{r}^s}|_{n_{s,r}=1}^{N_\mathbf{n}^{s,r}}\} \right)^t, \tag{3.81}$$

where $c_{i,m}$ for $i = 1 \rightarrow 4$ represent respectively the cubic cardinal B-spline coefficients for the v_v, ϵ, δ and tilted angle θ. This model space definition is the direct result of using an eikonal solver as forward engine, where we do not need to calculate explicitly the polarization vectors for anisotropic media.

Forward problem

Forward modelling in anisotropic adjoint slope tomography is based on the same principles as the isotropic adjoint slope tomography (section 3.2.4). It only requires a modification in traveltime calculation where we replace the isotropic eikonal solver with our TTI eikonal solver introduced in chapter 2. In our eikonal solvers, the first arrival traveltime maps are solution of a PDE constrained by a Dirichlet boundary condition at the injection point. For each scatterer we solve two TTI eikonal equations with the associated source and receiver as the injection points

$$\begin{cases} H(\mathbf{x}, \nabla t_s(\mathbf{x})) = 0, \\ t_s(\mathbf{x}_s) = 0, \end{cases} \tag{3.82}$$

and

$$\begin{cases} H(\mathbf{x}, \nabla t_r(\mathbf{x})) = 0, \\ t_r(\mathbf{x}_r) = 0, \end{cases} \tag{3.83}$$

where $t_s(\mathbf{x})$ and $t_r(\mathbf{x})$ are first-arrival traveltime maps originated from source s and receiver r, respectively. Here H stands for the Hamiltonian representation of 2D TTI eikonal equation (Alkhalifah, 2003; Waheed et al., 2014),

$$H(\mathbf{x}, \nabla t(\mathbf{x})) = A(\nabla_{\hat{x}} t(\mathbf{x}))^2 + C(\nabla_{\hat{z}} t(\mathbf{x}))^2 + E(\nabla_{\hat{x}} t(\mathbf{x}))^2 (\nabla_{\hat{z}} t(\mathbf{x}))^2 - 1. \tag{3.84}$$

Here $\widehat{\nabla t(\mathbf{x})} = (\nabla_{\hat{x}} t(\mathbf{x}), \nabla_{\hat{z}} t(\mathbf{x}))$ is the spatial derivative of the traveltime map $t(\mathbf{x})$ in the 2D local rotated coordinate system defined by tilted angle map θ for 2D TTI media. In other word, by introducing the rotation operator $\mathcal{R}$,

$$\widehat{\nabla t(\mathbf{x})} = \begin{pmatrix} \nabla_{\hat{x}} t(\mathbf{x}) \\ \\ \nabla_{\hat{z}} t(\mathbf{x}) \end{pmatrix} = \mathcal{R} \nabla t(\mathbf{x}) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \\ sin\ \theta & \cos\theta \end{pmatrix} \begin{pmatrix} \nabla_x t(\mathbf{x}) \\ \\ \nabla_z t(\mathbf{x}) \end{pmatrix}. \tag{3.85}$$

Therefore, we can rewrite the Hamiltonian (3.84) in a matricial form as

$$H(\mathbf{x}, \nabla t(\mathbf{x})) = A((\mathcal{R}\nabla \mathbf{t})_x)^2 + C((\mathcal{R}\nabla \mathbf{t})_z)^2 + E((\mathcal{R}\nabla \mathbf{t})_x (\mathcal{R}\nabla \mathbf{t})_z)^2 - 1. \tag{3.86}$$

Here the subscripts x and z indicate the first and the second elements of the vector $\mathcal{R}\nabla\mathbf{t}$, respectively, and the coefficients A, C and E are defined as (Thomsen, 1986; Tsvankin, 1997)

$$\begin{aligned} A =& v_v^2(1+2\epsilon) = v_h^2, \\ C =& v_v^2 = \frac{v_{nmo}}{1+2\delta}, \\ E =& -2v_v^4(\epsilon-\delta) = -2v_v^2 v_{nmo}^2\eta, \end{aligned} \tag{3.87}$$

where the anelliptic parameter η is defined as

$$\eta = \frac{\epsilon-\delta}{1+2\delta}. \tag{3.88}$$

Here for the sake of brevity we omit the coordinate dependency of the parameters. We compute the traveltime maps with an iterative factored eikonal solver based on the fast sweeping method (FSM) (Zhao, 2005) which is introduced in detail in chapter 2 and Tavakoli F. et al. (2015). Although in tomographic methods we deal with smooth velocity models, this forward modelling engine can calculate precisely traveltimes in geologically complex models. The extension of this forward modelling engine to 3D TTI models is presented by Waheed et al. (2015).

After solving eqs. (3.82) and (3.83) for source s and receiver r, we calculate two-way traveltime $T_{s,r,n_{s,r}}$ for the nth scatterer of pair (s,r) by

$$T_{s,r,n_{s,r}} = Q_{n_{s,r}}(\mathbf{t}_s + \mathbf{t}_r). \tag{3.89}$$

$Q_{n_{s,r}}$ is a sampling operator for extracting the traveltime at the scatterer positions, $\mathbf{x}_{n_{s,r}}$. Here we use windowed sinc function (Hicks, 2002) while other sampling operators also can be considered.

In order to calculate the slope at the source and receiver positions, we apply the same approach that used in the isotropic case (section 3.2.4). Considering the reciprocity between the source-scatterer and receiver-scatterer, we design a finite difference operator calculating the horizontal component of the slowness vector at the source and receiver positions. For scatterer $n_{s,r}$ the inputs of this operator are traveltime maps from neighbouring sources of the source s and neighbouring receivers of receiver r. This gives the following expressions for the slope at source ($p_{s,n_{s,r}}$) and receiver ($p_{r,n_{s,r}}$)

$$p_{s,n_{s,r}} = \frac{\partial t_s(\mathbf{x}_s)}{\partial x_s} \approx Q_{n_{s,r}}(\mathbf{t}_{s+1} - \mathbf{t}_{s-1})/2\Delta s, \tag{3.90}$$

$$p_{r,n_{s,r}} = \frac{\partial t_r(\mathbf{x}_r)}{\partial x_r} \approx Q_{n_{s,r}}(\mathbf{t}_{r+1} - \mathbf{t}_{r-1})/2\Delta r. \tag{3.91}$$

Here Δs and Δr are source and receiver intervals, respectively. Thanks to the reciprocity rule, the finite difference operator samples the traveltime maps far from the singular points of the eikonal equation (source and receiver positions). The smooth velocity in the background (as long as there is no multi-pathing) ensures a trustworthy approximation for slopes at the source and receiver positions. Through a test in Appendix C, we also investigate the effect of

the source/receiver interval lengths on the precision of slope estimation with eqs (3.90)-(3.91) where we conclude if the subsurface model be smooth, the effect of increasing these intervals (in a reasonable range such as $\Delta s, \Delta r < 500$ m) is negligible.

This finite-difference estimation of the slopes provides some qualitative insight on the role of the slopes in the positioning of the scatterers. The difference between traveltimes from the neighbouring sources and receivers at the scatterer position provides some constraints on the dip of the so-called migration facet (or dip bar). The dip of these migration facets constrains in turn the lateral positions of the scatterers, which is poorly constrained by traveltime only (as any scatterer along an isochrone surface will honour the same traveltime). We discussed this issue within an example in section 3.2.5. The finite difference approximation of slopes brings to light the relationship between adjoint slope tomography and double-difference tomography (Yuan et al., 2016).

Multi-parameter Inverse problem

We consider a l^2-norm misfit function between observed, denoted by symbol *, and calculated data

$$\begin{aligned} C(\mathbf{m}) &= \frac{1}{2\sigma^2_{T_{s,r}}} \sum_{s=1}^{N_\mathbf{s}} \sum_{r=1}^{N^s_\mathbf{r}} \sum_{n_{s,r}=1}^{N^{s,r}_\mathbf{n}} (T_{s,r,n_{s,r}}(\mathbf{m}) - T^*_{s,r,n_{s,r}})^2 \\ &+ \frac{1}{2\sigma^2_{p_s}} \sum_{s=1}^{N_\mathbf{s}} \sum_{r=1}^{N^s_\mathbf{r}} \sum_{n_{s,r}=1}^{N^{s,r}_\mathbf{n}} (p_{s,n_{s,r}}(\mathbf{m}) - p^*_{s,n_{s,r}})^2 \\ &+ \frac{1}{2\sigma^2_{p_r}} \sum_{r=1}^{N_\mathbf{r}} \sum_{s=1}^{N^r_\mathbf{s}} \sum_{n_{s,r}=1}^{N^{s,r}_\mathbf{n}} (p_{r,n_{s,r}}(\mathbf{m}) - p^*_{r,n_{s,r}})^2, \end{aligned} \tag{3.92}$$

where $\sigma^2_{T_{s,r}}, \sigma^2_{p_s}$, and $\sigma^2_{p_r}$ are the elements of a diagonal covariance matrix, which weight the relative contribution of each data classes in the misfit function (Tarantola, 1987). In this study we do not consider any regularization term in the misfit function. We use a Newton-based local optimization algorithm to update iteratively the model space,

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \Big(\frac{\partial^2 C}{\partial \mathbf{m}^2}(\mathbf{m}_k) \Big)^{-1} \frac{\partial C}{\partial \mathbf{m}}(\mathbf{m}_k). \tag{3.93}$$

Here the search direction $\Big(\frac{\partial^2 C}{\partial \mathbf{m}^2}(\mathbf{m}_k) \Big)^{-1} \frac{\partial C}{\partial \mathbf{m}}(\mathbf{m}_k)$ is the *Newton direction*, and a line search method satisfying the Wolfe conditions is used to set the value of the step length $\alpha_k \in \mathbb{R}_+$. We use the L-BFGS method (Byrd et al., 1995), implemented in the SEISCOPE optimization toolbox (Métivier and Brossier, 2016), to approximate the Hessian operator ($\frac{\partial^2 C}{\partial \mathbf{m}^2}$) and find the minimizer of the misfit function C. This method belongs to the family of quasi-Newton methods where, instead of computing explicitly the $n \times n$ Hessian matrix, we recursively estimate it from a few vectors of length n (Nocedal and Wright, 2006).

In this multi-parameter inversion, the different orders of magnitude of the different parameter classes make the optimization poorly scaled. In general, some optimization methods like Newton's algorithms are less sensitive to poor scaling while algorithms like steepest descent are highly sensitive (Nocedal and Wright, 2006). As in the first iteration of L-BFGS method the initial value for Hessian is identity (i.e. the search direction is steepest descent), this method is also sensitive to poor scaling. As a remedy, we use a constant scaling factor for ϵ, δ and theta to bring them to the same order of magnitude of v_v and scatterer positions. Without this scaling the optimization is highly prone to converge toward a local minimum.

Gradient computation with the adjoint-state method

In order to calculate the gradient of the misfit function with respect to the model parameters we use the adjoint-state method (section 3.1.5) which provides a matrix-free formulation for the gradient calculation without explicit building of the sensitivity matrix. This property makes the adjoint-state method well suited for large-scale inverse problems at the expense of restriction on optimization method selection; one can only apply the optimization approaches which do not require the sensitivity matrix (such as Quasi-Newton optimization methods) (Chavent, 2009). We use the Lagrangian formalism of the adjoint state method to impose the constraints (i.e. eqs (3.82),(3.83), (3.89)-(3.91)) to the misfit function C. This leads to the Lagrangian definition as

$$
\begin{aligned}
\mathcal{L}(\mathbf{m},\mathbf{u},\bar{\mathbf{u}}) \;=\; & \mathcal{H}(\mathbf{u},\mathbf{m}) \\
& - \sum_{s=1}^{N_\mathbf{s}} \sum_{r=1}^{N_\mathbf{r}^s} \sum_{n_{s,r}=1}^{N_\mathbf{n}^{s,r}} \mu_{s,r,n_{s,r}} \left(T_{s,r,n_{s,r}} - Q_{n_{s,r}}(\mathbf{t}_s + \mathbf{t}_r) \right) \\
& - \sum_{s=1}^{N_\mathbf{s}} \sum_{r=1}^{N_\mathbf{r}^s} \sum_{n_{s,r}=1}^{N_\mathbf{n}^{s,r}} \xi_{s,n_{s,r}} \left(p_{s,n_{s,r}} - Q_{n_{s,r}}(\mathbf{t}_{s+1} - \mathbf{t}_{s-1})/2\Delta s \right) \\
& - \sum_{r=1}^{N_\mathbf{r}} \sum_{s=1}^{N_\mathbf{s}^r} \sum_{n_{s,r}=1}^{N_\mathbf{n}^{s,r}} \xi_{r,n_{s,r}} \left(p_{r,n_{s,r}} - Q_{n_{s,r}}(\mathbf{t}_{r+1} - \mathbf{t}_{r-1})/2\Delta r \right) \\
& - \frac{1}{2} \sum_{s=1}^{N_\mathbf{s}} \Big\langle \lambda_s(\mathbf{x}) \mid H(\mathbf{x}, \nabla t_s(\mathbf{x}) \Big\rangle_\Omega - \frac{1}{2} \sum_{r=1}^{N_\mathbf{r}} \Big\langle \lambda_r(\mathbf{x}) \mid H(\mathbf{x}, \nabla t_r(\mathbf{x}) \Big\rangle_\Omega .
\end{aligned}
\tag{3.94}
$$

Here Ω stands for the subsurface domain. The Lagrange multipliers or adjoint state variables and the state variables are gathered in vector $\bar{\mathbf{u}}$ and $\mathbf{u}$, respectively, as

$$
\bar{\mathbf{u}} = (\mu_{s,r,n_{s,r}}, \xi_{s,n_{s,r}}, \xi_{r,n_{s,r}}, \boldsymbol{\lambda}_s, \boldsymbol{\lambda}_r), \qquad \mathbf{u} = (T_{s,r,n_{s,r}}, p_{s,n_{s,r}}, p_{r,n_{s,r}}, \mathbf{t}_s, \mathbf{t}_r). \tag{3.95}
$$

Because the state variables are considered independent of the model parameters, we rewrote the misfit function $C(\mathbf{m})$ as $\mathcal{H}(\mathbf{u},\mathbf{m})$ in the Lagrangian. Here the unknowns are the adjoint state variables which can be calculated by zeroing the partial derivative of the Lagrangian with

respect to the state variables. Here is the solution of $\mu_{s,r,n_{s,r}}$, $\xi_{s,n_{s,r}}$ and $\xi_{r,n_{s,r}}$:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial T_{s,r,n_{s,r}}} = 0 \quad &\rightarrow \quad \mu_{s,r,n_{s,r}} = \frac{1}{\sigma^2_{T_{s,r}}}(T_{s,r,n_{s,r}} - T^*_{s,r,n_{s,r}}) = \frac{\Delta T_{s,r,n_{s,r}}}{\sigma^2_{T_{s,r}}}, \\
\frac{\partial \mathcal{L}}{\partial p_{s,n_{s,r}}} = 0 \quad &\rightarrow \quad \xi_{s,n_{s,r}} = \frac{1}{\sigma^2_{p_s}}(p_{s,n_{s,r}} - p^*_{s,n_{s,r}}) = \frac{\Delta p_{s,n_{s,r}}}{\sigma^2_{p_s}}, \\
\frac{\partial \mathcal{L}}{\partial p_{r,n_{s,r}}} = 0 \quad &\rightarrow \quad \xi_{r,n_{s,r}} = \frac{1}{\sigma^2_{p_r}}(p_{r,n_{s,r}} - p^*_{r,n_{s,r}}) = \frac{\Delta p_{r,n_{s,r}}}{\sigma^2_{p_r}}.
\end{aligned}$$

The two other adjoint state variables λ_s and λ_r respectively are solution of the adjoint equations $\partial\mathcal{L}/\partial\mathbf{t}_s = 0$ and $\partial\mathcal{L}/\partial\mathbf{t}_r = 0$. Gradient of Lagrangian (3.94) with respect to $\mathbf{t}_s$ is given by

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mathbf{t}_s} = &\sum_{r=1}^{N_\mathbf{r}^s} \sum_{n_{s,r}=1}^{N_\mathbf{n}^{s,r}} Q^t_{n_{s,r}} \mu_{s,r,n_{s,r}} \\
&- \frac{1}{2\Delta s} \sum_{r=1}^{N_\mathbf{r}^{s+1}} \sum_{n_{s+1,r}=1}^{N_\mathbf{n}^{s+1,r}} Q^t_{n_{s+1,r}} \xi_{s+1,n_{s+1,r}} \\
&+ \frac{1}{2\Delta s} \sum_{r=1}^{N_\mathbf{r}^{s-1}} \sum_{n_{s-1,r}=1}^{N_\mathbf{n}^{s-1,r}} Q^t_{n_{s-1,r}} \xi_{s-1,n_{s-1,r}} \\
&+ \sum_{s=1}^{N_\mathbf{s}} (\nabla \cdot (\lambda_s(\mathbf{x})\, \mathcal{U}_s))_\Omega + \sum_{s=1}^{N_\mathbf{s}} (\lambda_s(\mathbf{x})\, \mathcal{U}_s.\mathbf{n})_\Gamma,
\end{aligned} \tag{3.96}$$

where Γ denotes the boundaries of Ω and $\mathcal{U}_s = \nabla_{\nabla t_s(\mathbf{x})} H(\mathbf{x}, t_s(\mathbf{x}))$ and $\mathcal{U}_r = \nabla_{\nabla t_r(\mathbf{x})} H(\mathbf{x}, t_r(\mathbf{x}))$ are the group velocity vectors along the ray-paths initiated from source s and receiver r, respectively. In order to derive the last two terms in eq. (3.96), we use an integration by part as reviewed in Appendix A. Considering a Dirichlet boundary condition as $\lambda_s(\mathbf{x})\mathcal{U}_s \cdot \mathbf{n} = 0$ over Γ we obtain

$$\begin{aligned}
(\nabla \cdot (\lambda_s(\mathbf{x})\, \mathcal{U}_s))_\Omega = &- \sum_{r=1}^{N_\mathbf{r}^s} \sum_{n_{s,r}=1}^{N_\mathbf{n}^{s,r}} Q^t_{n_{s,r}} \mu_{s,r,n_{s,r}} \\
&+ \frac{1}{2\Delta s} \sum_{r=1}^{N_\mathbf{r}^{s+1}} \sum_{n_{s+1,r}=1}^{N_\mathbf{n}^{s+1,r}} Q^t_{n_{s+1,r}} \xi_{s+1,n_{s+1,r}} \\
&- \frac{1}{2\Delta s} \sum_{r=1}^{N_\mathbf{r}^{s-1}} \sum_{n_{s-1,r}=1}^{N_\mathbf{n}^{s-1,r}} Q^t_{n_{s-1,r}} \xi_{s-1,n_{s-1,r}}.
\end{aligned} \tag{3.97}$$

Similarly the adjoint state variable λ_r satisfies

$$
\begin{aligned}
(\nabla \cdot (\lambda_r(\mathbf{x})\, \mathcal{U}_r))_\Omega = &- \sum_{s=1}^{N_\mathbf{s}^r} \sum_{n_{s,r}=1}^{N_\mathbf{n}^{s,r}} Q^t_{n_{s,r}} \mu_{s,r,n_{s,r}} \\
&+ \frac{1}{2\Delta r} \sum_{s=1}^{N_\mathbf{s}^{r+1}} \sum_{n_{s,r+1}=1}^{N_\mathbf{n}^{s,r+1}} Q^t_{n_{s,r+1}} \xi_{r+1,n_{s,r+1}} \\
&- \frac{1}{2\Delta r} \sum_{s=1}^{N_\mathbf{s}^{r-1}} \sum_{n_{s,r-1}=1}^{N_\mathbf{n}^{s,r-1}} Q^t_{n_{s,r-1}} \xi_{r-1,n_{s,r-1}}.
\end{aligned} \tag{3.98}
$$

We use a conservative finite difference operator [4] alongside the FSM to solve the linear eqs (3.97) and (3.98). The right hand sides are defined at the scatterer positions and solution of these equations represents a flow directing from the scatterers toward the sources and receivers. In similar fashion to the isotropic formulation (eqs 3.70 and 3.71) or adjoint first arrival traveltime tomography (Taillandier et al., 2009), here we solve a transport like equation but for 2D TTI media which requires designing a suitable conservative finite difference operator (details in Appendix B). The role of the right hand side source terms is fully explained in section 3.2.4.

The gradient of the misfit function, or equivalently the gradient of the Lagrangian (3.94), with respect to the subsurface parameters is the weighted summation of the adjoint fields λ_s and λ_r:

$$
\nabla_{v_v(\mathbf{x})} C = - \sum_{s=1}^{N_\mathbf{s}} \lambda_s(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_s(\mathbf{x}))}{2\, \partial v_v(\mathbf{x})} - \sum_{r=1}^{N_\mathbf{r}} \lambda_r(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_r(\mathbf{x}))}{2\, \partial v_v(\mathbf{x})}, \tag{3.99}
$$

$$
\nabla_{\epsilon(\mathbf{x})} C = - \sum_{s=1}^{N_\mathbf{s}} \lambda_s(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_s(\mathbf{x}))}{2\, \partial \epsilon(\mathbf{x})} - \sum_{r=1}^{N_\mathbf{r}} \lambda_r(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_r(\mathbf{x}))}{2\, \partial \epsilon(\mathbf{x})}, \tag{3.100}
$$

$$
\nabla_{\delta(\mathbf{x})} C = - \sum_{s=1}^{N_\mathbf{s}} \lambda_s(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_s(\mathbf{x}))}{2\, \partial \delta(\mathbf{x})} - \sum_{r=1}^{N_\mathbf{r}} \lambda_r(\mathbf{x}) \frac{2\, \partial H(\mathbf{x}, \nabla t_r(\mathbf{x}))}{2\, \partial \delta(\mathbf{x})}, \tag{3.101}
$$

$$
\nabla_{\theta(\mathbf{x})} C = - \sum_{s=1}^{N_\mathbf{s}} \lambda_s(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_s(\mathbf{x}))}{2\, \partial \theta(\mathbf{x})} - \sum_{r=1}^{N_\mathbf{r}} \lambda_r(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_r(\mathbf{x}))}{2\, \partial \theta(\mathbf{x})}. \tag{3.102}
$$

The gradients of the Hamiltonian $H(\mathbf{x}, \nabla t(\mathbf{x}))$, for a given traveltime map $t(\mathbf{x})$, with respect to the subsurface parameters can be calculated as follow

$$
\frac{\partial H(\mathbf{x}, \nabla t(\mathbf{x}))}{\partial v_v(\mathbf{x})} = 2v_v(1+2\epsilon)((\mathcal{R}\nabla t)_x)^2 + 2v_v((\mathcal{R}\nabla t)_z)^2 - 8v_v^3(\epsilon-\delta)((\mathcal{R}\nabla t)_x(\mathcal{R}\nabla t)_z)^2, \tag{3.103}
$$

$$
\frac{\partial H(\mathbf{x}, \nabla t(\mathbf{x}))}{\partial \epsilon(\mathbf{x})} = 2v_v^2((\mathcal{R}\nabla t)_x)^2 - 2v_v^4((\mathcal{R}\nabla t)_x(\mathcal{R}\nabla t)_z)^2, \tag{3.104}
$$

4. Conservative finite difference operators honour the conservation law.

$$\frac{\partial H(\mathbf{x},\nabla t(\mathbf{x}))}{\partial \delta(\mathbf{x})} = 2v_v^4((\mathcal{R}\nabla t)_x(\mathcal{R}\nabla t)_z)^2, \tag{3.105}$$

$$\begin{aligned}\frac{\partial H(\mathbf{x},\nabla t(\mathbf{x}))}{\partial \theta(\mathbf{x})} &= 2A(\mathcal{R}\nabla t)_x(\mathcal{R}'\nabla t)_x + 2C(\mathcal{R}\nabla t)_z(\mathcal{R}'\nabla t)_z \\ &+ 2E(\mathcal{R}\nabla t)_x(\mathcal{R}\nabla t)_z.\Big((\mathcal{R}'\nabla t)_x(\mathcal{R}\nabla t)_z + (\mathcal{R}\nabla t)_x(\mathcal{R}'\nabla t)_z\Big),\end{aligned} \tag{3.106}$$

where $\mathcal{R}'$ is the gradient of rotation matrix with respect to θ and the index $(\dots)_x$ and $(\dots)_z$ designate the first and second elements of the related vector. We omit the coordinate dependency of traveltime maps for the sake of brevity. In section 3.3.4 through a synthetic example, we illustrate the role of these weights on the shape of gradients and the sensitivity of the gradients to offset (reflection angle) variations. In Appendix D, we validate the accuracy of calculated gradients with the adjoint state method against the finite difference approach.

The dominant terms of the weights (3.103)-(3.106) shape the sensitivity kernel of the associated gradients in eqs (3.99)-(3.102). For weight (3.103) associated with $\nabla_{v(\mathbf{x})}C$ in eq. (3.103), two first terms are dominant and this makes the gradient with respect to v_v sensitive to both vertical and horizontal propagations. In contrast, $\nabla_{\epsilon(\mathbf{x})}C$ is mostly sensitive to the horizontal paths because the dominant term in the associated weight (3.104) is tied to the horizontal derivatives of traveltimes. For $\nabla_{\epsilon(\mathbf{x})}C$ the highest amplitude for sensitivity kernel is for the oblique propagations where the single term in the weight (3.105) is maximum.

Remember, in the model space definition, expression (3.81), the subsurface parameters are defined on the spline nodes and therefore the final step in the gradient calculations is projection of the calculated gradients on the Cartesian grid, eqs (3.99)-(3.102), to cubic B-spline bases, $\mathbf{c}_i = c_{i,m}|_{m=1}^{M}|_{i=1}^{4}$. This can be achieved by applying the chain rule of derivatives

$$\nabla_{\mathbf{c}_1}C = \mathbf{B}^t\,\nabla_{\mathbf{v}_v}C, \tag{3.107}$$

$$\nabla_{\mathbf{c}_2}C = \mathbf{B}^t\,\nabla_{\boldsymbol{\epsilon}}C, \tag{3.108}$$

$$\nabla_{\mathbf{c}_3}C = \mathbf{B}^t\,\nabla_{\boldsymbol{\delta}}C, \tag{3.109}$$

$$\nabla_{\mathbf{c}_4}C = \mathbf{B}^t\,\nabla_{\boldsymbol{\theta}}C, \tag{3.110}$$

where "$\mathbf{B}$" stands for the cubic cardinal B-spline operator ($\mathbf{v}_v = \mathbf{B}\mathbf{c_1}$, $\boldsymbol{\epsilon} = \mathbf{B}\mathbf{c_2}$, $\boldsymbol{\delta} = \mathbf{B}\mathbf{c_3}$, $\boldsymbol{\theta} = \mathbf{B}\mathbf{c_4}$) and $(\cdots)^t$ denotes the adjoint operator.

The gradient with respect to the scatterer coordinates is given by

$$\nabla_{\mathbf{x}_{n_{s,r}}}C = \frac{Q_{n_{s,r}}}{\partial \mathbf{x}_{n_{s,r}}}\Big(\mu_{s,r,n_{s,r}}(\mathbf{t}_s+\mathbf{t}_r) + \frac{\xi_{s,n_{s,r}}}{2\Delta s}(\mathbf{t}_{s+1}-\mathbf{t}_{s-1}) + \frac{\xi_{r,n_{s,r}}}{2\Delta r}(\mathbf{t}_{r+1}-\mathbf{t}_{r-1})\Big). \tag{3.111}$$

Here the term $\partial Q_{n_{s,r}}/\partial \mathbf{x}_{n_{s,r}}$ is calculated by analytical derivative of windowed sinc function (Hicks, 2002). As mentioned before one can choose other simpler sampling operators at the expense of accuracy of the operator gradient. Direct contribution of the traveltime maps of neighbouring sources and receivers is the result of our finite difference scheme in slope calculation. The different steps required to compute the gradients are outlined in Algorithm 3.

Algorithm 5: Gradient algorithm

$\mathbf{v}_v = \mathcal{B}\ \mathbf{c}_1$
$\epsilon = \mathcal{B}\ \mathbf{c}_2$
$\delta = \mathcal{B}\ \mathbf{c}_3$
$\theta = \mathcal{B}\ \mathbf{c}_4$
for $s, r = 1$ to $N_\mathbf{s}$,$N_\mathbf{r}$ **do**
    Compute $\mathbf{t}_s(\mathbf{x})$, $\mathbf{t}_r(\mathbf{x})$
end for
for $s, r = 1$ to $N_\mathbf{s}$, $N_\mathbf{r}$ **do**
    for $r, s = 1$ to $N^s_\mathbf{r}$, $N^r_\mathbf{s}$ **do**
        for $n_{s,r} = 1$ to $N^{s,r}_\mathbf{n}$ **do**
            Compute $T_{s,n_{s,r}}$, $p_{s,n_{s,r}}$, $p_{r,n_{s,r}}$
        end for
    end for
end for
Compute misfit function C
for $s, r = 1$ to $N_\mathbf{s}$,$N_\mathbf{r}$ **do**
    Compute λ_s, λ_r
end for
Compute $\nabla_{\mathbf{v}_v} C$, $\nabla_\epsilon C$, $\nabla_\delta C$, $\nabla_\theta C$
Compute $\nabla_\mathbf{x} C$
Compute $\nabla_{\mathbf{c}_i} C|_{i=1}^4 = \mathbf{B}^t(\nabla_{\mathbf{v}_v} C, \nabla_\epsilon C, \nabla_\delta C, \nabla_\theta C)$

Implementation in practice

The implementation of the anisotropic slope tomography is similar to the strategy we applied in isotropic formulation in section 3.2 which is based on the implementation of classic stereotomography (Billette et al., 2003). The anisotropic slope tomography includes three main steps: 1) initialization of the model parameters, 2) localization of the scatterers while subsurface parameters are fixed to their initial values, 3) multi-scaling joint inversion of the parameters. Detail of this three-steps procedure can be found in section 3.2.4. We can consider all the subsurface parameters as the optimization parameters (those we aim to update) or consider some of them as the passive parameters and keep them fixed during the inversion. During the third step, after convergence in each scale we apply the subdivision technique (Virieux and Farra (1991, their appendix)) to refine the spline grid for the subsurface optimization parameter(s) to insert shorter wavelength components of the subsurface to the inversion. This mitigates the ill-posedness of the inverse problem. Algorithm 6 summarizes the anisotropic slope tomography workflow.

3.3.4 Synthetic examples

In this section we assess the method against two synthetic examples: a circular anomaly in a homogeneous VTI background and a part of BP TTI salt model (Shah, 2007). For both tests,

Algorithm 6: Anisotropic adjoint slope tomography workflow. N_{ms}: number of multi-scale step; $\mathcal{S}$: B-spline subdivision operator; $\mathbf{M}_{final}$: the final updated subsurface optimization parameters on Cartesian grid. For sake of clarity, the model parameters $\mathbf{m}$, the subsurface optimization parameters on B-spline nodes $\mathbf{c}^{opt}$, the subsurface passive parameters on B-spline nodes $\mathbf{c}^{opt}$, the misfit function C and its gradient ∇C are indexed by the scale step i and the iteration number k. The model parameters consist of scatterer positions, subsurface passive and optimization parameters.

Initialization of scatterer positions (Billette et al., 2003, Their appendix A)
Initialize the subsurface optimization parameters on B-spline nodes $\mathbf{c}_0^{opt}$
Preliminary re-localization of scatterers in $\mathbf{c}_0^{opt}$ and $\mathbf{c}^{psv}$
for $i = 1$ to N_{ms} **do**
  $\mathbf{c}_{0,i}^{opt} = \mathcal{S}\,\mathbf{c}_{N_{it},i-1}^{opt}$
  for $k = 1$ to N_{it} **do**
    Perform forward modelling for $\mathbf{m}_{k-1,i}$
    Calculate the cost function $\rightarrow C_{k-1,i}$
    Calculate the gradients for optimization parameters (Algorithm 5) $\rightarrow \nabla C_{k-1,i}$
    call L-BFGS($\mathbf{m}_{k-1,i}$,$C_{k-1,i}$,$\nabla C_{k-1,i}$)
    $\mathbf{m}_{k,i} = \mathbf{m}_{k-1,i} + \Delta\mathbf{m}_{k-1,i}$
  end for
end for
$\mathbf{M}_{final} = \mathbf{B}_{N_{ms}}\mathbf{c}_{N_{ms},N_{it}}^{opt}$

we generate the observables with our forward modelling engine where we consider a regular distribution of the reflection/diffraction points in the true subsurface models. Compared to datasets generated by picking in the data domain, this artificial dataset provides a more uniform distribution in depth of the scatterers, and hence a better-controlled subsurface illumination. In a preliminary stage, our aim here is to mitigate the impact of the picking and the related illumination issues on the inversion results to focus on the ability of the method to update multiple subsurface parameter classes associated with anisotropy.

VTI circular anomaly

As a first example we consider a 5 km $\times$ 21 km VTI medium which includes a smooth circular anomaly for v_v, ϵ and δ. The background model is a constant-gradient model for v_v and is homogeneous for δ and ϵ (Fig. 3.22). We define 155 reflection/diffraction points on five horizontal lines with 200 m and 700 m interval in the horizontal and vertical directions, respectively. Also, there is a line of sources/receivers at 500 m depth with 200 m spacing. Each reflection/diffraction point is sampled by five source-receiver pairs with offsets 0.8, 1.6, 2.4, 3.2 and 4 km such that the midpoint of the source-receiver is aligned with the reflection/diffraction point. This results in 755 scatterers (black circles in Fig. 3.22a). Here the goal is to recover the circular anomaly of v_v, ϵ and δ starting the inversion from their true background models. The scatterer initial positions are randomly distributed around the true positions with a maximum deviation of 400 m (Fig. 3.23a). The forward modelling is perform in a finite-difference

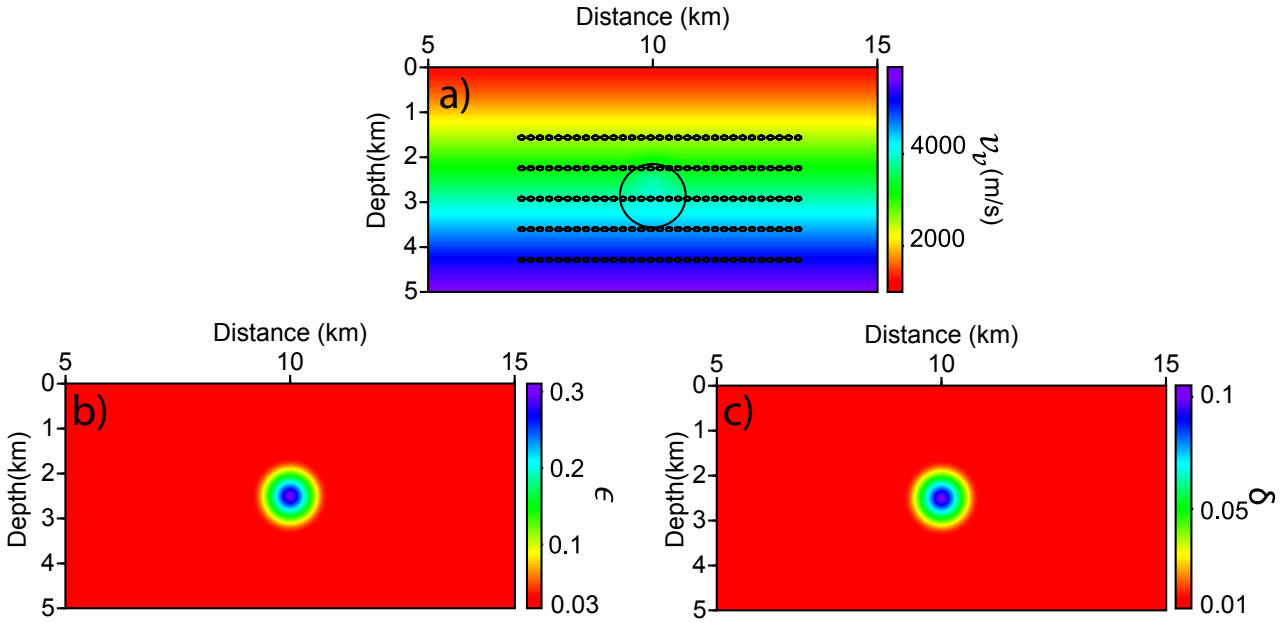

Figure 3.22 – VTI circular anomaly example. True model for a) v_v, b) ϵ, and c) δ. The big circle in a) represents the velocity anomaly position and small dots are reflection/diffraction points (the true scatterer positions).

grid with a grid interval of 50 m in both the vertical and horizontal directions. The values for $(\sigma_{T_{s,r}}, \sigma_{P_s}, \sigma_{P_s})$ are $(10^{-3}, 10^{-6}, 10^{-6})$. In this example we multiply ϵ and δ by constant scaling factor 1000 to make their order of magnitudes in the same range as v_v and scatterer positions.

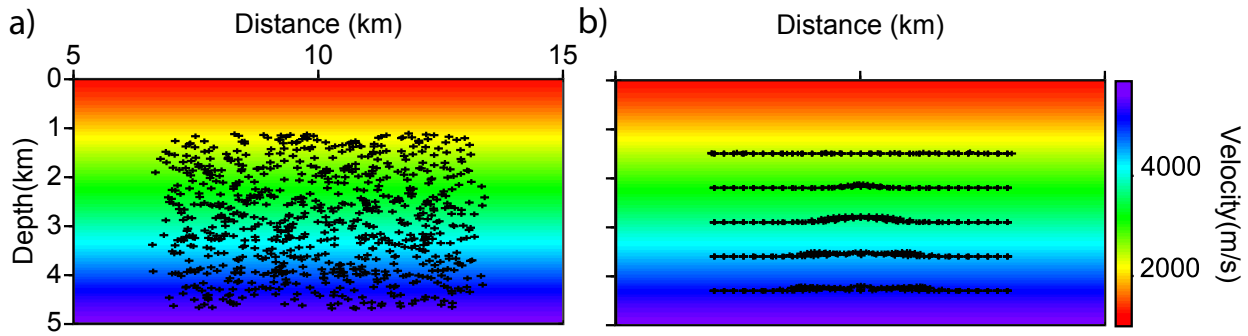

Figure 3.23 – VTI circular anomaly example. v_v and the scatterer positions after a) Initialization and b) after Localization step. The initial v_v model is the true background.

After the localization step (Fig. 3.23b) the scatterers are positioned nearby their true positions. This results because we use the true subsurface background models as the initial models. However, there are some mis-positioned scatterers on the top line because a 4km-offset source receiver pair leads to pathological scattered paths formed by two diving wave paths that pass the turning point. This means there is no reflection with offset 4 km from the top line reflec-

tion/diffraction points and consequently the calculated data for the top line scatterers with this offset are erroneous. In the current implementation of our method, these scatterers play the role of noisy scatterers. In the perspective section (section 5.2), we shall discuss about taking advantage of diving wave propagation by inserting diving wave picks among the observables. In the framework of classic formulation, Prieux et al. (2013) implemented this extension.

In Fig. 3.24(a) we show the adjoint fields λ_s and λ_r at the first iteration of the joint inversion for a scatterer located at 2.2 km depth and a source-receiver offset of 4 km. The associated gradients with respect to v_v, ϵ and δ are shown in Fig. 3.24(b-d), respectively. For the gradient with respect to v_v (Fig. 3.24b) the sensitivity for both vertical and horizontal propagations is in a same order of magnitude. Here, the strong vertical sensitivity is due to surface to depth propagation and source and receiver footprints. For the gradient with respect to ϵ (Fig. 3.24c) and δ (Fig. 3.24d) the sensitivity is for the horizontal and midrange propagation paths. These observations are consistent with the explanations of eqs (3.103)-(3.105); the gradients with respect to subsurface parameters are the weighted adjoint fields and the dominant terms in these weights shapes the sensitivity kernel associated with each parameter class.

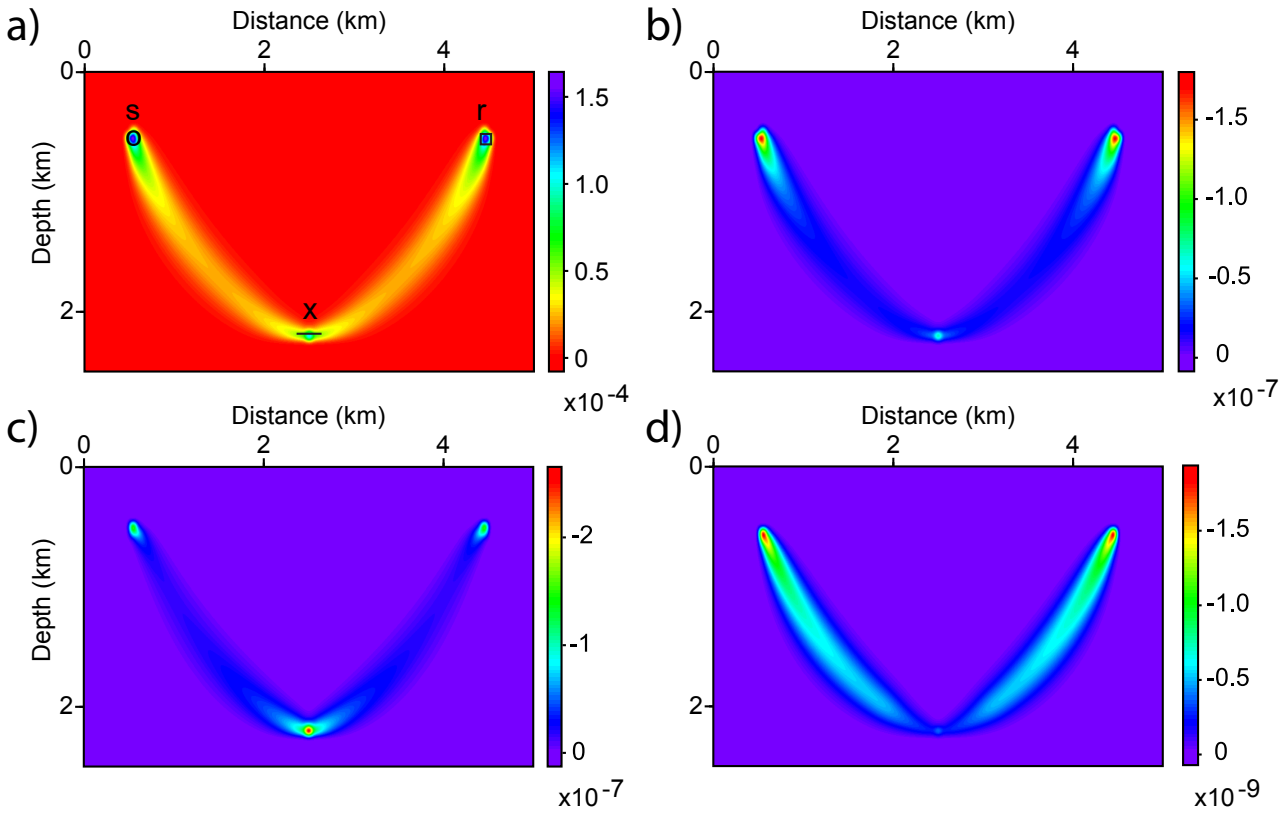

Figure 3.24 – VTI circular anomaly example. a) Adjoint fields λ_s and λ_r for one 4km-offset source-receiver pair (s, r) associated with a scatterer x at 2.2 km depth. Gradient with respect to b) v_v, c) ϵ and d) δ in first iteration. Compared to the a) and b) δ panel has amplitudes that are two order of magnitude smaller.

Figs 3.25-3.27 show the recovered circular anomaly after 110 iterations of the joint update of all the model parameters (v_v, ϵ, δ and scatterer positions). We use the multi-scaling approach where, during 5 multi-scale steps, spline-node intervals decrease from 250 m to 65 m and from 1000 m to 130 m, in the vertical and horizontal directions, respectively. In order to regularize the optimization problem we smooth the gradients and apply a hard constraint on the inverted value of ϵ and δ to keep them always positive. The updated parameters reveal the leakage between all the parameter classes.

Fig. 3.25 presents the inverted v_v and scatterer positions where the cross-talk between these parameters results in a slight mismatches between the true and updated scatterer positions below the circular anomaly. The upward shifts of these scatterers compensate partially the underestimated v_v anomaly. The oblique paths crossing the underestimated v_v at the center of anomaly generate some negative perturbations. Due to the cross-talk between the scatterer positions and v_v, these negative values perturb the scatterer positioning on the left and right side of the anomaly.

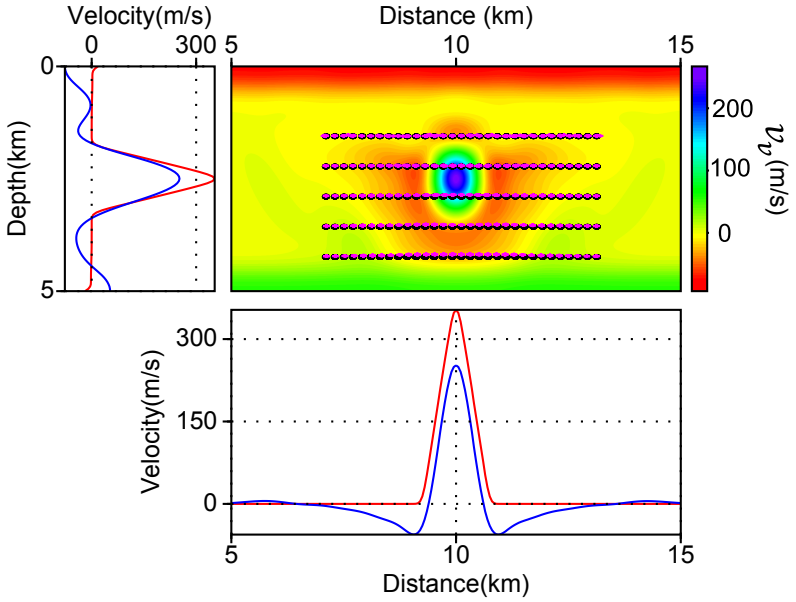

Figure 3.25 – VTI circular anomaly example. Updated v_v circular anomaly and scatterer positions after 170 joint inversion iterations. Diagrams show the direct comparison between true (red) and updated (blue) values along the lines cross cutting the anomaly.

Fig. 3.26 shows the updated model for ϵ. The shape of ϵ sensitivity kernel (Fig. 3.24c) leads to a horizontal smearing for the anomaly. Because of the overlap between the sensitivity kernel of v_v and ϵ for horizontal propagation paths (compare the weights in eqs. (3.103) and (3.104), both include the horizontal derivatives), we can state that the over estimated ϵ above the anomaly mostly is a compensation for the underestimated velocity. Due to the acquisition, the overlap between sensitivity kernel of v_v and δ is more influential (vertical propagation in this acquisition is dominant), and in result the erroneous δ values above the anomaly relevantly are more considerable than those in ϵ.

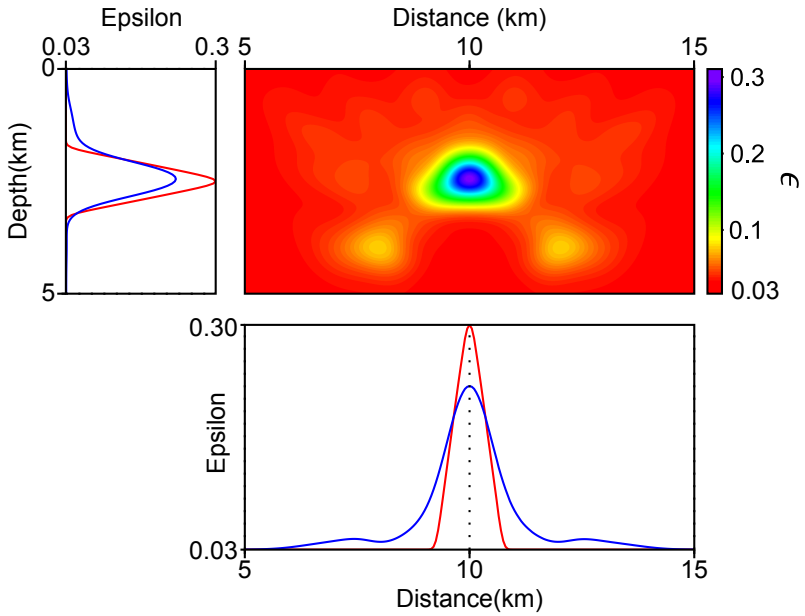

Figure 3.26 – VTI circular anomaly example. Inverted ϵ circular anomaly after 170 joint inversion iterations. Diagrams show the direct comparison between true (red) and inverted (blue) values along the lines cross cutting the anomaly.

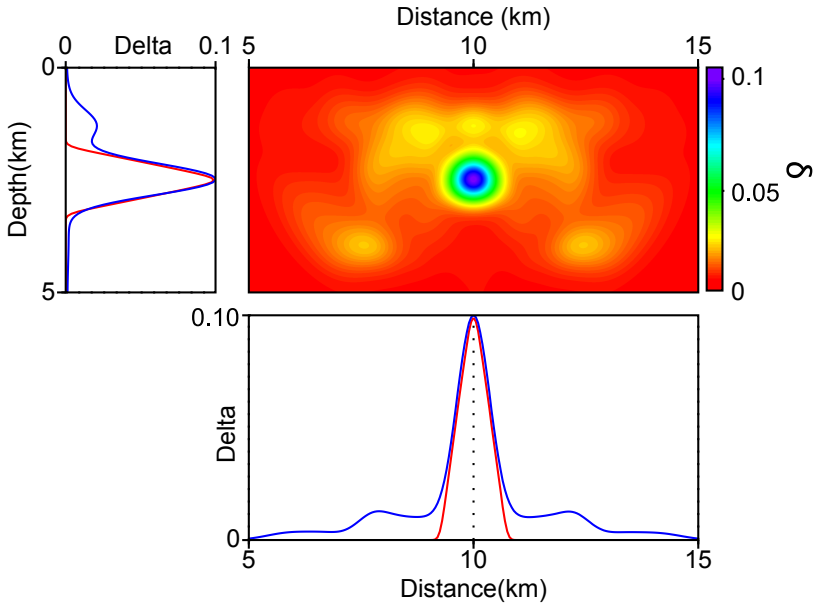

Figure 3.27 – VTI circular anomaly example. Inverted δ circular anomaly after 170 joint inversion iterations. Diagrams show the direct comparison between true (red) and inverted (blue) values along the lines cross cutting the anomaly.

As discussed in the explanation of Fig. 3.24(d), δ sensitivity kernel mostly influences the propagations with small polar angles. Compared to the shape of the updated anomaly for ϵ, there is not considerable smearing for the updated δ anomaly, Fig.3.27. The overestimated values on the top and sides are a compensation for the underestimated velocity. Also the mispositioned scatterers on the shallow part are generating some artefacts on updated δ.

A hard constraint on the ϵ and δ, which keep the updated values positive, makes a blind zone below the updated anomalies where the updated values could be negative. It is worth mentioning that a portion of updated values for ϵ and δ are due to the cross-talk of model parameters. Both mispositioned scatterers and the underestimated velocity can amplify the inverted values for ϵ and δ. In section 3.4 we shall discuss more in detail the leakage between different parameters and the possible remedies such as scaling and parametrization. Note that, these reliable constructions for all the model parameters is achieved for a slope tomographic data set with only few noisy picks and enough illumination of the subsurface by scatterers with large-reflection angle, while in real applications (as we shall discuss in chapter 4), presence of noisy picks and lack of long wavelength propagations can prevent updating the Thomsen's parameters.

Sensitivity of the gradients to offset (reflection angle)

In order to illustrate the sensitivity of $\nabla_{\mathbf{v}_v} C$, $\nabla_\epsilon C$ and $\nabla_\delta C$ to offset variations (or here equivalently the reflection angle), we compare these gradients for different offsets through the following example.

We consider the VTI example in Fig 3.22 but with a homogeneous background for vertical velocity, $v_v = 1500$ m/s, and only the circular anomaly of vertical velocity. We consider one reflection/diffraction point at $(x, z) = (10, 2.3)$ which is sampled by five source-receiver pairs with offsets $1, 2, 4, 6$ and 9 km (Fig. 3.28). We calculate the observables by applying our forward modelling on the true model parameters.

The initial slope tomographic model consists of the true scatterer position and the true background subsurface parameters (no v_v circular anomaly). The absence of the vertical velocity anomaly generates the two-way traveltime and slopes residuals. After solving the state equations (3.97) and (3.98), we calculate the gradients through eqs (3.99)-(3.101).

Figs 3.29-3.31 present the gradients with respect to v_v, ϵ and δ, respectively, for the five mentioned source-receiver pair offsets. As expected by the physics of the parameters and the calculated weights in eqs (3.103)-(3.105), gradients with respect to v_v show same range of sensitivity for all the offsets, while ϵ is responsive only to the long offset propagations (or wide angle reflection). These demonstrations imply that for updating ϵ parameter we do need long-offset acquisitions. Parameter δ has more sensitivity to the midrange offsets. Note that, the sensitivity of v_v and δ to offset variations is close to each other and this shows there is a cross-talk between these parameters for a wide range of offsets.

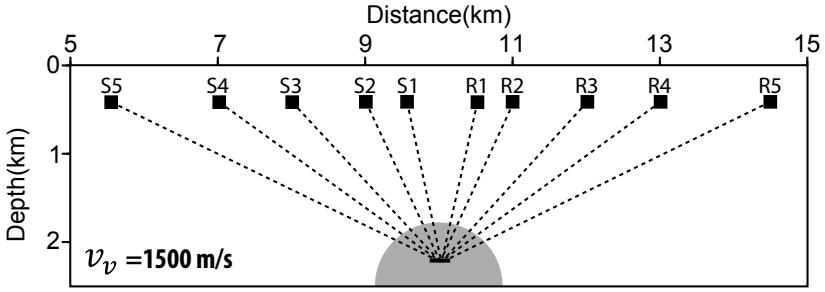

Figure 3.28 – Sensitivity of gradients to offset. The true v_v model which consists of a homogeneous background and an anomaly (the dark half-circle). The scatterer at depth (black facet) is sampled by five source-receiver pairs with offsets $1, 2, 4, 6$ and 9 km. the dashed line designates the rays sampling the scatterers within the initial model, a homogeneous VTI model with $v_v = 1500$ m/s, $\epsilon = 0.03$ and $\delta = 0.01$.

TTI BP salt

As a second synthetic example, we consider the smoothed BP salt model. We build true v_v, ϵ and δ maps by smoothing the original subsurface models with different Gaussian smoothing operators. The smoothing length is 250 m for v_v and 1000 m for ϵ and δ, respectively. Fig. 3.32 shows the true subsurface models which are 11.2 km $\times$ 30 km in size and discretized with a grid interval of 50 m. We consider 370 reflection/diffraction points with 500 m and 1000 m horizontal and vertical interval, respectively. Each of these points are sampled by five source-receiver pairs at the surface with offsets of $1, 3, 5, 8$ and 10 km. This builds a dataset with 1850 scatterers with a rather complete coverage of short and long offsets. However, the maximum reflection angle in the shallow part is larger than in the deep part (compare the schematic ray paths in Fig. 3.32a). We set our goal retrieving the v_v and ϵ models while δ and θ are kept fixed to their true values. The values for $(\sigma_{T_{s,r}}, \sigma_{P_s}, \sigma_{P_s})$ are $(10^{-3}, 10^{-6}, 10^{-6})$. In this example we multiply ϵ with a fixed scaling factor, 4000, to make its order of magnitude same as v_v and scatterer positions.

We initialize the model parameters v_v and ϵ with constant-gradient models (Fig. 3.33a-b). The initial position of the scatterers are random distributions around the true positions with the maximum offset of 500 m (Fig. 3.33). We prefer to avoid the localization step to have a better assessment of the method in joint inversion of the scatterer positions, v_v and ϵ. We apply regularization by smoothing the gradients with a Gaussian operator of window length 1000 m in the horizontal and vertical directions.

During each step of our multi-scaling approach, we first invert jointly for v_v and scatterer positions. After satisfying the convergence criterion (which can be a fixed number of iterations or a specific convergence rate) we restart the inversion by joint inversion of v_v, ϵ and the scatterer positions using the final v_v model and scatterer positions from the previous stage as the starting point for the new inversion. Here the convergence criterion for the joint inversion of v_v and scatterer positions in each scale plays an important role for handling the cross-talk between ϵ and v_v. We stop the joint inversion of the v_v and scatterer positions when the update in v_v and

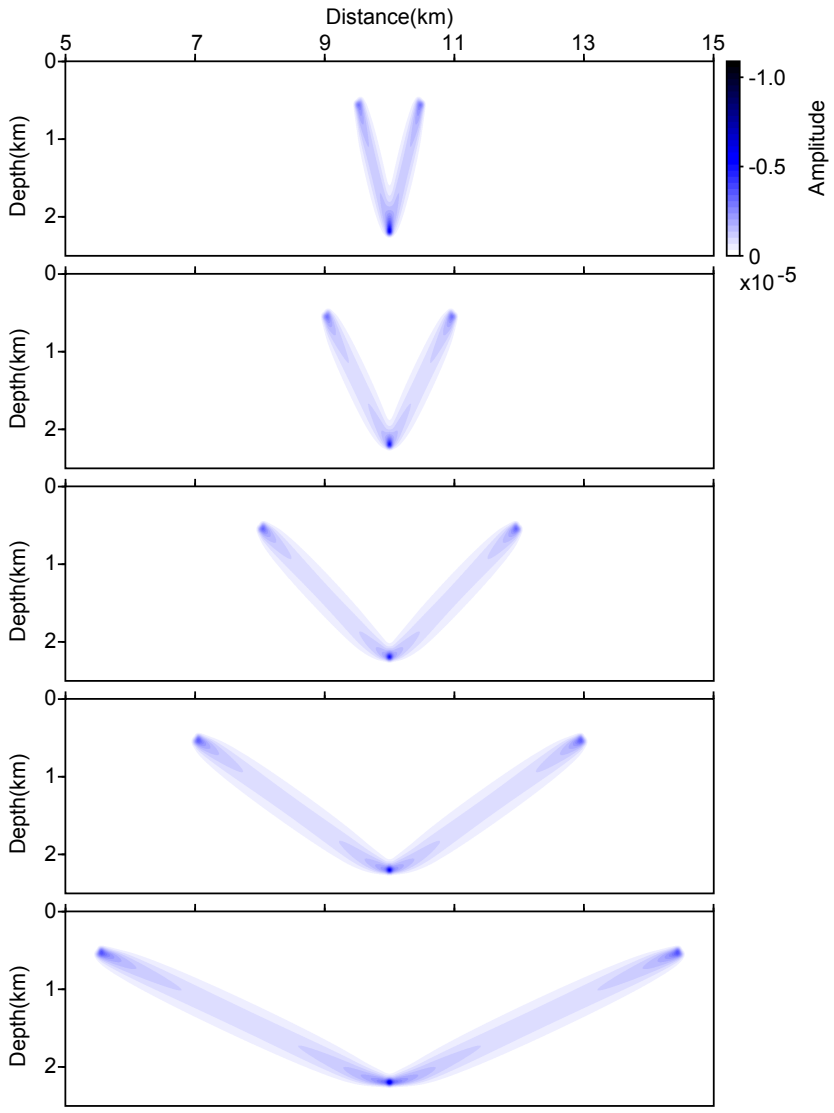

Figure 3.29 – Sensitivity of gradients to offsets. Gradient with respect to v_v for different offsets. Panels show gradient for one scatterer in a homogeneous VTI media for offsets $1, 2, 4, 6$ and 9km, respectively from top to bottom.

scatterer positions become small compared to previous iterations. This can be assessed through the analysis of the convergence rate.

We define three scales such that the vertical and horizontal intervals of spline nodes vary from 250 m to 70 m. Fig. 3.34 shows final updated models of v_v, ϵ and scatterer positions after 405 iterations within three scales (Fig. 3.35). Diagrams in Figs 3.36 and 3.37 show direct comparison between the exact and final updated v_v and ϵ. The v_v model is well reconstructed up to 8 km depth but in deeper parts lack of long wavelength propagation leads to an erroneous

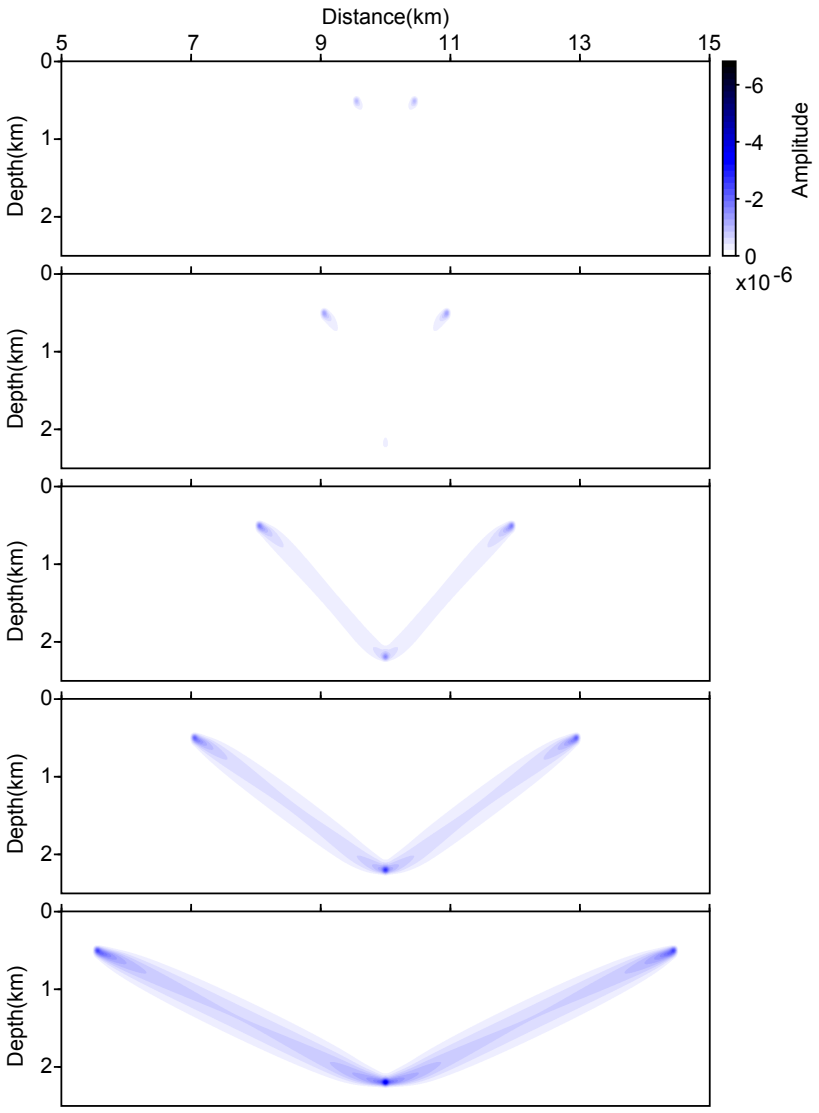

Figure 3.30 – Sensitivity of gradients to offsets. Gradient with respect to ϵ for different offsets. Panels show gradient for one scatterer in a homogeneous VTI media for offsets $1, 2, 4, 6$ and 9km, respectively from top to bottom.

reconstruction of v_v. This deficiency also affects reconstruction of ϵ (Fig. 3.34b).

The final updated ϵ map (Fig. 3.34b) presents a reliable reconstruction down to 3 km depth. Here wide reflection angles of shallow scatterers (< 3 km) provide enough illumination for retrieving ϵ. Moreover, a dense coverage of adjoint fields in the central part makes the updated ϵ even at deeper part (> 3 km) close to the true values (Fig. 3.37). This shows that without a priori information, slope tomography can retrieve an acceptable approximation of ϵ when a wide-aperture illumination is provided by a long-offset acquisition. The cross-talk between the

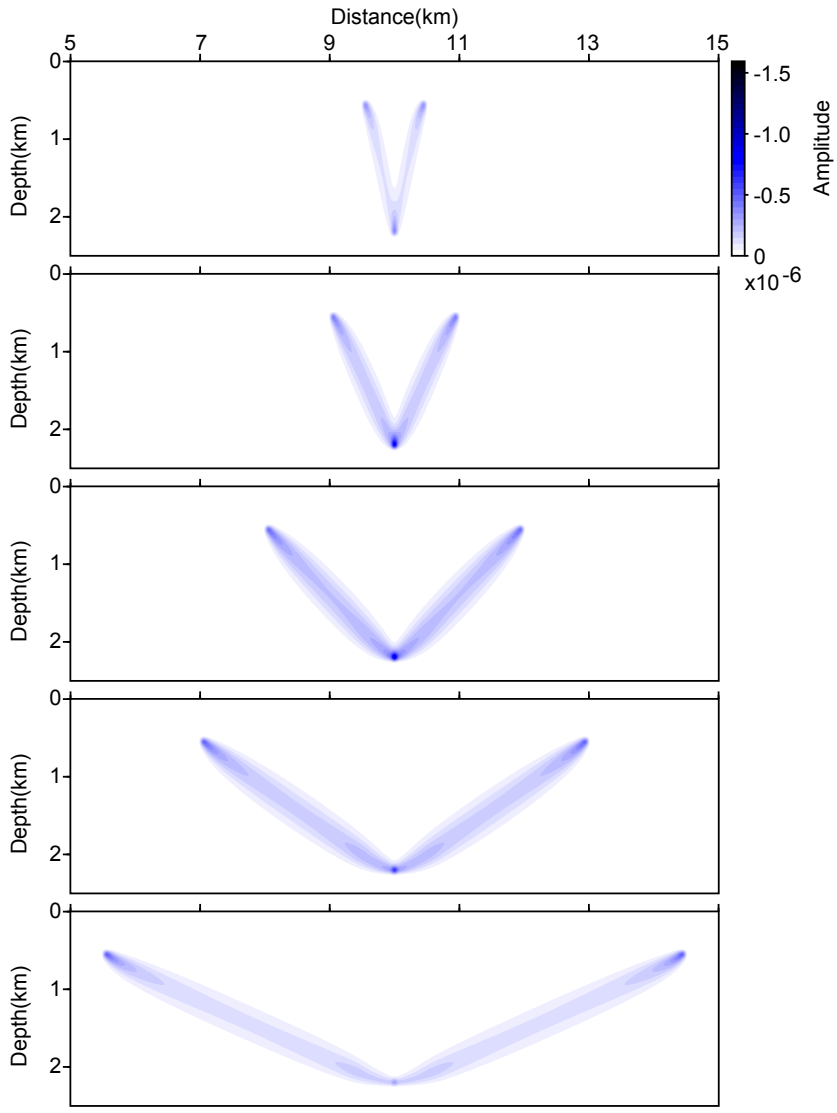

Figure 3.31 – Sensitivity of gradients to offsets. Gradient with respect to δ for different offsets. Panels show gradient for one scatterer in a homogeneous VTI media for offsets $1, 2, 4, 6$ and 9km, respectively from top to bottom.

v_v, ϵ and scatterer positions is the most influential issue which prevents a better reconstruction of ϵ. We observe that increasing the density of the reflection/diffraction points for this acquisition does not improve the resolution and accuracy of the inverted ϵ. This emphasizes the importance of wide azimuth reflections and diving waves in reconstruction of the ϵ model.

The upward shift for the updated positions below 9 km depth (Fig. 3.34a) is the direct consequence of v_v and ϵ underestimation. However, in the shallow part there are some erroneous positioning of the scatterers associated with long-offset source-receiver pairs. For offsets greater

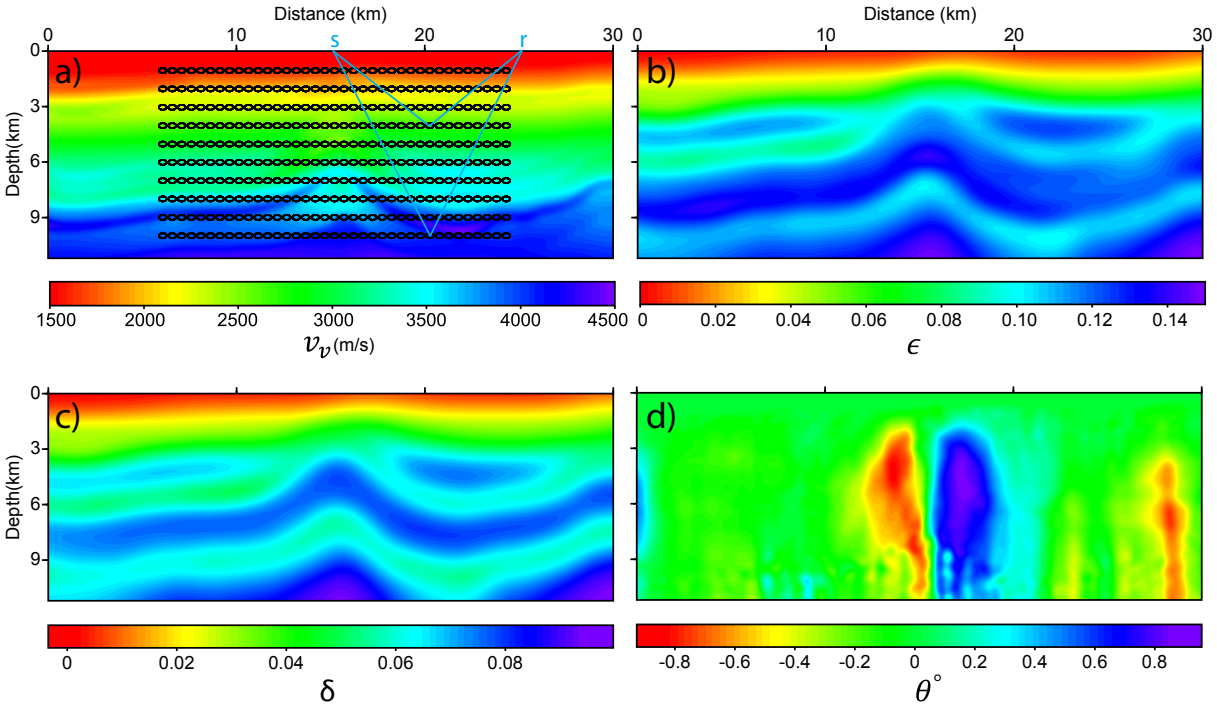

Figure 3.32 – TTI BP salt example. a) True models for v_v, b) ϵ, c) δ and d) tilt angle. The black circles in a) designate the true position for 370 reflection points. Each reflection point is sampled with five source-receiver pairs with offsets 1, 3, 5, 8 and 10km.

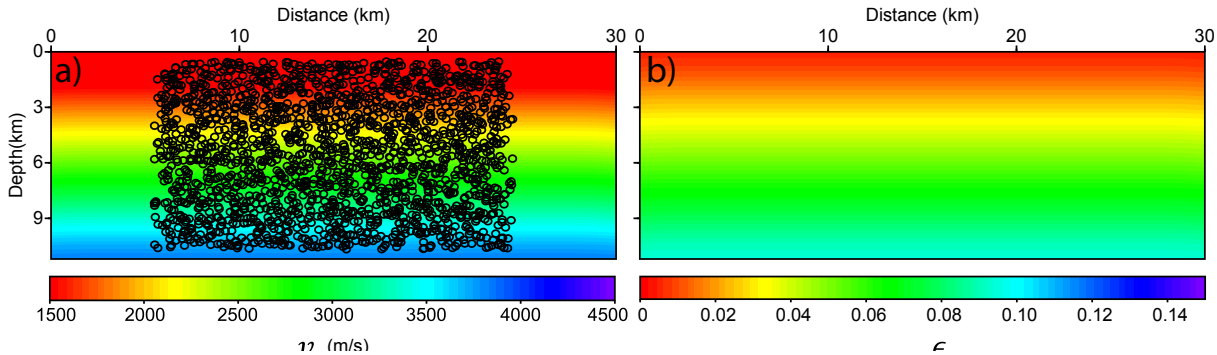

Figure 3.33 – TTI BP salt example. Initial models for a) v_v and b) ϵ. The black circles on a) represents the random initial scatterer positions.

than 8 km we observe propagation of diving waves (Fig.3.38). This implies that the observables built by our forward modelling engine for shallow scatterers (and with large reflection angle) are erroneous. These scatterers play the role of noise in our synthetic data set.

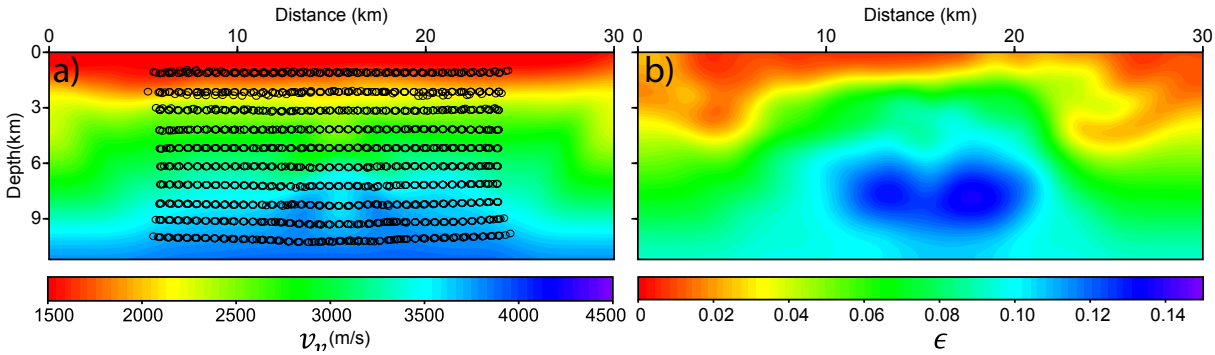

Figure 3.34 – TTI BP salt example. Inverted model for a) v_v and b) ϵ. The black circles on a) represent the final inverted scatterer positions.

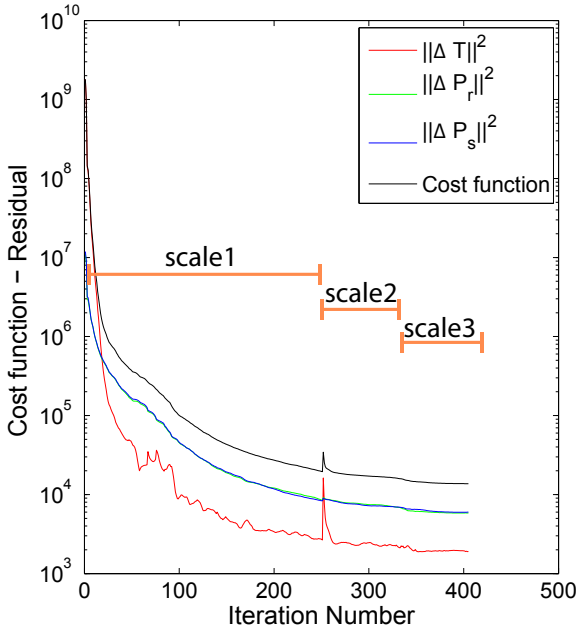

Figure 3.35 – TTI BP salt example. Convergence Diagram plotted with a logarithmic scale. There are 3 scales within 405 iterations. The curves show the misfit function with respect to iteration number for the traveltime (red), slope at source (blue), slope at the receiver (green), and total misfit function (black). Summation of the residual curves (red, green and blue) is the total misfit function.

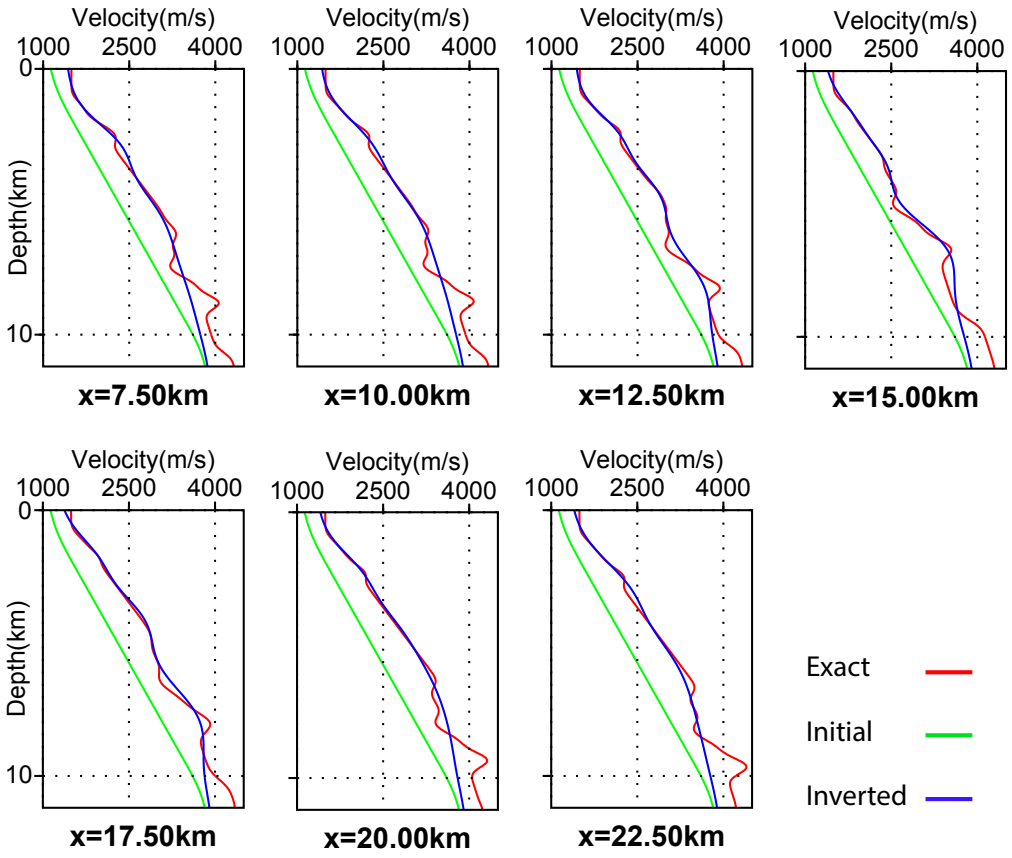

Figure 3.36 – TTI BP salt example. Direct comparison between the exact (red) and inverted (blue) v_v by anisotropic slope tomography in Fig. 3.34(a). The green line designates the initial velocity model.

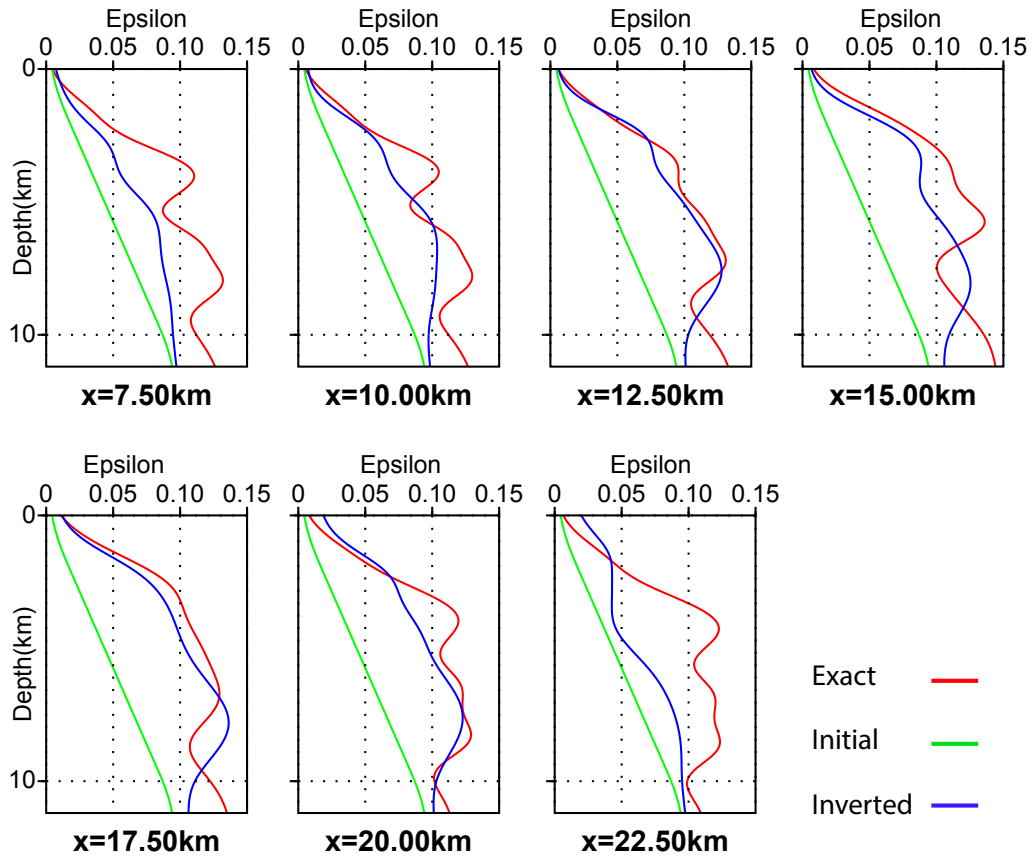

Figure 3.37 – TTI BP salt example. Direct comparison between the exact (red) and inverted (blue) ϵ by anisotropic slope tomography in Fig. 3.34(b). The green line designates the initial ϵ model.

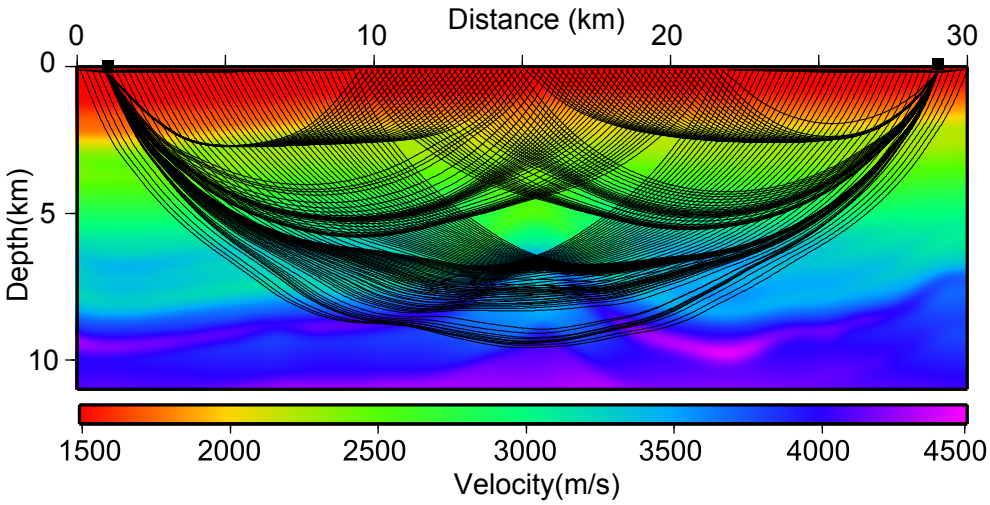

Figure 3.38 – TTI BP salt example. Isotropic ray tracing for true BP v_v model. There is two sources at the surface (black squares). Diving waves are generated for offsets $> 7 - 8$ km.

3.4 On the footprint of parameter cross-talk and parametrization

3.4.1 Assessment of parameter cross-talk

To gain more quantitative insight on the parameter cross-talk between different parameter classes we design the following tests. We consider a similar true model as the VTI circular anomaly example (Fig. 3.22, section 3.3.4) but without circular anomaly inclusion for ϵ and δ. The non-zero value for ϵ and δ in background keeps our tests in the framework of VTI media. The inversion goal shall be the reconstruction of the v_v anomaly. The anisotropic parameters (ϵ, δ) and scatterer positions (X_{sct}) are initialized by their true values and initial v_v model is the true constant-gradient background (v_v model in Fig. 3.23a). We perform several inversion schemes which differ in frozen model parameters (namely passive parameters) and unfrozen model parameters (namely optimization parameters). The passive parameters always are fixed to their true values. Therefore, for all the following tests v_v is always an optimization parameter and ϵ, δ and scatterer positions can be a passive or optimization parameter. In order to assess the inversion solutions, in the following we compare the difference between the final updated optimization parameter(s) and the associated true background value (perturbation models). The perturbation model for a passive parameters is zero.

Test series 1: scatterer positions as passive parameter

As a first series of tests, we consider the scatterer positions as passive parameters and fixed to their true positions while ϵ and δ can be a passive or optimization parameter. Fig 3.39(a) shows the true v_v anomaly (difference between the true velocity model and the constant-gradient background). Panels of the row (b) are the solution of a mono-parameter (v_v) inversion scheme where both ϵ and δ are passive parameters. Inverted v_v anomaly (Fig 3.39b1) is nearly identical to the true one (Fig. 3.39a).

The joint inversion of v_v and ϵ reveals a small leakage between these two parameters (Fig. 3.39c1-c3). Because of this leakage, we observe some smearing for the anomaly in the vertical direction, as well as underestimation of v_v at the central part of the anomaly. The smearing, which generates overestimated v_v outside the inclusion, is a compensation for underestimated v_v in the central part. This underestimated values are generated by adjoint ray tubes sampling the erroneous inverted ϵ at the center.

The joint inversion of v_v and δ, while ϵ is a passive parameter, generates some artefact on the inverted δ model (Fig. 3.39d1-d3). Comparison with the previous inversion for v_v-ϵ (Fig. 3.39c1-c3) emphasizes that leakage between the optimization parameter δ-v_v is more significant than for v_v-ϵ. Again, here the inverted v_v anomaly is smeared and the updated values at the center of anomaly are underestimated (Fig. 3.39d1). This underestimation is more noticeable than the previous test in Fig. 3.39(c1). This is mostly because of the erroneous inverted δ map and cross-talk between optimization parameters. Due to the shape of δ sensitivity kernel, which is more responsive to the mid-range propagations, the smearing of the v_v anomaly in the

vertical direction is more significant than during the epsilon update (Fig. 3.39c1). Compared to the joint inversion of v_v and ϵ (Fig. 3.39c1-c3), here the smearing is generated by narrower reflection angles and this results in a slight difference between the shape of the inverted v_v anomaly in two tests.

Fig. 3.39(e1-e3) shows the perturbation models after joint inversion of v_v, ϵ and δ (three optimization parameters). Here, the error for inverted v_v anomaly is larger than two previous joint inversions of in Fig. 3.39(c1) and Fig. 3.39(d1). In addition to leakages between v_v-δ and v_v-ϵ, the cross-talk between ϵ and δ also takes part in the underestimation of v_v at the center of anomaly. Smearing and the amplitude of the artefacts in the updated δ and ϵ are more significant than two previous tests (compare second and third column of rows (e), (d) and (c) in Fig. 3.39). In order to understand better the source of these artefacts we plot in Figs. 3.40(a-c) the gradient w.r.t v_v, ϵ and δ at the first iteration of this joint inversion. Here the gradient with respect to v_v (Fig. 3.40a) has sensitivity to the both vertical and horizontal propagations while the gradient map of ϵ (Fig. 3.40b) is only sensitive to wide angle reflections. The gradient with respect to δ (Fig. 3.40c) includes high amplitudes for the midrange reflection angles. These patterns for the gradients coincide with the smearing in v_v anomaly and the artefacts on the updated ϵ and δ maps.

Test series 2: scatterer positions as optimization parameter

As a second series of tests, we consider the same configuration in the previous tests but here scatterer positions are always optimization parameters and they are initialized with the true positions.

Fig. 3.41 includes solutions of different inversion schemes. Compared with Fig. 3.39, there are two extra columns in the right, column 4 and 5, which present the perturbation of updated scatterer positions from their true positions in vertical (z) and horizontal (x) directions. The dots are plotted at the true X_{sct} and their size and colors indicate the horizontal and vertical perturbations of the scatterer positions. These scatterer perturbation maps are all in the same scale, and note that, each dot represents the superimposition of perturbations from five independent scatterers sampled by five different offsets.

Fig. 3.41(f1-f5) shows the final perturbation models of v_v and X_{sct} joint inversion where ϵ and δ are passive parameters. The underestimated velocity at the center of v_v anomaly is compensated by mis-positioning of the scatterers in the vertical and horizontal directions. Because of the acquisition layout and leakage between v_v and X_{sct} we observe again a vertical smearing. For X_{sct} the horizontal errors are higher than vertical ones. Thanks to the v_v gradient (sensitive to both vertical and horizontal propagation), for many scatterers the ambiguity between z components of X_{sct} and v_v is resolved well. But for the x components, lack of horizontal propagations results in erroneous values for the horizontal positioning of X_{sct}.

By performing joint inversion of v_v, ϵ and X_{sct} while δ is a passive parameter we observe some artefacts on the updated ϵ (Fig. 3.41g1-g5). These erroneous ϵ updates are originated primarily from cross-talk between v_v-ϵ (such as the test in Fig. 3.39c1-c2) and have been further strengthened by adding new degrees of freedom in the inversion though the update of ϵ-X_{sct}.

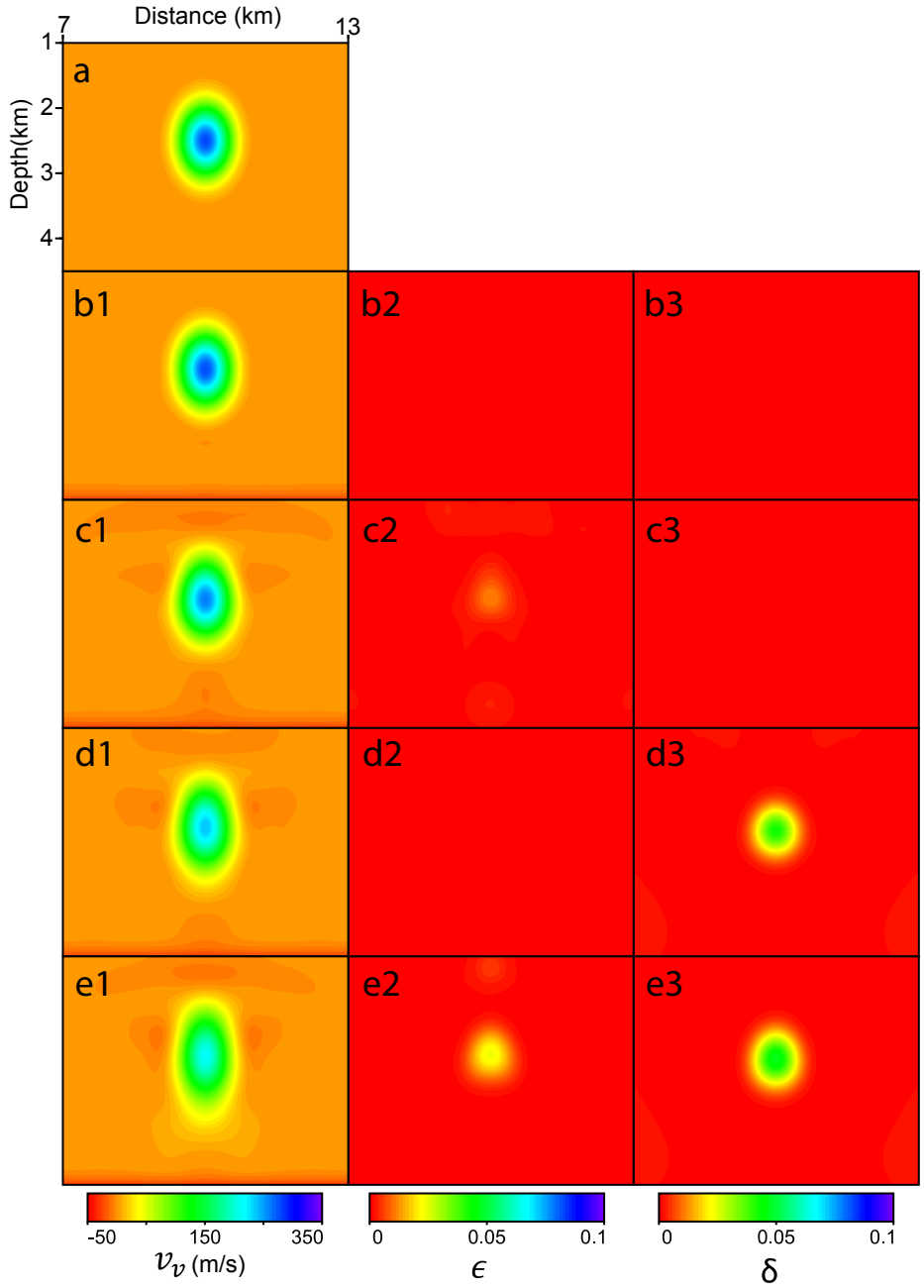

Figure 3.39 – Assessment of parameter cross-talks. a) True v_v circular anomaly. Difference maps for updated v_v, ϵ and δ (respectively column one, two and three) for different joint inversion schemes. Here optimization and passive parameters are (passives are in brackets): b1-b3) v_v, $[\epsilon, \delta, X_{sct}]$, c1-c3) v_v, ϵ, $[\delta, X_{sct}]$, d1-d3) v_v, δ, $[\epsilon, X_{sct}]$ and e1-e3) v_v, ϵ, δ, $[X_{sct}]$.

Compared to the joint inversion of v_v-X_{sct} (Fig. 3.41f4-f5), here the perturbations of the z coordinate of the scatterers are more significant and these vertical perturbations increase with depth as the wide aperture illumination decreases. This results because adding ϵ to the optimization

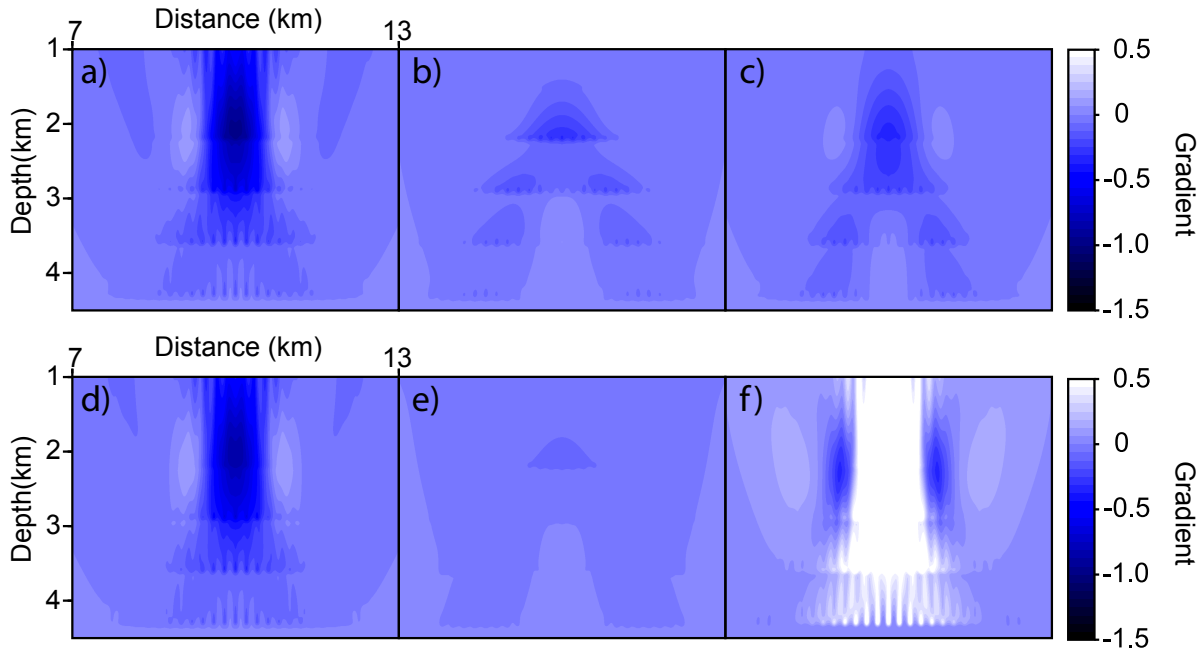

Figure 3.40 – Gradient at first iteration of a joint inversion with X_{sct} as passive parameter. Top, parametrization (v_v, ϵ, δ): gradient w.r.t a) v_v, b) ϵ, c) δ. Bottom, parametrization (v_{nmo}, v_h, δ): gradient with respect to d) v_{nmo}, e) v_h, f) δ.

parameters increases the ill posedness and strengthen the error in v_v, ϵ and z. During the early iterations of joint inversion, the missing v_v anomaly generates both positive v_v and ϵ perturbations as well as an upward shift of the scatterers. The mispositioned scatterers sampled with wide reflection angles generate strong artefacts in the updated ϵ (such as trails at the bottom left and right) while if we keep X_{sct} as passive parameters (like the test in Fig.3.39c1-c3) this error is far less. The erroneous updated ϵ disables the inversion to recover well the v_v anomaly and consequently perturbs X_{sct} in the vertical direction.

Fig. 3.41(h1-h5) shows the perturbation models for joint inversion of v_v, δ and X_{sct} while ϵ is a passive parameter. Compared to the previous joint inversion scheme for v_v, ϵ and X_{sct} (Fig. 3.41g1-g5), here estimation of v_v at the center of anomaly is slightly closer to the true values and this can be the result of less perturbation in the vertical positioning of X_{sct}. During the inversion when the scatterers are perturbed upward to compensate absence of v_v anomaly, the sensitivity kernels of δ and v_v are both boosted at the anomaly region. Because of overlap between sensitivity pattern for v_v and δ, the underestimated v_v mostly emerges as artefacts on δ map rather than perturbation of X_{sct}. Compared with the joint inversion of v_v, ϵ, and X_{sct} (Fig. 3.41), this results in less perturbation in z component of X_{sct} and consequently less smearing for the v_v anomaly. The horizontal smearing in v_v is balanced by higher errors in x component of scatterer positions.

Fig. 3.41(i1-i5) shows the perturbation models for joint inversion of all the parameters v_v, ϵ,

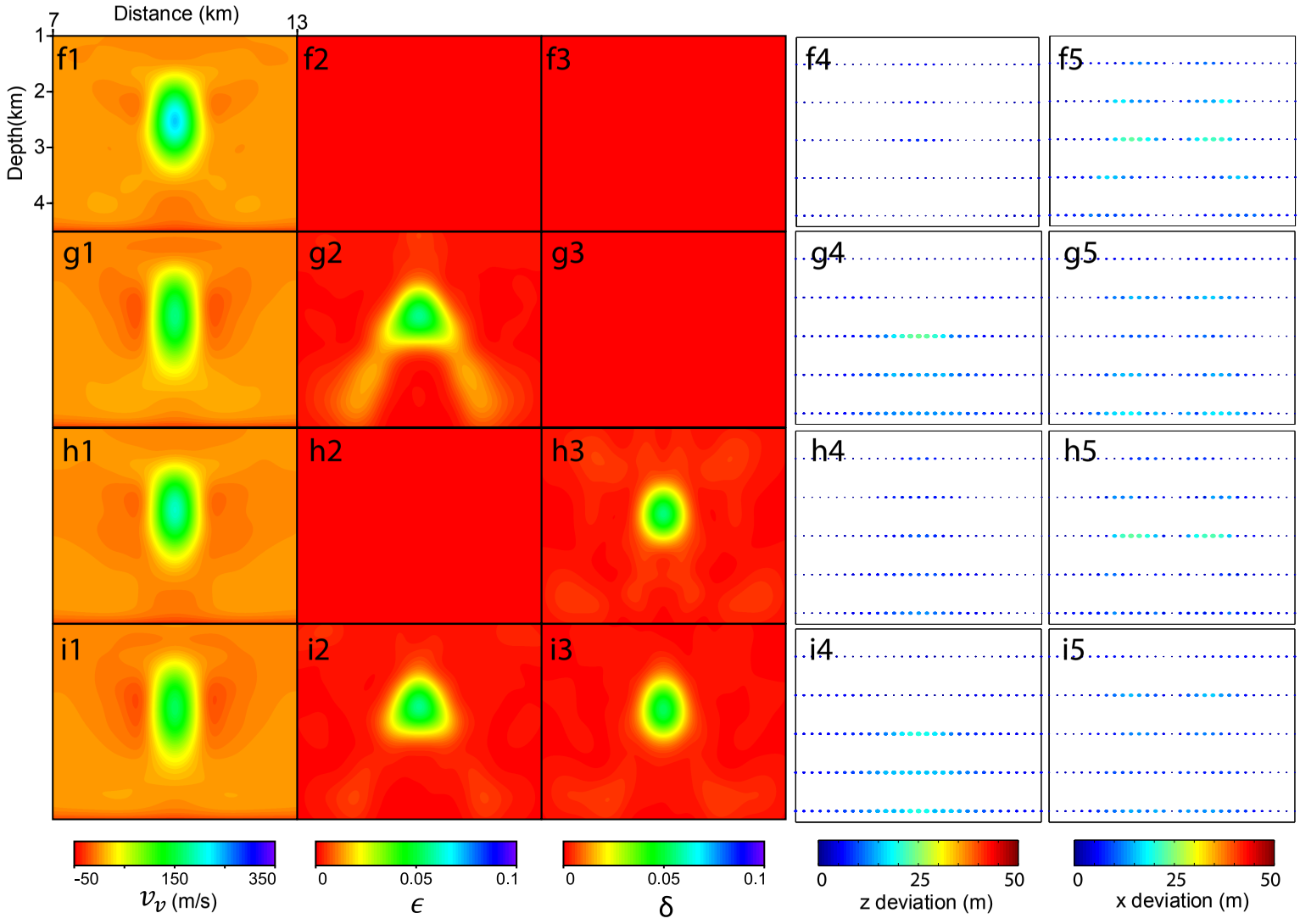

Figure 3.41 – Assessment of parameter cross-talks. Difference maps for updated v_v, ϵ, δ, z and x components of X_{sct} (respectively from first to fifth column) for different joint inversion schemes. Here optimization and passive parameters are (passives are in brackets): f1-f5) v_v, X_{sct},$[\epsilon, \delta]$, g1-g5) v_v, X_{sct}, ϵ, $[\delta]$, h1-h5) v_v, X_{sct}, δ, $[\epsilon]$ and i1-i5) $v_v, X_{sct}, \epsilon, \delta$.

δ and X_{sct}. The updated v_v anomaly presents the same pattern of smearing and underestimation as previous testes. Compared to the joint inversion of v_v-ϵ-X_{sct} (Fig. 3.41g1-g5), the ϵ artefacts are less strong. These artefacts and the mis-positioning of the scatterers generate smaller data residuals for the deep scatterers with wide reflection angles and this reduces the amplitude of the oblique adjoint ray tubes passing the anomaly. Here during the early iterations of the joint inversion, emergence of a strong δ and ϵ artefacts makes underestimation and smearing of the v_v anomaly stronger than other tests. The perturbed z components are mostly relates to the scatterers sampled by narrow reflection angles where estimated v_v and δ play a major role in accurate scatterer positioning.

Test 3: offset range and multi-scaling effect

There are two other issues worth discussing here: first, what is the effect of increasing reflection angles (offset) on the leakage patterns? and second, does the multi-scaling approach mitigate the cross-talk between parameters or non-uniqueness solution of the inversion? To answer these questions we perform the following tests.

Consider joint inversion of all the parameters, i.e. v_v, ϵ, δ and X_{sct} where each reflection point is sampled by five offsets $0.8, 1.6, 2.4, 3.2$ and 4 km (same as test 2 in Fig 3.42i1-i5). We perform joint inversion of all the parameters for two new offset ranges: a shorter range ($0.4, 0.8, 1.2, 1.6$ and 2 km) and a longer range ($1.2, 2.4, 3.6, 4.8$ and 6 km). Fig 3.42 shows the final perturbation models for these offset ranges. The larger offset range (Fig 3.42m1-m5) boosts the sensitivity of ϵ and in result the artefact on the ϵ map are stronger compared to the shorter offset ranges (Fig 3.42i2 and n2). Larger offsets increases the illumination of the subsurface along subhorizontal wave paths and this resolves better the ambiguity between the horizontal positioning of scatterers and subsurface parameters (compare Fig 3.42m5, i5 and n5). Since the vertical positioning of the scatterers is influenced by both v_v and δ and there is a strong cross-talk between these two parameters, modifying the offset range does not impact significantly the vertical positioning of the scatterers (compare Fig 3.42m4, i4 and n4). Accordingly, modifying the offset range does not also affect significantly the amplitudes of the δ perturbations. Shorter offset ranges favours the sensitivity of v_v in the vertical direction and considering the mis-positioned z components and cross-talk between v_v-δ, this leads to more smearing of v_v for the shorter offset ranges. Therefore, improving the illumination helps better controlling the scatterer positions but tends to increase leakage between subsurface parameters (in particular v_v and ϵ).

As a last test we demonstrate the effect of multi-scaling approach on mitigating the cross-talk between different parameter classes. In the multi-scaling approach the inversion starts on a coarse grid of spline nodes for subsurface parameters and during the inversion the number of nodes are increased, while for mono-scaling approach the inversion is performed on a fixed Cartesian grid. In Fig. 3.43 we compare the perturbation models from joint inversion of all parameters in multi-scaling approach (same as test 2 Fig 3.41i1-i5) and mono-scaling approach on Cartesian grid (Fig. 3.43p1-p5). The comparison reveals that the δ and ϵ perturbations are stronger in the mono-scaling approach than in the multi-scaling one (compare Fig. 3.43i2-i3 and p2-p3). This is correlated with smaller perturbations of both the vertical and horizontal

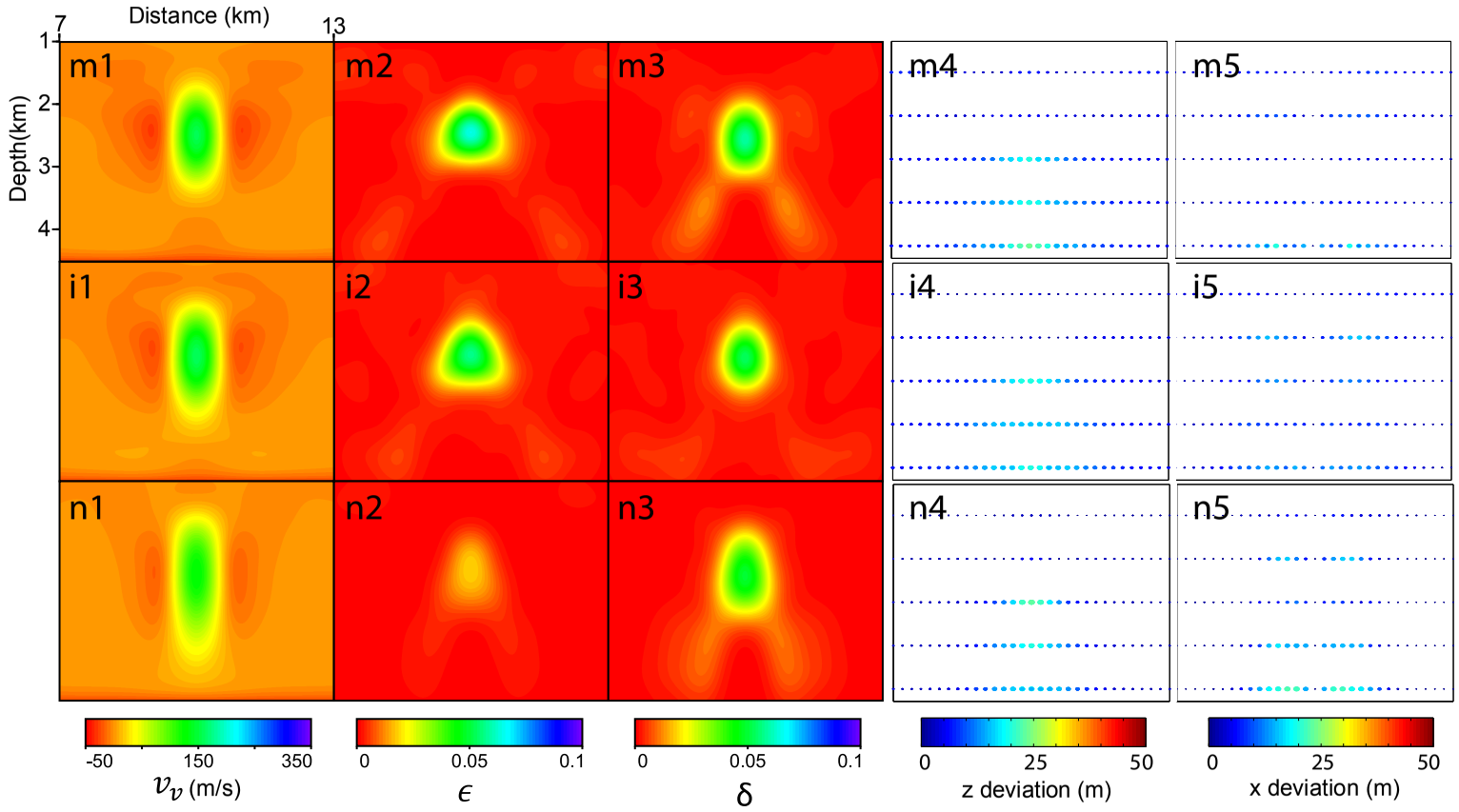

Figure 3.42 – Assessment of parameter cross-talks. Offset range effect. Difference maps for updated v_v, ϵ, δ, z and x components of X_{sct} (respectively from first to fifth column) for joint inversion of all the parameters in different offset ranges for scatterers. Offset ranges are m1-m5) $(1.2, 2.4, 3.6, 4.8, 6)$ km, i1-i5 $(0.8, 1.6, 2.4, 3.2, 4)$ km and n1-n5) $(0.4, 0.8, 1.2, 1.6, 2)$ km. Maps in i1-i5 are same as Fig. 3.41(i1-i5).

positioning of scatterers when multi-scaling is applied. The perturbation of z and x components of X_{sct} in multi-scaling approach is less. The negative perturbations on the left and right sides of the v_v inclusion is stronger when mono-scaling is applied and these contribute in erroneous δ and ϵ updates as well as the scatterer position. The multi-scaling approach manages to insert lower wavenumber at lower scales and by adding more nodes during upper scales it retrieves the high frequencies. During the lower scales, the number of subsurface nodes are less and this reduces the ill-posedness of the inverse problem and prevents from convergence to a local minimum. For realistic subsurface models, using a mono-scale approach leads the inversion toward a local minimum during early iterations.

Therefore, a combination of a proper offset distribution for scatterers and multi-scaling approach can be the first remedy for the problem of cross-talk between parameters.

3.4.2 Impact of subsurface parametrization upon anisotropic slope tomography

In addition to the scaling which can steer the inversion, reparametrization of the model space is another tool to mitigate the cross-talk between the different parameters and make the inversion better-conditioned. Each parametrization is a set of independent parameter classes that fully describe the properties of the subsurface governing wave propagation. The choice of parametrization in waveform inversion methods is a key issue and is driven by three main specifications: 1) the trade-off between spatial resolution and parameter cross-talks which can be assessed by the so-called radiation pattern that control amplitude versus angle variation of the partial derivative (scattered) wavefield 2) choice of a suitable set of optimization parameters to prevent over parametrization, and 3) scaling of different parameter classes to improve the conditioning of the inverse problem. In the framework of FWI there are different studies which investigate influence of parametrization (Forgues and Lambaré, 1997; Operto et al., 2013; Gholami et al., 2013b; Plessix and Cao, 2011; He and Plessix, 2017; Rusmanugroho et al., 2017). However, these concerns can be extended to other imaging methods which does not apply a full waveform modelling, such as anisotropic tomography. In the following we seek to gain preliminary insights on the sensitivity of anisotropic adjoint slope tomography to the subsurface parametrization.

Parametrization of anisotropic slope tomography primarily relies on different representation of anisotropy in the medium. Among different set of parameters that describes a VTI medium are (v_v, ϵ, δ), (v_{nmo}, v_h, δ), $(v_{nmo}, \eta, \epsilon)$ and (v_h, η, δ). Considering the relations $v_{nmo} = v_v\sqrt{1+2\delta}$, $v_h = v_v\sqrt{1+2\epsilon}$ and $\eta = \epsilon - \delta/1+2\delta$, these sets fully define the subsurface properties. We can calculate the gradient of misfit function with respect to these parameter sets by

$$\nabla_{p_1(\mathbf{x})} C = -\sum_{s=1}^{N_\mathbf{s}} \lambda_s(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_s(\mathbf{x}))}{2\,\partial p_1(\mathbf{x})} - \sum_{r=1}^{N_\mathbf{r}} \lambda_r(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_r(\mathbf{x}))}{2\,\partial p_1(\mathbf{x})}, \tag{3.112}$$

$$\nabla_{p_2(\mathbf{x})} C = -\sum_{s=1}^{N_\mathbf{s}} \lambda_s(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_s(\mathbf{x}))}{2\,\partial p_2(\mathbf{x})} - \sum_{r=1}^{N_\mathbf{r}} \lambda_r(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_r(\mathbf{x}))}{2\,\partial p_2(\mathbf{x})}, \tag{3.113}$$

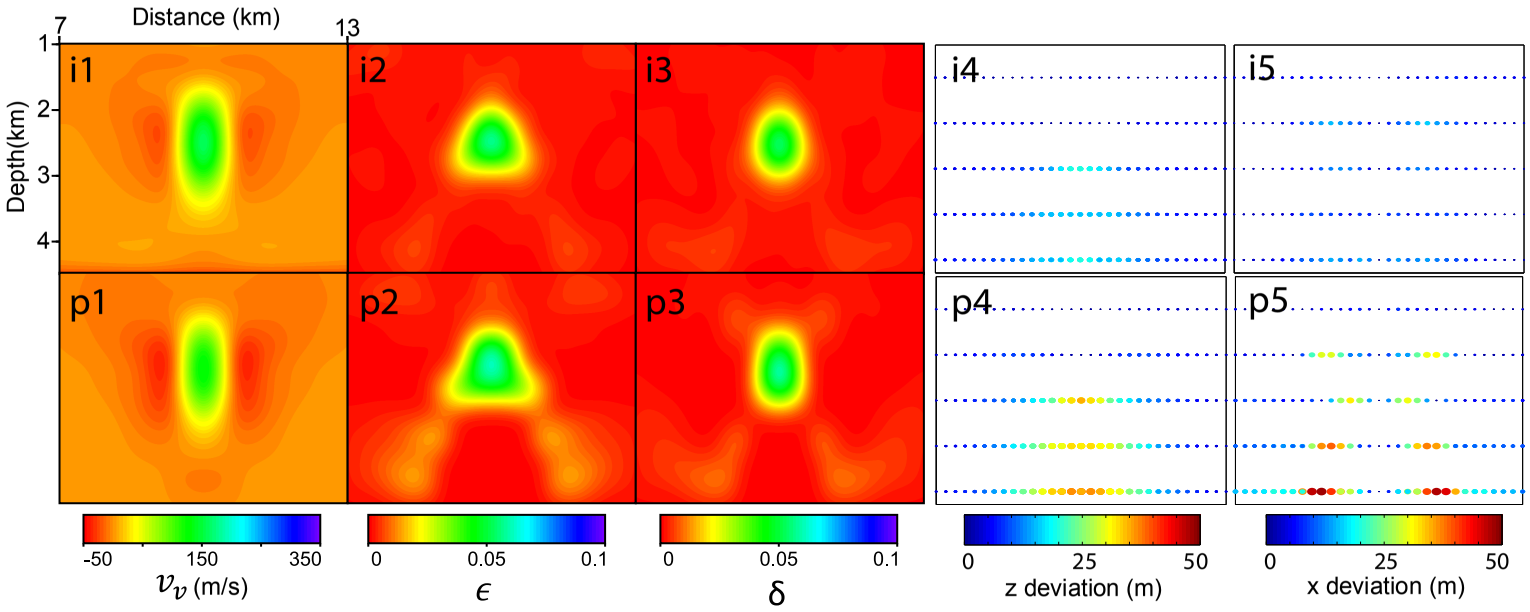

Figure 3.43 – Assessment of parameter cross-talks. Multi-scaling effect. Difference maps for updated v_v, ϵ, δ, z and x components of X_{sct} (respectively from first to fifth column) for joint inversion of all the parameter in i1-i5) multi-scaling approach and p1-p5) mono-scaling on Cartesian grid.

$$\nabla_{p_3(\mathbf{x})} C = -\sum_{s=1}^{N_\mathbf{s}} \lambda_s(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_s(\mathbf{x}))}{2\,\partial p_3(\mathbf{x})} - \sum_{r=1}^{N_\mathbf{r}} \lambda_r(\mathbf{x}) \frac{2\,\partial H(\mathbf{x}, \nabla t_r(\mathbf{x}))}{2\,\partial p_3(\mathbf{x})}, \tag{3.114}$$

where (p_1, p_2, p_3) stand for the parameters defining the VTI model (similar to eqs (3.99)-(3.101) for (v_v, ϵ, δ) parametrization). Here Hamiltonian H is formulated with the same parameters under question.

It is worth drawing the analogy between radiation pattern in full waveform modelling and the sensitivity kernels (or gradients) in our formulation. Radiation pattern in FWI is defined by $\partial B/\partial \mathbf{m}$ where B stands for the forward modelling and $\mathbf{m}$ the model parameters. Here, $\partial H/\partial \mathbf{m}$ builds the spatially-dependent coefficients that weight the sensitivity kernel of the subsurface parameter update along the scatterer-to-surface propagation paths (eqs. (3.103)-(3.106)). It is important to stress that the scatterers in slope tomography do not play the same role as the virtual scattering sources in FWI. In the first case, the scatterers are used as secondary sources in depth to compute the sensitivity kernel of slope tomography along transmission paths connecting the scatterer to the sources and receivers and update the subsurface along these paths accordingly. In FWI, they are used as the second-order source of the Fréchet derivative wavefield to compute the parameter update at the scatterer position. In this case, the scattering angle controls the resolution of the parameter under question. Using the term of "radiation pattern" in slope tomography is not appropriate since there is no amplitude information here. But, despite the fact that slope tomography deals only with the kinematic part of the data, there are many similarities between the radiation patterns of PP transmission waves in FWI (He and Plessix, 2017) and sensitivity to the propagation direction in slope tomography. In fact, slope tomography can be viewed as a kinematic equivalent of RWI (reflection waveform inversion) where the reflector at depth are used as secondary sources to update the velocities along up-going paths connecting the reflector to the surface.

In this study we do not assess all the possible parametrization for slope tomography. We only consider (v_{nmo}, v_h, δ) parametrization as a possible, not the best, alternative to (v_v, ϵ, δ) parametrization. This parametrization can be interesting since it provides a decoupling between the vertical and horizontal propagation while there is no need to scale two velocities. There are other widely-used parametrization such as (v_{nmo}, η, δ) (Alkhalifah and Tsvankin, 1995), which favours application like amplitude-preserving-migration (Kiyashchenko et al., 2004). For this parametrization again we need scaling of v_{nmo} or η.

With parametrization (v_{nmo}, v_h, δ) TTI eikonal equation (3.84) can be written as

$$H(\mathbf{x}, \nabla t(\mathbf{x})) = A(\nabla_{\hat{x}} t(\mathbf{x}))^2 + C(\nabla_{\hat{z}} t(\mathbf{x}))^2 + E(\nabla_{\hat{x}} t(\mathbf{x}))^2 (\nabla_{\hat{z}} t(\mathbf{x}))^2 - 1, \tag{3.115}$$

where

$$\begin{aligned} A =& v_h^2, \\ C =& \frac{v_{nmo}^2}{(1+2\delta)}, \\ E =& v_{nmo}^2 \frac{(v_{nmo}^2 - v_h^2)}{(1+2\delta)}. \end{aligned} \tag{3.116}$$

The equivalent of these coefficients in (v_v, ϵ, δ) parametrization are $A = v_v^2(1+2\epsilon)$, $C = v_v^2$

and $E = -2v_v^4(\epsilon - \delta)$.

In order to derive the gradients of the misfit function with respect to the parameters (v_{nmo}, v_h, δ), similar to eqs. (3.99)-(3.106), we need only to calculate gradient of Hamiltonian (3.115) with respect to (v_{nmo}, v_h, δ) as follow

$$\frac{\partial H(\mathbf{x}, \nabla t(\mathbf{x}))}{\partial v_{nmo}(\mathbf{x})} = \frac{2v_{nmo}}{1+2\delta}\Big(((\mathcal{R}\nabla t)_z)^2 + (2v_{nmo}^2 - v_h^2)((\mathcal{R}\nabla t)_x(\mathcal{R}\nabla t)_z)^2\Big), \tag{3.117}$$

$$\frac{\partial H(\mathbf{x}, \nabla t(\mathbf{x}))}{\partial v_h(\mathbf{x})} = 2v_h\Big(((\mathcal{R}\nabla t)_x)^2 - \frac{2v_{nmo}^2 v_h}{1+2\delta}((\mathcal{R}\nabla t)_x(\mathcal{R}\nabla t)_z)^2\Big), \tag{3.118}$$

$$\frac{\partial H(\mathbf{x}, \nabla t(\mathbf{x}))}{\partial \delta(\mathbf{x})} = -\frac{2v_{nmo}^2}{(1+2\delta)^2}\Big(((\mathcal{R}\nabla t)_z)^2 + (v_{nmo}^2 - v_h^2)((\mathcal{R}\nabla t)_x(\mathcal{R}\nabla t)_z)^2\Big). \tag{3.119}$$

The gradient with respect to tilted angle θ is same as eq. (3.106).

In the expression (3.117) the dominant term is linked to the gradient of traveltime in the vertical direction and this means we expect more sensitivity to vertical propagations for v_{nmo} gradient. But the equivalent of this expression in (v_v, ϵ, δ) parametrization, i.e. $\dfrac{\partial H}{\partial v_v}$, has two dominant terms, associate with gradient of traveltime in the vertical and horizontal directions, which makes the sensitivity kernels of v_v responsive to both vertical and horizontal propagation.

The gradient of the Hamiltonian with respect to v_h presents same sensitivity pattern as the gradient with respect to ϵ in eq. (3.104); the sensitivity is significant along the subhorizontal paths of the adjoint fields. But, for gradient of the Hamiltonian with respect to δ there is a considerable difference between two parametrizations. In (v_v, ϵ, δ) parametrization the sensitivity of the δ gradient, eq. (3.105), is more along intermediate reflection angles while for (v_{nmo}, v_h, δ) parametrization this sensitivity is mostly for vertical propagations. This is why, from mathematical and pragmatical point of view, during the inversion for (v_{nmo}, v_h, δ) parametrization we prefer to keep δ as a passive parameter to remove the cross-talk between v_{nmo} and δ. In Fig. 3.40(d-f) we plot the gradients at first iteration of a joint inversion for v_{nmo}, v_h and δ (the subsurface model is inferred from the test in Fig. 3.39(e1-e3)). These gradients highlight the difference between the discussed weights for two parametrizations. In Fig. 3.40(f) the white zone on the δ kernel are the clipped values because of the high positive amplitudes. Here, we use the same scaling factor for δ which has been used in Fig. 3.40(c). We keep the color-scales in Fig. 3.40 in a same range to emphasize the difference between two parametrizations. For this example we observe the relative amplitude of gradient with respect to v_{nmo} is higher than v_h and the sensitivity pattern for v_{nmo} and δ are same.

Now we compare the results obtained with (v_{nmo}, v_h, δ) parametrization with those inferred from (v_v, ϵ, δ) parametrization for the example in Fig. 3.41(g1-g5) where δ is a passive parameter and v_v, ϵ and X_{sct} are the optimization parameters. Here, we consider v_{nmo}, v_h and X_{sct} as the optimization parameters. We apply the relations $v_{nmo} = v_v\sqrt{1+2\delta}$ and $v_h = v_v\sqrt{1+2\epsilon}$ to convert the true v_v, ϵ and δ models to true v_{nmo} and v_h models. The true v_v is a constant-gradient background with a circular anomaly at the center, and the true models for ϵ and δ are two homogeneous models. This implies the true v_{nmo} and v_h models include a constant-gradient background and a circular anomaly at the center (Fig. 3.44s1-s2). In the following the pertur-

bation models for v_{nmo} and v_h are defined as the difference between final updated and the true background models for v_{nmo} and v_h.

We initialize v_{nmo} and v_h by their true background models (i.e. true models without the circular anomalies) and X_{sct} with true positions. Fig. 3.45(t1-t5) shows the final perturbation models of v_{nmo}, v_h, δ and X_{sct} components for joint inversion in (v_{nmo}, v_h, δ) parametrization. The updated v_{nmo} anomaly is underestimated at the center, and because of the acquisition layout and vertical sensitivity of gradient with respect to v_{nmo}, there is a vertical smearing for the anomaly. The gradient with respect to v_h is sensitive to wide reflection angles and this leads to smearing of the updated v_h anomaly along some oblique adjoint ray tubes (Fig. 3.45t2). But, lack of scatterers with wide reflection angles results in the underestimation of v_h and horizontal perturbation for X_{sct}. This horizontal mispositioning of the scatterers also contributes in erroneous v_h updates. On the other hand, enough illumination with vertical propagations well resolves the ambiguity between v_{nmo} and vertical positioning. Because of v_{nmo} and v_h decoupling in (v_{nmo}, v_h, δ) parametrization, the errors in the updated velocities mostly relates to cross-talk between the velocities and X_{sct}.

Fig. 3.44(q1-q2) shows the v_{nmo} and v_h models inferred from the v_v and ϵ model and the background delta of the test in Fig. 3.41(g1-g2) where we performed a joint inversion of v_v, ϵ and X_{sct}. Both parametrizations result in a similar v_{nmo} anomaly but the updated v_h and X_{sct} are different. The artefact on updated ϵ in (v_v, ϵ, δ) parametrization (Fig. 3.41g2) generates an overestimated v_h at the centrer of anomaly (Fig. 3.44q2) and some artefacts at the deep part.

We also compare the updated ϵ and v_v in both parametrizations. We apply the relations $v_{nmo} = v_v\sqrt{1+2\delta}$ and $v_h = v_v\sqrt{1+2\epsilon}$ to derive v_v and ϵ from the updated v_{nmo} and v_h in (v_{nmo}, v_h, δ) parametrization (Fig. 3.45z1-z2). Compared to the updated v_v in (v_v, ϵ, δ) parametrization, the calculated v_v anomaly from v_{nmo} is a closer estimation of the true anomaly. However, the calculated ϵ from the updated v_h and v_{nmo} in (v_{nmo}, v_h, δ) parametrization results in some negative values for ϵ (Fig. 3.41z2, here negative values are clipped).

Therefore we can conclude, for a suitable parametrization one should consider two issues: 1) the acquisition layout, and 2) the requested parameters. When we need estimation of v_{nmo} in a reflection data set with short azimuth acquisition, a parametrization which is more sensitive to the short-offset reflections, like (v_{nmo}, v_h, δ) is favourable. However, this parametrization is inappropriate if ϵ model is requested as the final imaging product. Here, v_{nmo} and v_h have mostly a vertical and horizontal sensitivity, respectively. This results in a poor estimation of ϵ through $\epsilon = ((v_h/v_{nmo})^2(1+2\delta) - 1)/2$. In other words, we use two quantities (v_{nmo}, v_h) without any overlap on their sensitivity kernels to extract another parameter (this is another reason for poor reconstruction in Fig. 3.41g2). Concerning this issue, (v_{nmo}, η, δ) parametrization can be reliable in ϵ reconstruction because v_{nmo} has a full aperture sensitivity.

3.4.3 Conclusion

We extend the adjoint slope tomography to anisotropy through a factored TTI eikonal solver in forward modelling and a matrix-free inversion scheme based on the adjoint state method. The eikonal solver provides a uniform sampled traveltime map where there is no shadow zone while

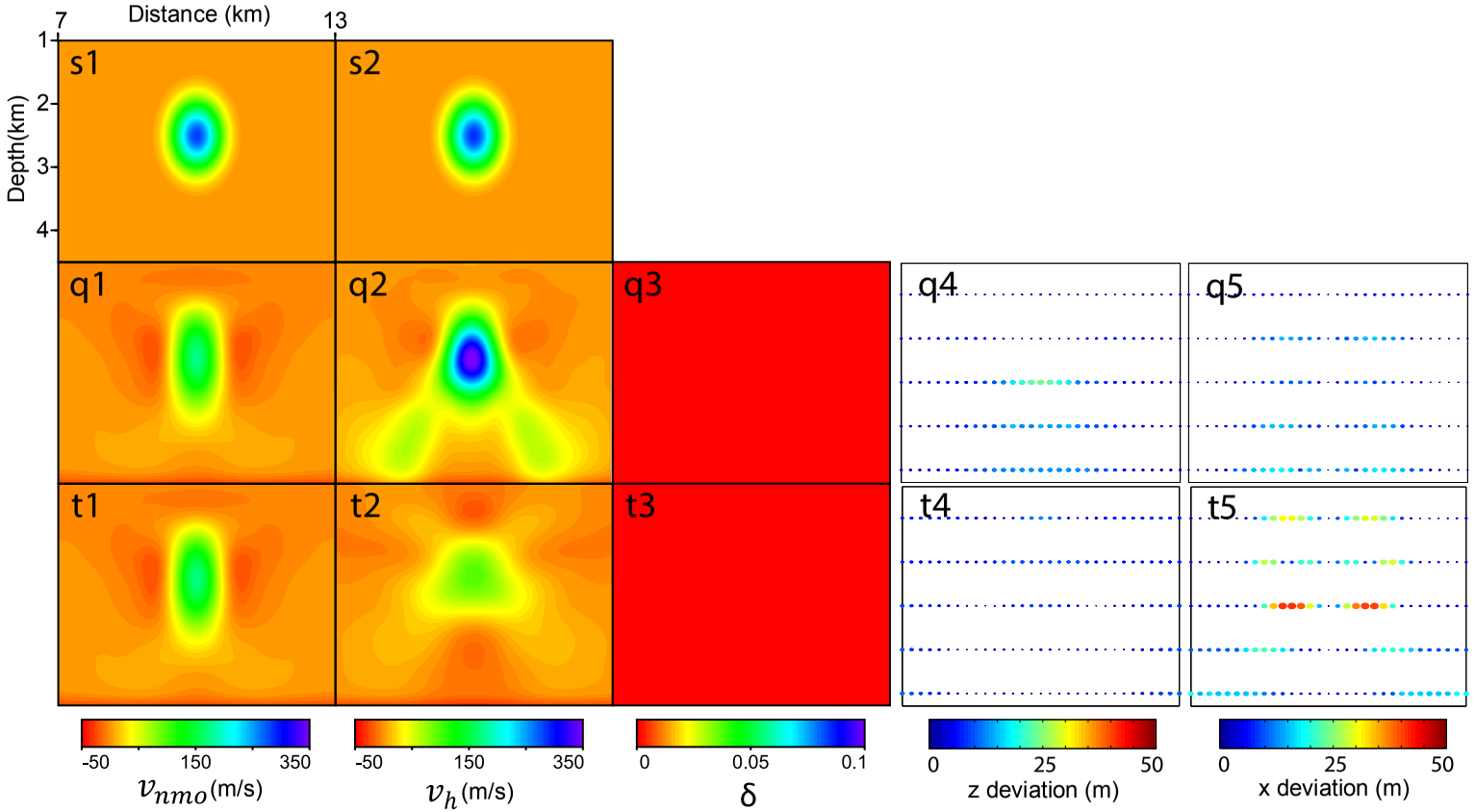

Figure 3.44 – Parametrization test. True circular anomalies of s1) v_{nmo} and s2) v_h for the example in Fig. 3.41(g1-g5). Difference maps for updated v_{nmo}, v_h, δ, z and x components of X_{sct} (respectively form first to fifth column) in joint inversion for parametrization q1-q5) (v_v, ϵ, δ) and t1-t5) (v_{nmo}, v_h, δ). For both tests the only passive parameter is δ.

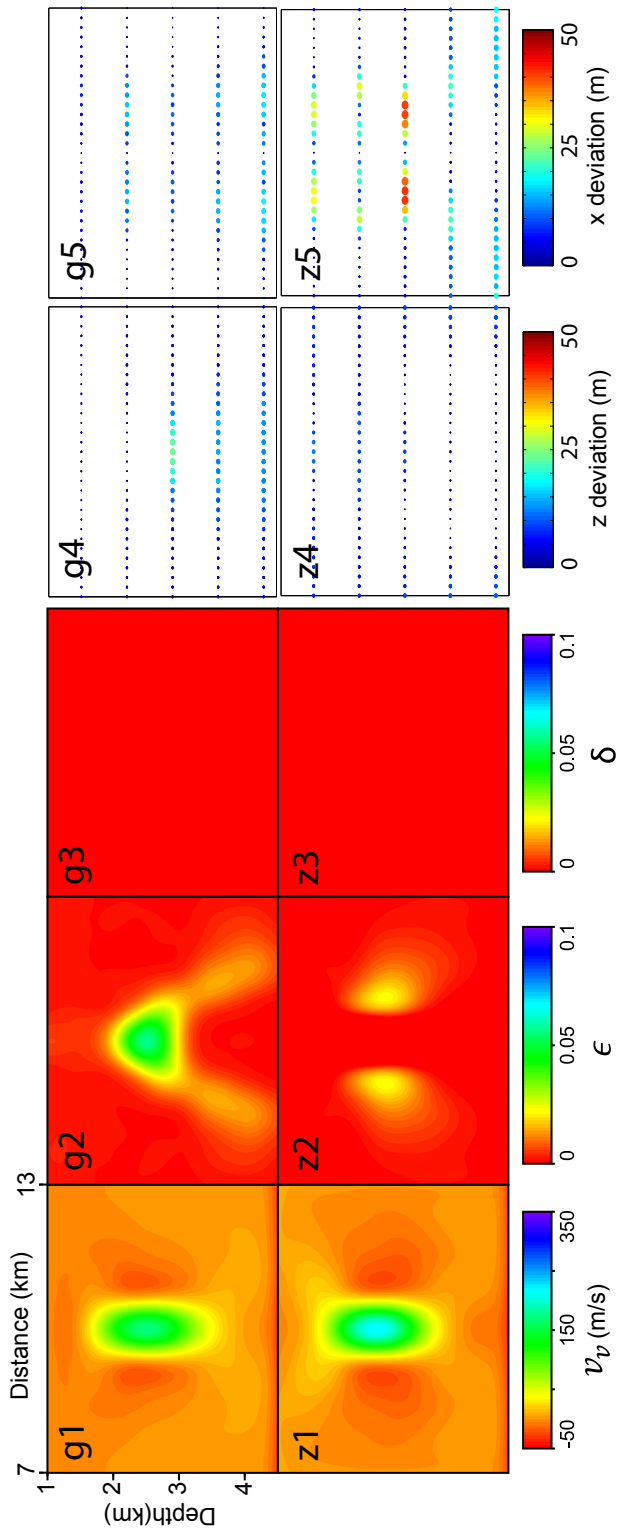

Figure 3.45 – Parametrization test. Difference maps for updated v_v, ϵ, δ, z and x components of X_{scl} (respectively form first to fifth column) in joint inversion for parametrization g1-g5) (v_v, ϵ, δ) and z1-z5) (v_{nmo}, v_h, δ). For both tests the only passive parameter is δ.

the implemented factorization method efficiently removes singularity of the point source. Instead of ascending regime for wave propagation in classic formulation, we solve the eikonal equation from source and receiver positions. By considering the reciprocity rule, we propose to calculate the slopes at source and receiver positions by applying a finite difference operator on traveltime maps from neighbouring sources and receivers of the source and receiver under question. This leads to a computational cost which depends on the number of sources and receivers rather than scatterer numbers. In this framework the model space includes the subsurface parameters (i.e. vertical velocity and anisotropy parameters) and scatterer positions, and the observables are scatterers two-way traveltimes and slopes at the source and receiver positions. We apply the L-BFGS algorithm in our multi-scaling inversion scheme to inexpensively exploit the Hessian information and mitigate the leakage between different parameter classes. However, scaling of parameters is crucial in avoiding convergence toward a local minimum. Through a realistic example, TTI BP salt model, we show that a dense and wide azimuth acquisition provide enough illumination for vertical velocity and epsilon reconstruction. The future extensions of this method include applying some regularization terms in the misfit function to reduce the leakage between parameters, updating jointly the transmission and reflection arrivals, and investigating other parametrizations like Voigt parameterization.

Chapter 4

Real data application

Chapter overview: In this chapter we present application of anisotropic slope tomography to a real data set. The seismic data is a 2D towed-streamer dataset from north-west of Australia and was provided by CGG. The picking procedure is performed by CGG in common image gathers and observables are deduced in the data domain by demigration. In the following, we first discuss about the provided data and picks, and since our method is based on picks in the unmigrated data domain (two-way traveltimes and slopes at source and receiver positions), we present an approach to transform the picks from migrated domain to data domain. Then, we compare the CGG velocity model with velocity model inferred from anisotropic slope tomography where the Thomsen's parameters δ and ϵ have been used in a passive way. As a quality control of our results we perform VTI reverse time migration with the CGG and our velocity models to assess the flatness of the common image gathers and focus of energy on final migrated image.

4.1 2D BroadSeis data

The 2D towed-streamer real data set is provided by CGG and acquired in the north-west of Australian continent (Fig. 4.1), Carnarvon Basin. The region under exploration includes several gas reservoirs at depths greater than 3 km in Triassic Mungaroo formation. This interval consists of the fluvial sandstones and non-marine brackish siltstone. The discovered gas accumulations in the Mungaroo formation are both in fluvial section and the marine-reservoirs and the deeper Locker Shale is interpreted as the source rock. On the top of the Mungaroo formation there are Brigadier formation and the Murat siltstone which include several rotated horst and graben structures due to the continental break up of the North Western Australian margin. These horst blocks form the hydrocarbon traps (Wellington, 2016). There are many reports which explain geology and exploration history in this basin. For a complete list of these publications the reader is referred to the Australian government-Geoscience Australia website (http://www.ga.gov.au/). A discussion on the geological interpretation of the imaging results is out of the scope of our

study, that will focus on the methodological aspects of the slope tomography (inversion set-up and quality control of the results from the imaging perspective).

The acquisition line length is approximately 57 km and there are three wells nearby the line (Fig. 4.1) which only include the P-wave velocity information. This acquisition consists of 2479 shots and a streamer of 8250 m length. The source and hydrophone intervals are ≈ 25 and 12.5 m, respectively, and the sources are at 5 m depth while the depth of hydrophones varies from 8 m to 57.5 m (BroadSeis technology). Compared to the conventional flat streamer acquisition, the BroadSeis data (Fig. 4.2) includes a broader frequency bandwidth, $2.5 - 155$ Hz (Fig. 4.4), and a higher signal to noise ratio for deep targets provided by the recording of low frequencies. Moreover, the curved shape of the streamer helps to remove more efficiently the receiver ghosts (http://www.cgg.com/).

In addition to the picks, CGG provided us the processed data for migration (Fig. 4.3) where the maximum offset is 7.2 km, the multiples are eliminated and the first arrival muted. During this processing the low frequencies (< 2.5 Hz) are boosted and some notches from frequency spectrum are removed (Fig. 4.4). The other provided information by CGG are the well logs (only include P-wave velocity), water depth which varies from 100 m to more than 1.2 km, and a 2D TTI model updated by non-linear slope tomography (Fig. 4.5).

4.2 Adjoint slope tomography in practice

Our formulation for the adjoint slope tomography is based on picks in the unmigrated data domain (two-way traveltimes and slopes at source and receiver positions) where we have directly access to the kinematic information and for each source-receiver pair there are several picks. But, the provided picks by CGG are from depth migrated domain and therefore we need some considerations before inserting them to our algorithm. In the following, we briefly review the advantages and disadvantages of depth-migrated domain picking. Then, we explain how we deal with these picks and finally we discuss some practical aspects and solution of adjoint slope tomography for the real data.

4.2.1 Picks from depth-migrated domain

Formulation of the classic slope tomography and also our method is based on picking of local coherent events in a pre-stack unmigrated domain. The main problems of picking in this domain are the low signal to noise ratio of the coherent events and lack of control on the distribution of the picks in depth (Lambaré, 2008; Chauris et al., 2002a; Billette et al., 2003; Lambaré et al., 2004). These deficiencies result in erroneous picks, uneven subsurface illumination by scatterers, and consequently poor reconstruction for subsurface parameters. Among the alternative picking approaches there are picking in depth-migrated domain (Chauris et al., 2002a; Nguyen et al., 2008), picking in pre-stack time migrated (PreSTM) domain (Lambaré et al., 2007) and picking in post-stack time domain (Neckludov et al., 2006; Lavaud et al., 2004).

Pre-stack depth migrated (PreSDM) domain (Fig. 4.6) is the most natural domain for velocity model building (Lambaré, 2008). Thanks to the focused energy on CIGs (Fig. 4.7),

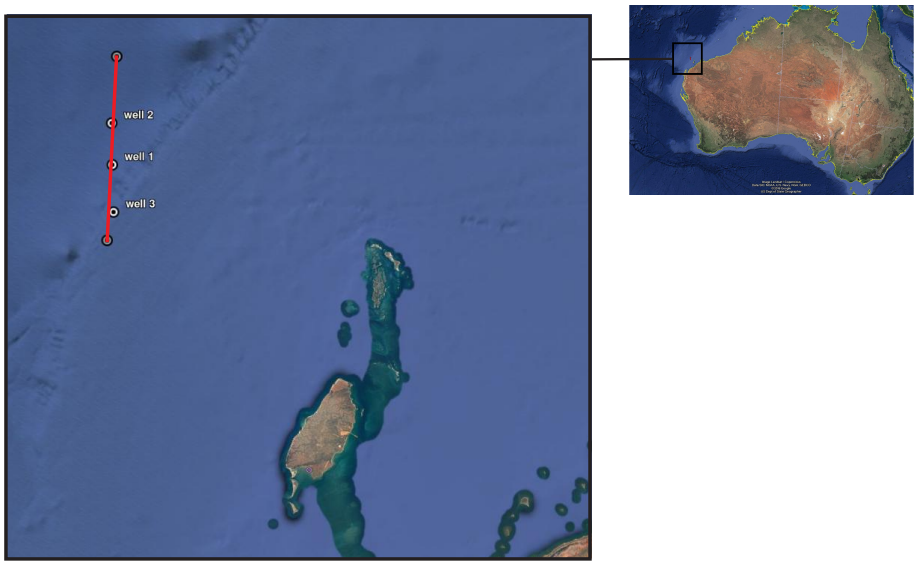

Figure 4.1 – Location of 2D BroadSeis data from north-west of Australia continent, Carnarvon basin. Length of the line is around 60 km.

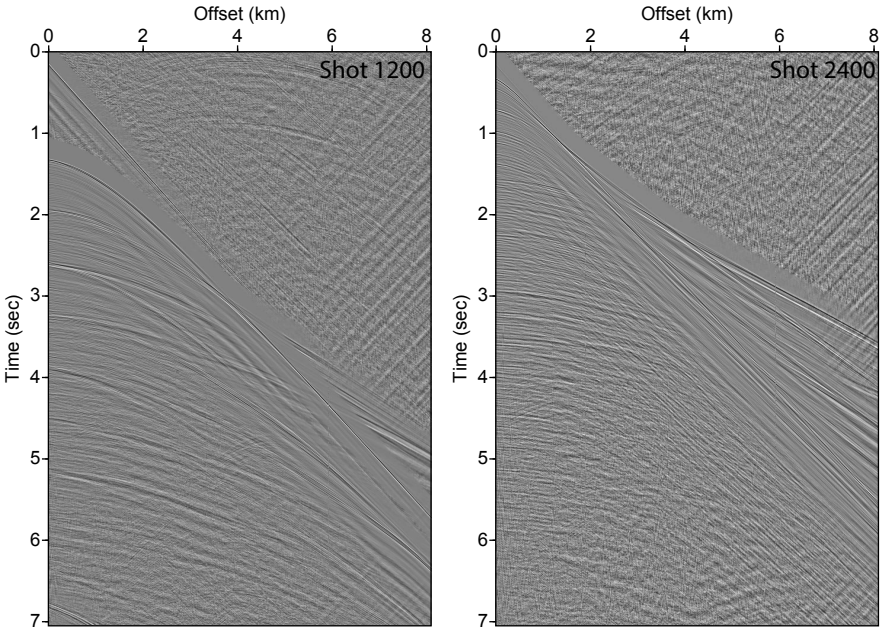

Figure 4.2 – Raw common-shot gathers after applying automatic gain control: shot number 1200 and 2400. Maximum offset is 8.2 km.

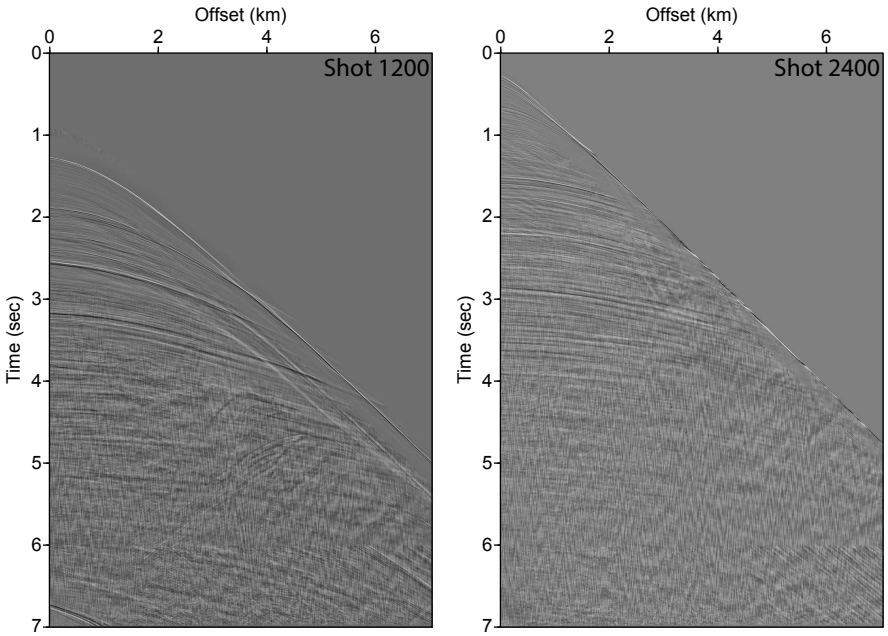

Figure 4.3 – Common-shot gathers after preprocessing and automatic gain control: shot number 1200 and 2400. Maximum offset is 7.2 km.

picking local coherent events in CIGs provides more reliable information and denser sampling of medium with a control on the pick distributions in depth. In pre-stack depth migrated cube, the local coherent events are characterized with their positions (distance and depth), offset, and slopes on common migrated offset gather and common image gather (Fig. 4.7) (Chauris et al., 2002a). The slope in the common-offset domain represents the structural dip and the slope in the common image gather is the residual move out. Chauris et al. (2002a,b) demonstrated the direct link between local coherent events in the depth-migrated and unmigrated time domains. Through demigration and migration one can extract the coherent event attributes in each of these domains from the other. Demigration can be defined as the inverse of the migration process where the reflection energy is back traced to the recording domain (time-offset domain) (Santos et al., 2000). Therefore, as long as the subsurface parameters for migration and demigration process are same, by applying demigration on the depth-migrated domain picks we can retrieve the kinematic invariants (i.e. two-way traveltimes and slopes at the source and receiver positions) (Guillaume et al., 2008).

In an ideal setting (e.g. a homogeneous medium without noise) demigration exactly retrieves the observed kinematic in the seismic records. However, in real application due to the demigration/migration artefacts, noise and picking errors, perfect retrieval of the kinematic information, in particular the source and receiver positions, is difficult. Therefore, the calculated kinematic invariants from depth migrated picks can be contaminated with errors. Moreover, since in the migration process the ties between arrival energies and their associated source-receiver is lost, the retrieved positions for the sources and receives for each pick can differ from the true source and receiver positions.

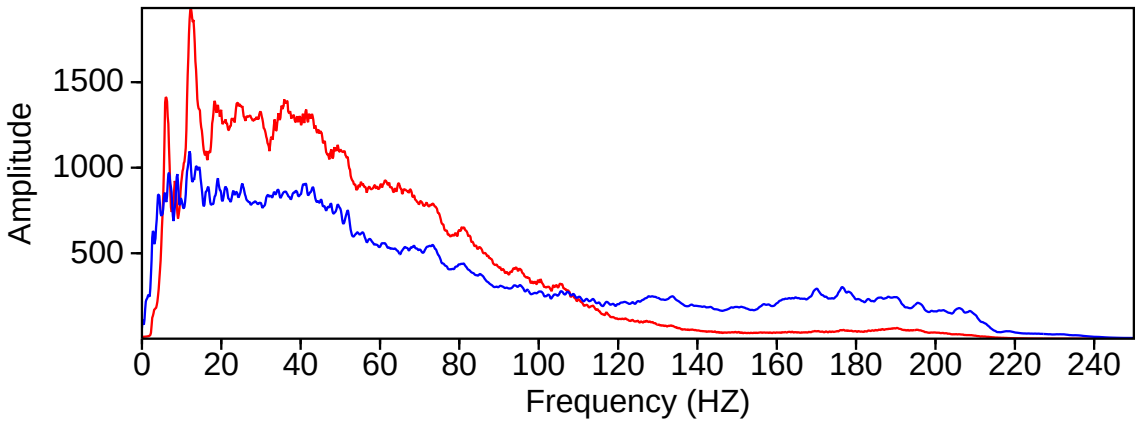

Figure 4.4 – Frequency bandwidth for shot 1200. Red designates the bandwidth for raw common-shot gathers (Fig. 4.2) and blue for processed one (Fig. 4.3).

As an alternative to demigration, Nguyen et al. (2008) proposed deriving the kinematic invariants for picks in depth migrated domain through the migration of attributes. Compared to demigration, they showed migration of attributes results in a less erroneous kinematic invariants. As another picking method, (Lambaré et al., 2007) introduced picking in pre-stack time migrated domain where the demigration procedure is easier than pre-stack depth migration domain and there is no compromise on available kinematics.

4.2.2 Dealing with kinematic invariant from demigration

The provided picks by CGG are the output of a semi-automatic picking tool in PreSDM cube and the kinematic invariants are calculated by demigration. As mentioned above, demigration dedicates to each picked event a source-receiver pair which does not necessarily coincide with the source-receiver pair of the real acquisition. Therefore, the retrieved source and receiver positions are not regularly distributed (Fig. 4.8). Moreover, picks that should have been assigned to a single source-receiver pair are now associated to specific source-receiver pairs. This issue is not a problem for classic stereotomography, but it reduces dramatically the performance of our algorithm. A part of our method efficiency is related to the number of required forward modelling which is equal to $N_s + N_r$ where N_s and N_r stand for the number of distinct shot and receiver gathers, respectively. In our approach we suppose for each source-receiver pair there are more than one pick (which in the framework of picking in pre-stack time domain is totally natural) and in results the number of specific sources-receivers is far less than total picks. Therefore, the provided PreSDM picks, which implies the number of specific source-receiver pairs is in the same range of picks number, makes application of our method costly.

In order to cope with this problem, we design an aggregation scheme for which we define a hypothetical regular distribution of sources and receivers. In Fig. 4.8 we schematically present the retrieved sources and receivers by demigration which do not respect any regular acquisition

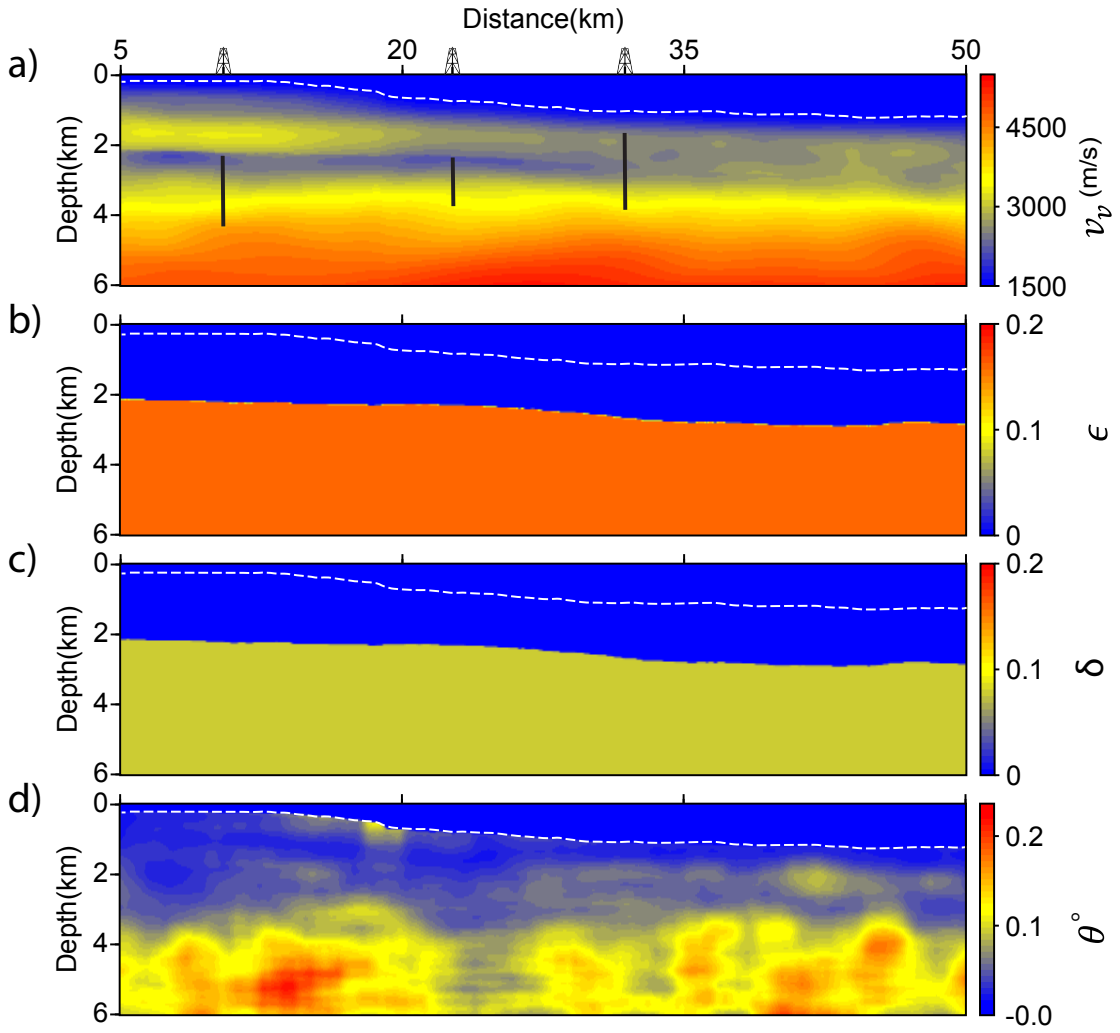

Figure 4.5 – 2D TTI model by CGG: a) v_v, b) ϵ, c) δ and d) θ. The black line segments are the covered depth by the wells. The white dashed line designates the seabed.

layout. We consider a maximum value for deviation from the hypothetical sources and receivers, and then we explore the whole set of picks to identify those whose associated source-receiver pair has a distance less than the maximum deviation from a hypothetical source-receiver pair. These source-receiver pairs on the hypothetical acquisition are the new positions for the defined sources and receivers by demigration. Density of this virtual acquisition defines the number of picks we choose from all the PreSDM picks.

The total number of picked events by CGG is more than $2,000,000$. Here in our aggregation scheme, we define a regular distribution of sources/receivers with 100 m interval, and we choose the maximum deviation from hypothetical sources and receiver of 5 m. We do think this deviation in the source and receiver positions is negligible at the scale of tomographic methods. Consequently, there is around 550 unique positions for the sources and receivers. This scheme results in more than $50,000$ picked event which it is approximately 2.5% of all the provided

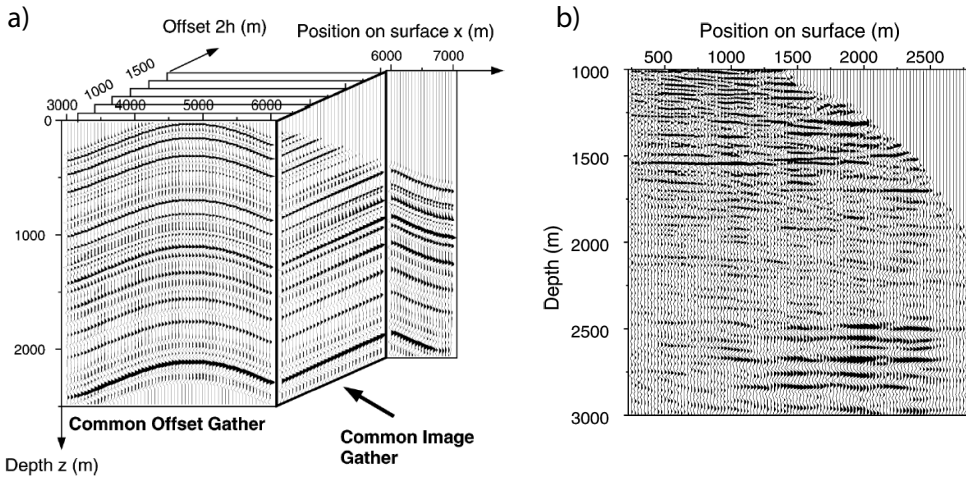

Figure 4.6 – a) Migrated volume by prestack depth migration of common-offset gathers. b) CIG from a real data set (Figure from Chauris et al. (2002a)).

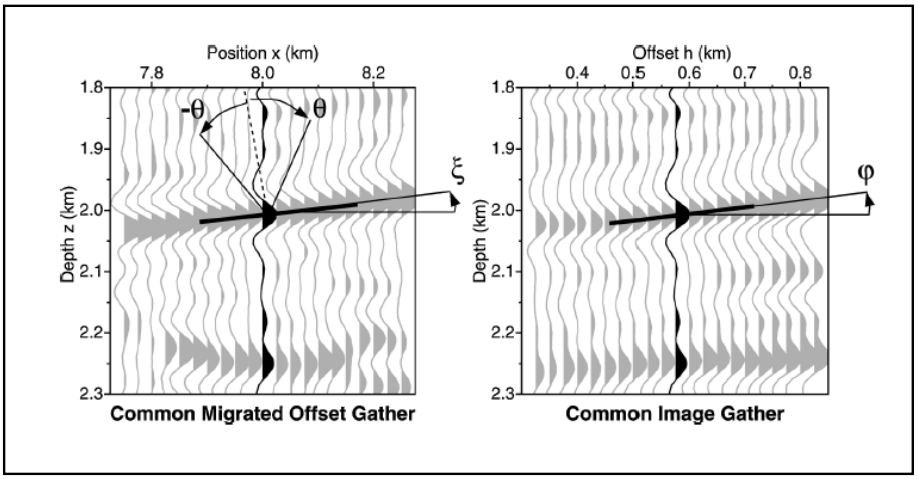

Figure 4.7 – Each local coherent event in prestack depth migrated domain characterized by (x, z, h, ξ, φ) (figure from Chauris et al. (2002a))

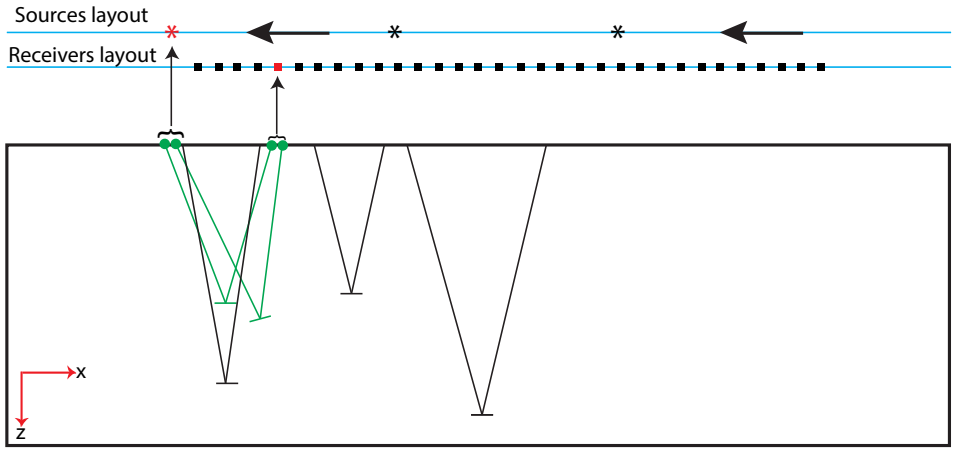

Figure 4.8 – Schematic sketch of source-receiver positioning by demigration. Within an aggregation scheme on the retrieved sources and receivers by demigration, we cluster some of them (green •) to a close source or receiver (red ■ and $*$) position on a regular distribution of sources-receivers .

picks by CGG.

Fig. 4.9(a) shows distribution of offsets for all the picks provided by CGG. The maximum offset is around 7.2 km and picking is performed approximately for every 150 m in the offset dimension of PreSDM cube for a selected common image point. The histogram in Fig. 4.9(a) is skewed left where over 51% of the picks have less than 1 km offset, 25% have an offset between 1 and 2 km, 14% with offset between 2 and 3 km and only 10% include offsets more than 3 km. Lack of picks from long offset reflections makes the reconstruction of epsilon challenging and because of that here we only try to invert for the vertical velocity. Through our aggregation scheme the distribution of selected picks is less skewed left (Fig. 4.9b) but we lose some offsets (because of the offset interval of the original picks). A hierarchical inversion which prioritises the contribution of picks according to their offsets is a tool to manage more efficiently different offset ranges (Prieux et al., 2013; Billette, 1998). In this study we do not consider any hierarchical inversion and we directly use the selected picks by our aggregation scheme (Fig. 4.9b).

4.2.3 Implementation and solution of the adjoint slope tomography

As mentioned before, the inversion includes three steps: initialization of the scatterers and subsurface models, localization and finally joint inversion of model parameters in a multi-scale scheme. Here, we aim to update v_v and scatterer positions while during inversion ϵ, δ and θ are fixed to CGG models (Fig. 4.5b-d). Here, the covariance matrix is defined by $(\sigma_{tt}, \sigma_{ps}, \sigma_{pr}) = (5 \times 10^{-3}, 10^{-6}, 10^{-6})$. In order to regularize the inversion, we apply a Gaussian smoothing filter on gradient with respect to v_v with a window length of 500 m in both directions.

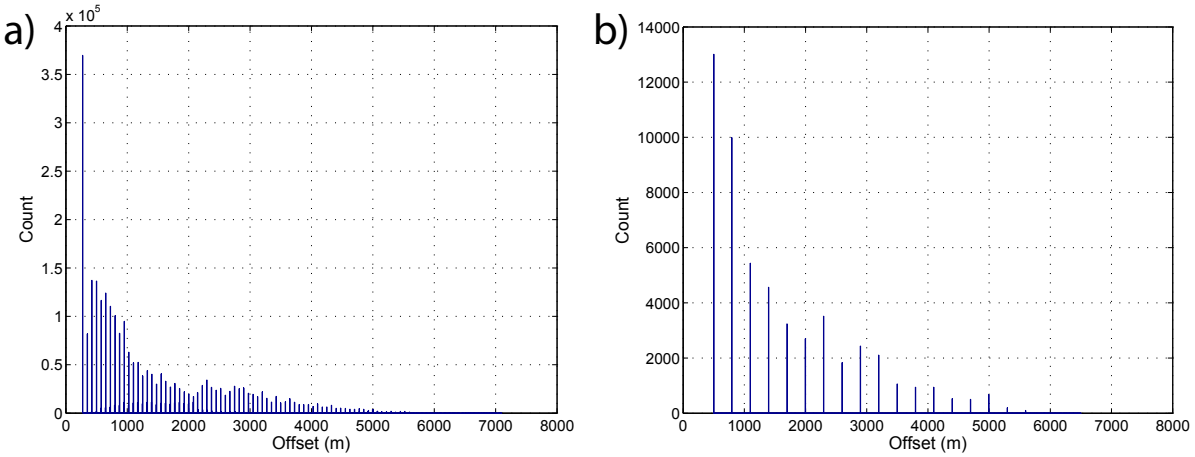

Figure 4.9 – Distribution of picks (scatterers) with respect to their source-receiver pair offsets for a) all the picks from CGG and b) selected picks by our aggregation approach.

In order to define the initial position for each scatterer, we consider two rays initiated from associated source and receiver such that they honour the observed slopes. By considering homogeneous background model, these rays are two straight lines whose total traveltime should be the picked two-way traveltime. The intersection point of these rays defines the initial position for the scatterer (details are in Appendix of Billette et al. (2003)). This implies that the initial positions only depend on the observables and they are independent of the initial v_v model that we shall define. This initialization approach in case of complex geological structure for subsurface can not propose a reliable initial positions for all the scatterers. For example, if a scatterer has observed source and receiver slopes of the same sign, this approach will not find any intersection between the two straight rays. But, this issue seems less important for this dataset in which there are not sharp dipping reflectors. Although, picking in depth migrated cube can provide the initial position of the scatterers, we did not have access to this information.

Fig. 4.10(a) shows the initial positions of the scatterers inferred from the above-mentioned approach. We consider the scatterers above seabed as noise and remove them from inversion. In the shallow-right part there is a downward shift for the scatterers. This can be attributed to the rapid velocity variation in this region while the initialization is based on the ray trajectories in homogeneous media. Fig. 4.11 schematically illustrate this problem: we define the initial position by observable supposing a homogeneous velocity on the background while the true velocity is increasing with depth. This issue affects less those scatterers on the shallow-left part because the homogeneous water above them is consistent with the assumption of the initialization.

We initialize the v_v by a constant gradient velocity model (Fig. 4.10b), i.e. $v = v_0 + a \times z_{abs}$, where z_{abs} is the depth from the seabed, $v_0 = 1500$ and $a = 0.5$. The only reason for this choice is based on the distribution of the scatters after localization step. A background velocity is favourable in which after localization the scatterers are not pulled/pushed abnormally toward top/bottom of the model. We choose a velocity-gradient background model because, compared to a homogeneous background, it mitigates the move of the scatterers toward the top or bottom of the model after the localization step. We limit the number of iterations in the localization

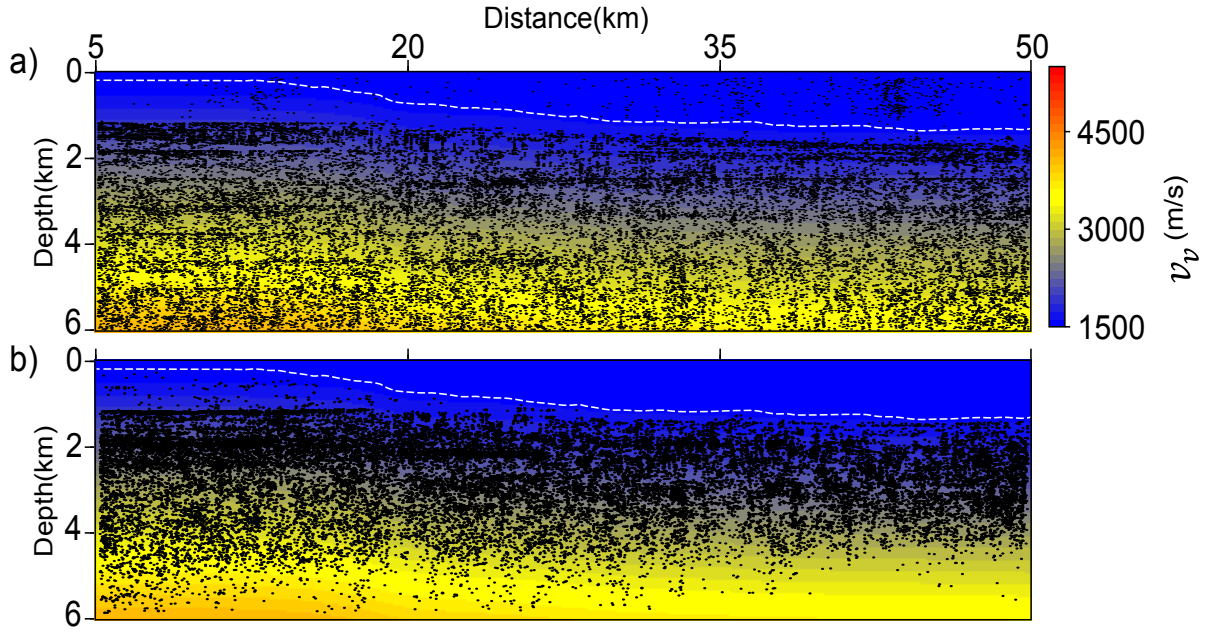

Figure 4.10 – BroadSeis data application. Velocity model and scatterer positions after (a) initialization and (b) localization. White dashed curve shows the seabed.

step to 10 to prevent convergence of the localization step toward a local minimum, which could result from the inaccuracy of the velocity-gradient background model. This stopping criterion of iteration is estimated by trial and error. Fig. 4.10(b) shows the scatterer positions after localization.

After localization we perform the joint inversion of v_v and scatterer positions. We consider six multi-scale steps with vertical and horizontal spline node intervals decreasing from 670 m to 80 m and from 2000 m to 60 m, respectively. The total number of L-BFGS iterations is 230 (Fig.4.14). For scales 5 and 6 we use the wells information as a hard constraint to ensure a correct velocity update at well locations. This constraint restricts the updated values to a range the width of which is scaled to the distance from the well; the nodes far from the wells are allowed to vary over a wider range and the nodes which are close to the wells vary within a narrower range. This constraint is only considered for nodes located at a maximum distance of 1 km from the wells. If we apply these local constraints in the initial scales the inversion converges to a local minimum. In first scales, inversion updates low frequencies of subsurface while wells include the high frequency and local information.

Fig. 4.12(a) and (c) show the final inverted velocity and scatterer positions by anisotropic slope tomography, respectively. In order to have a more quantitative assessment, in Fig. 4.13 we compare the logs which are extracted every 5 km from our inverted velocity and the CGG velocity model in Fig. 4.12. Due to lack of illumination near the end of the picked zone, we can not accurately update the velocity before 5 km and after 50 km distances. Lack of scatterer also affects the deep part of the model specifically after $x = 35$ km. The erroneous velocity leads to an upward shift for the scatterer in these regions (Fig. 4.12c). However, the updated velocity

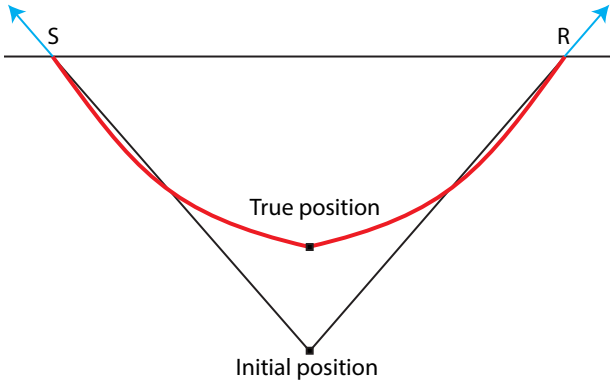

Figure 4.11 – Initial position for a scatterer is calculated by the straight rays (black lines) only based on the observables and this can push/pull the positions far from the true positions where there is fast varying velocity in the background. Red curve is true ray paths on the true velocity (schematic presentation).

model in general is close to the CGG velocity model and the low velocity zone at $2-3$ km is well recovered. In the following, we compare the CIGs computed by reverse time migration (RTM) for the CGG and our anisotropic slope tomographic models.

4.3 RTM with slope tomographic model

As a quality control of the tomographic results, we perform a VTI reverse time migration (RTM) of the BroadSeis data using the initial, CGG and our final slope tomography vertical velocity models presented in Fig. 4.10 and Fig. 4.12(b-a), respectively, while for all the migrations the Thomsen's parameters are the provided model by CGG (Fig. 4.5)

Here we use TOYxDAC-TIME (from SEISCOPE tools) to perform RTM. The goal is to build the CIGs for the initial, CGG and the slope tomography vertical velocity model and compare the flatness of the reflectors for two velocity models. In order to build the CIGs, we performed several migrations for all the shots such that in each migration only a specific range of offsets is considered. This results in a series of common-offset migrated images from which it is easy to extract common-image gathers in the depth-offset domain at arbitrary surface positions. Here the offset ranges are defined from $0.17-5$ km with 150 m increment.

Fig. 4.15 presents the calculated CIGs and the migrated section from VTI reverse time migration with the initial vertical velocity. Non of the reflectors on the CIGs are flat (Fig. 4.15a) and this means the initial vertical velocity model is far from the correct subsurface velocity. In result, the migrated section (Fig. 4.15b) includes unfocused reflectors with wrong positioning. Fig 4.16 shows a collection of CIGs calculated for the slope tomographic (top) and the CGG vertical velocity models (bottom). Comparison of the calculated CIGs with the initial velocity model (Fig. 4.15a) and those CIGs with adjoint slope tomography (Fig 4.16a), reveals clearly the improvements with updated tomographic vertical velocity model. The CIGs for both CGG

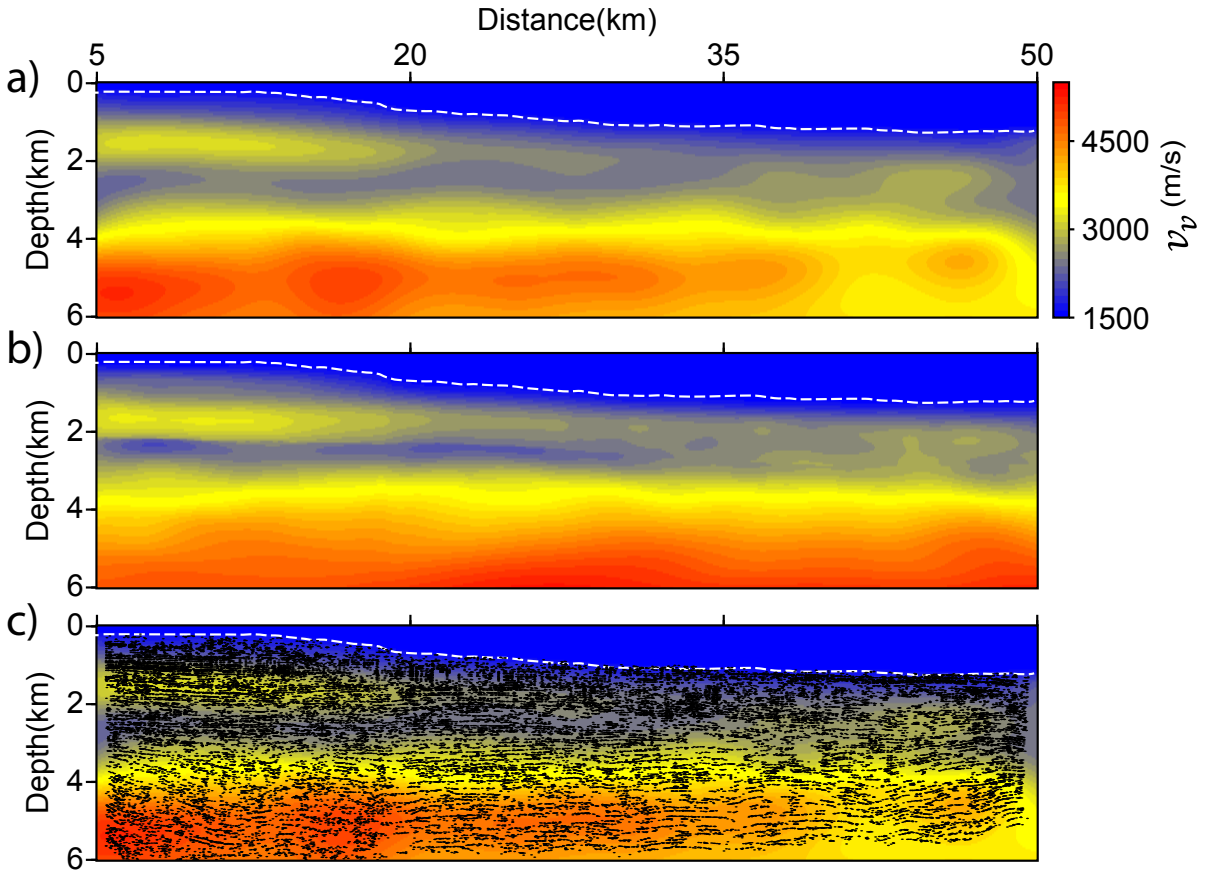

Figure 4.12 – BroadSeis data application. a) Final velocity model obtained by our anisotropic slope tomography. b) CGG velocity model. c) Superimposition of the final updated scatterers on our final velocity model (a). The white dash line delineates the sea bed.

and slope tomography velocity models represent acceptable and similar flatness for the reflectors. In Fig 4.16 we draw some ellipsoids which highlight some major differences between two CIG sets. However, both velocity models still need improvement to flatten some reflectors. This might be a result from the fact that the Thomsen's parameters have not been updated during the inversion.

Fig. 4.17 shows the final depth migrated images obtained by stacking the different offset contributions ($0.17-7.2$ km). In migrated section with slope tomographic velocity (Fig. 4.17a), after the distances > 48 km and for the deep part with distances < 10 km, the erroneous velocities shift the reflectors upward. Some of the differences between the two migrated sections are highlighted by two ellipsoids where we observe more continuous reflectors in the migrated section with slope tomographic model. However, because of the repeated deep faulted structures, the continuity in depth does not necessarily means a more accurate velocity model.

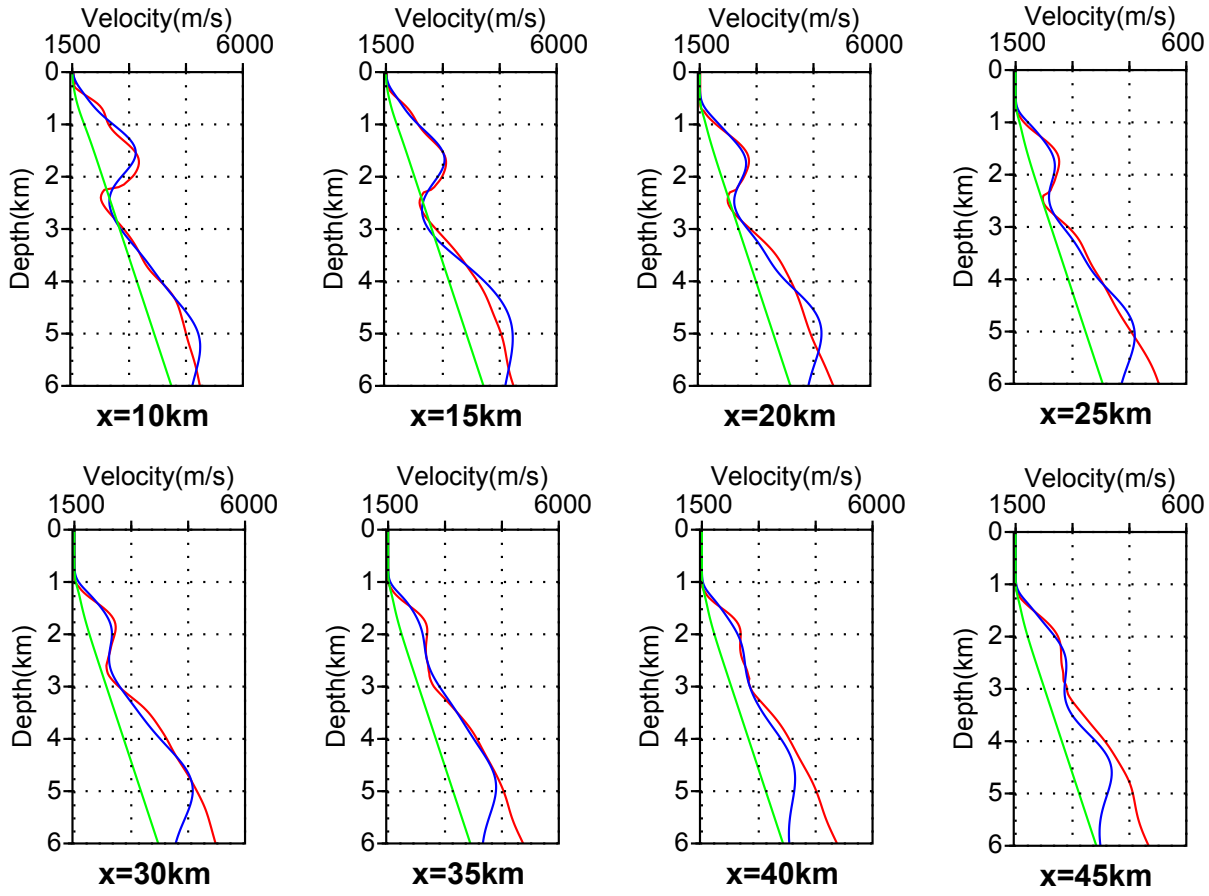

Figure 4.13 – BroadSeis data application. Direct comparison between the CGG velocity model (red) and the inverted velocity by anisotropic slope tomography (blue). The green line represents the initial velocity model.

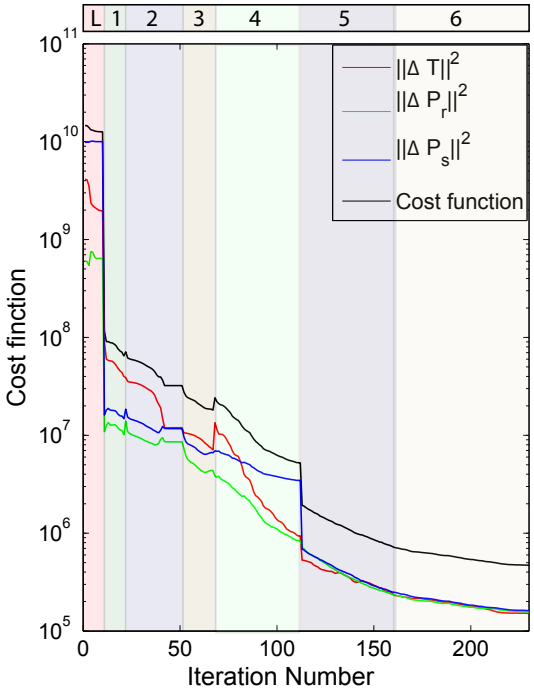

Figure 4.14 – BroadSeis data application. Convergence Diagram in logarithmic scale. There are 6 scales within 230 iterations. Diagrams designates the residuals for: traveltime (red), slope at source (blue), slope at the receiver (green), and total cost function (black). Summation of residual curves is the total cost function. On top letter "L" shows the localization step and numbers are the scales.

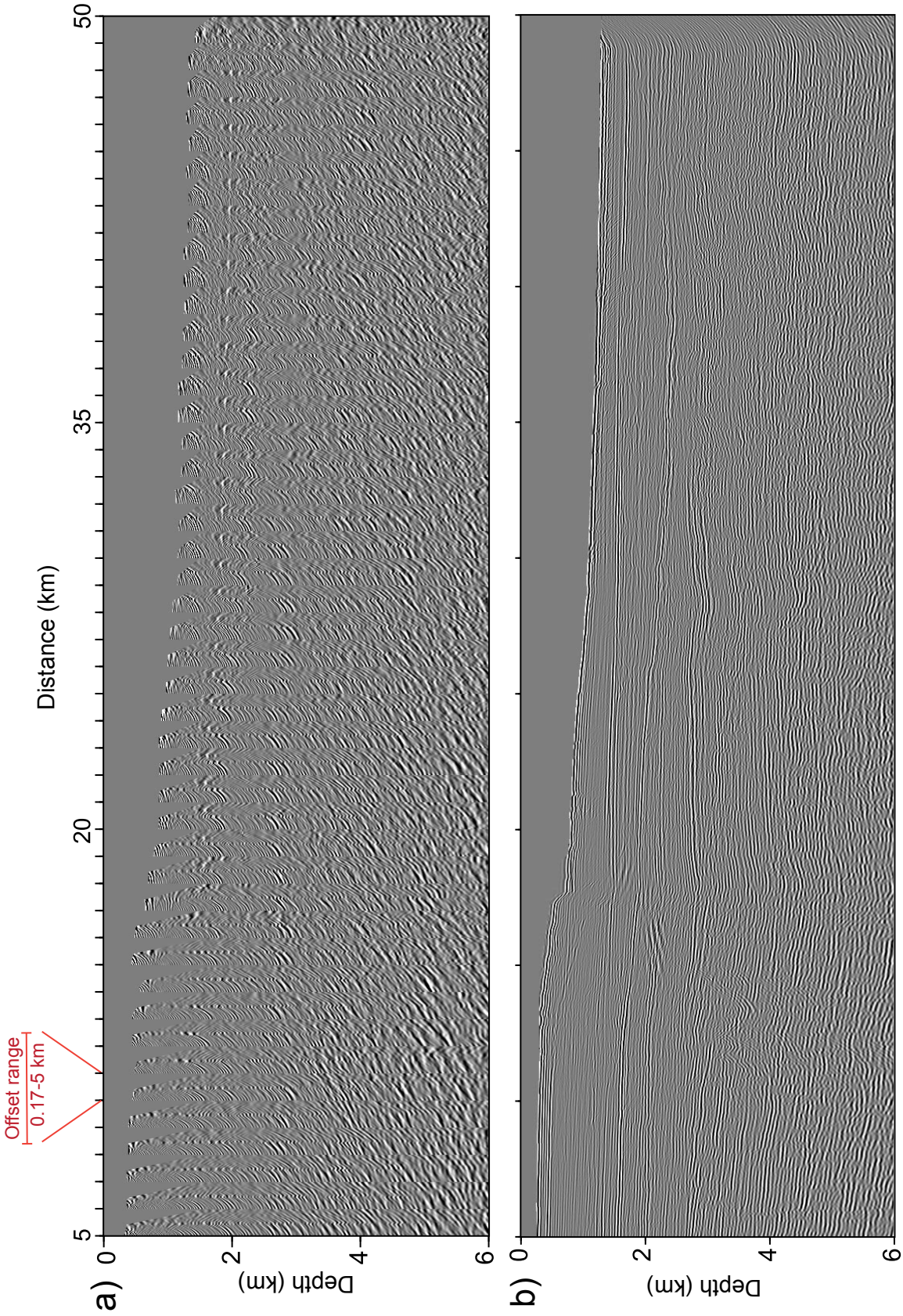

Figure 4.15 – BroadSeis data application. a) CIGs and b) migrated section from VTI reverse time migration with the initial vertical velocity model. The offset range for each CIG is $0.17 - 5$. Here CIGs start from 5 km to 50 km with the interval of 1 km.

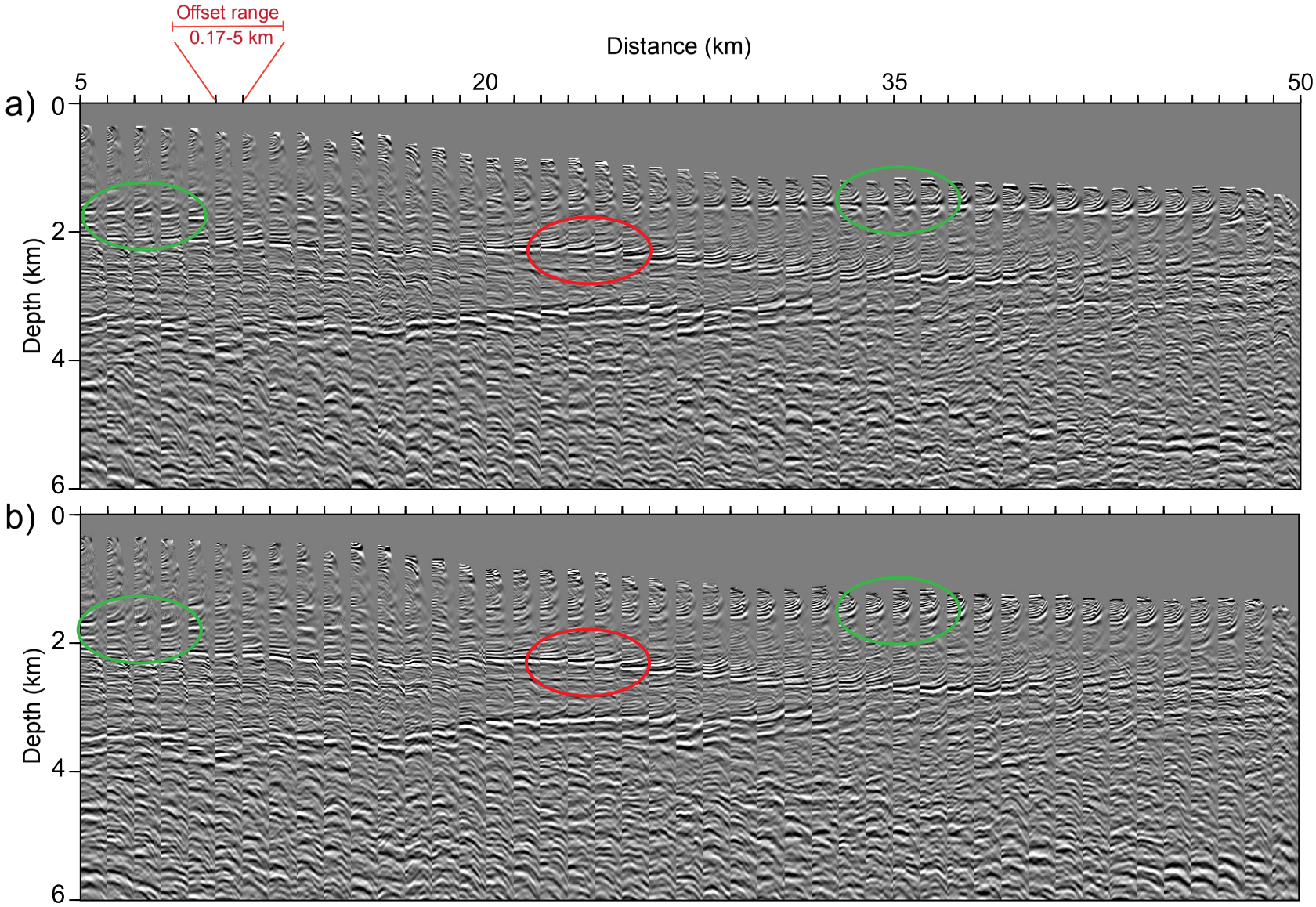

Figure 4.16 – BroadSeis data application. CIGs from VTI reverse time migration with the vertical velocity model from a) slope tomography and b) CGG. The offset range for each CIG is $0.17 - 5$. Green ellipsoid highlight some differences between two models. The green ellipsoids show some improvements and the red ellipsoid a slight deterioration. Here CIGs start from 5 km to 50 km with the interval of 1 km.

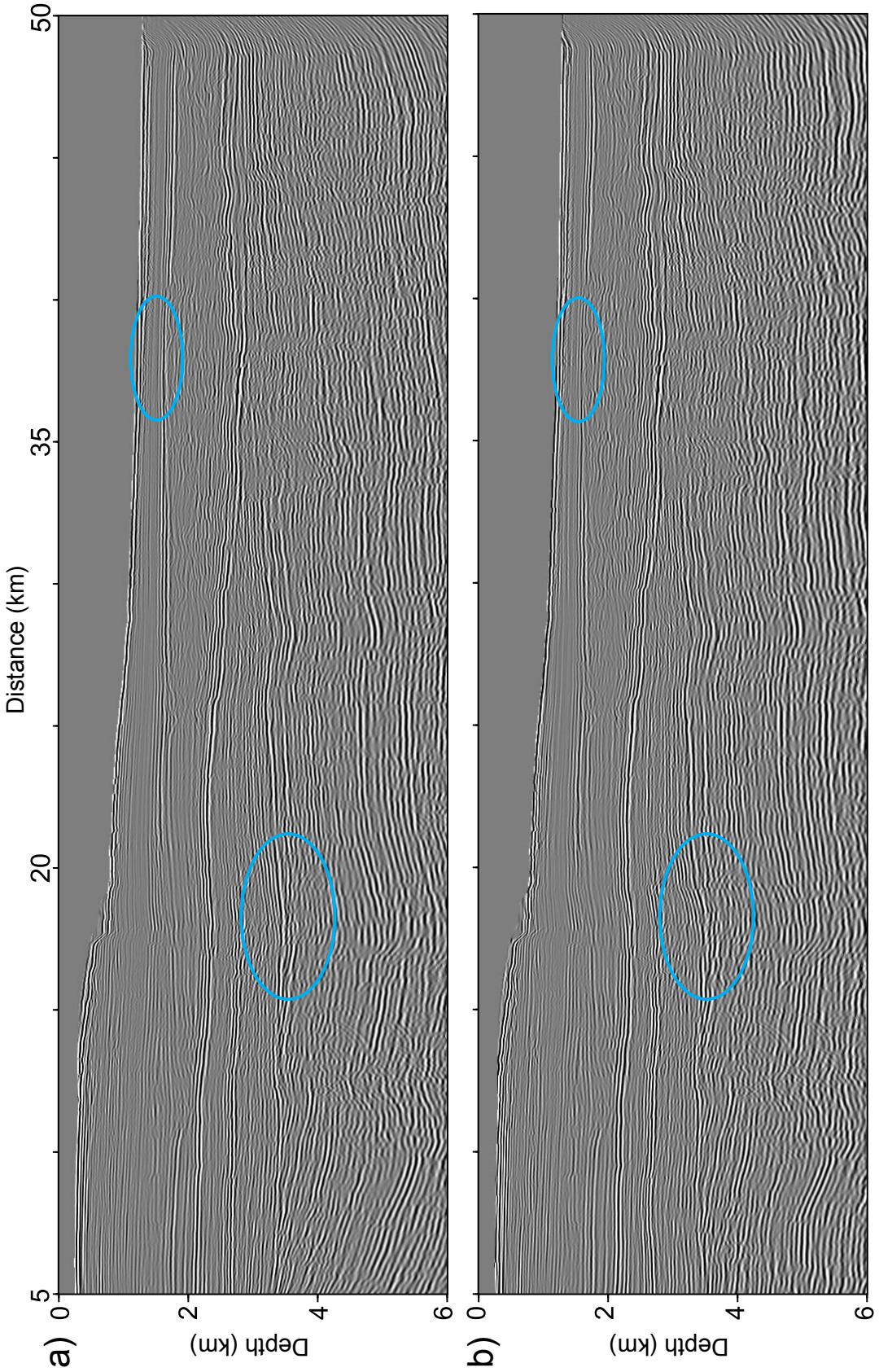

Figure 4.17 – BroadSeis data application. Migrated sections from VTI reverse time migration with a) slope tomography vertical velocity and b) CGG vertical velocity model. The blue ellipsoid highlight some key differences.

4.3.1 More picks or not?

Do more picks help to improve the inverted velocity model? To answer this question we performed some other tests on the provided picks. Selection of the picks within our aggregation scheme plays an important role in the final updated model. More picks necessarily does not improve the results. For real data sets, feeding the inversion with more picks can insert more noise in the inversion. For the cases that we aim to find the best solution by a single run of code (which is the case for the discussed results), according our experiences, we prefer to give preference to the selection of the picks rather than their quantity. But, in case of some consecutive implementations of the algorithm, one can use as many picks as possible and after converging to a solution, perform depth migration and re-pick the depth migrated cube, then redo the slope tomographic inversion for the new picks. In our real data set example we did not try to improve the updated vertical velocity model by using such a technique.

4.4 Conclusion

We present application of anisotropic slope tomography on a 2D offshore data. The picking is performed in depth migrated domain and the kinematic invariants are extracted by demigration. Since the retrieved source and receiver positions by demigration are not respecting a regular distribution, we design an aggregation scheme to cluster the them on a hypothetical regular distribution. Lack of long-offset reflections prevents us from reconstruction of ϵ model, and we update only vertical velocity. The updated slope tomographic method flattens well the reflectors on CIGs which we extracted from RTM.

Chapter 5

Conclusion and perspectives

5.1 General conclusion

I developed the "2D anisotropic adjoint slope tomography" approach as a new framework for slope tomography based on a reformulation of the classic stereotomography (Billette, 1998; Billette et al., 2003; Lambaré, 2008). In this method, I replace the ray tracing based forward modelling engine of the classic formulation with a 2D TTI eikonal solver to obtain a well sampled first-arrival traveltime map for complex anisotropic media and long-offset acquisitions. In the inverse problem, I calculate the misfit function gradients through the adjoint state method establishing a matrix-free framework for the inversion without explicit calculation of the Fréchet derivative matrix.

Forward modelling: The eikonal equation is solved from the source and receiver positions (descending propagation regime) which scales the overall computational cost of the method to the number of distinct sources and receivers rather than scatterers. The strategy to solve the 2D TTI eikonal equation is based on the separation of the elliptic and anelliptic terms in which resolution of the elliptic eikonal equation is achieved through the fast sweeping method while the fixed point iterations update the solution of the elliptic eikonal equation using the anelliptic term as a right-hand side. I also apply a factorization technique which efficiently removes the singularity at the point source and improves considerably the precision of the solutions. After resolution of the eikonal equation from the sources and receivers, a sampling operator extracts the one-way traveltimes at the scatterer positions and build the two-way traveltimes. In chapter 2 I assessed the accuracy of the proposed eikonal solver for different long-offset models including the 2D TTI BP-salt model. Compared to the ray based slope tomographic approaches, application of eikonal solvers in the forward modelling is more costly but we think the descending propagation regime in our formulation partially compensates the extra cost.

Calculation of slopes at the source and receiver positions is based on the finite difference approximation. These slopes are calculated from the neighbouring source and receiver traveltime maps. The reciprocity between source-scatterer and receiver scatterer allow the finite difference operator to sample the traveltime maps at the scatterer positions which are far from the singular points of the traveltimes maps, i.e. the source and receiver positions. As long as

there is no caustic in the traveltime maps, this simple and efficient approach results in enough accurate slopes in smooth isotropic and anisotropic media.

Isotropic inverse problem: As a direct consequence of applying eikonal solvers in the forward modelling, the model space definition in the adjoint slope tomography does not include the take-off angles as the eikonal solvers consider propagations in all the directions. This leads to a model space which consists of the scatterer positions and subsurface velocities. Moreover, because the real source and receiver positions are known and can be implemented in the eikonal solver, I do not need to consider them among the observables. Therefore, the data space includes only the two-way traveltimes and the slopes at the sources and receivers. Considering a l_2-norm misfit function between the observables and modelled data, I apply the L-BFGS optimization method to iteratively update the model parameters while mitigating the leakage between them by taking into account the Hessian information in a cheap way. Moreover, this approach takes advantage of a multi-scaling strategy, during which the subsurface velocity model is progressively refined, to handle the ill-posedness of the inverse problem. I use the adjoint state method to calculate the gradient of the cost function with respect to the model parameters. This requires mainly resolution of two adjoint state equations which back-propagate the traveltime and slope residuals from the scatterer positions toward the sources and receivers along the adjoint ray tubes. This leads to two adjoint fields which build the sensitivity kernels of the gradient with respect to the velocity. Gradient of misfit function with respect to the each scatterer position is derived by applying the spatial derivative of the sampling operator on the associated source-receiver pair and their neighbouring traveltime maps. Different examples, including velocity model estimation for the smooth marmousi model, assess the capability of the method in velocity macro-model building.

Extension to TTI media: Extension of adjoint slope tomography to TTI media involves considering the anisotropic properties of the subsurface among the model parameters. After solving the TTI eikonal equation from the sources and receivers, the traveltimes and the slopes can be achieved in the same way as isotropic formulation. The calculation of the misfit function gradient with respect to model parameters, again mainly require to determine two adjoint fields. I showed that, the sensitivity kernel of the subsurface parameters are the weighted summation of these adjoint fields and the properties of these kernels (such as sensitivity to the horizontal and vertical propagations) is directly linked to the definition of these weights. In the context of Thomsen parametrization of the subsurface, I investigate the cross-talk between different parameter classes. The sensitivity of vertical velocity has a full aperture sensitivity (i.e. sensitivity to the all propagation directions) and δ and ϵ are mostly sensitive to the mid-range and horizontal propagations, respectively. The leakage between the vertical velocity and δ prevents a reliable update for joint inversion of both parameters while this leakage between ϵ and vertical velocity is less significant. Updating ϵ requires the long wavelength propagations and in absence of such an acquisition (and prior information) the ambiguity between the horizontal positioning of the scatterers and ϵ is not resolvable. The leakage between the vertical velocity and ϵ can be better controlled if one uses a parametrization which decouples the sensitivity to the vertical and horizontal propagations. However, the leakage between the scatterer positions and the subsurface parameters plays a major role in the update of the subsurface parameters and this should be controlled through regularization of the scatterer positions.

Through a realistic synthetic example (TTI BP salt model), I showed that the adjoint slope tomography, for the fixed values of tilted angle and δ during the inversion, accurately updates the vertical velocity and retrieves a close estimation of the true ϵ in the shallow part (where I defined wide reflection angles for the scatterers). Moreover, a real data application shows the potential of the proposed method in velocity macro-model building for the real acquisition settings.

5.2 Perspectives

Here I mention the main perspectives for the future extensions of the adjoint slope tomography which are under investigation. Therefore, I only address the key points of these extensions. The applications and validations of the proposed extensions are among our future work-plans and publications.

5.2.1 Regularization of the scatterer positions

During the joint inversion of the subsurface parameters in anisotropic adjoint slope tomography, the updated parameters are considerably affected by the leakage between the scatterer positions and subsurface parameters (compare Figs 3.39 and 3.41) and a regularization technique can be a tool to control these cross-talks. In the context of classic stereotomography, Costa et al. (2008) proposed the structurally motivated smoothing constraint in the direction of potential reflectors, and Barbosa et al. (2008) regularized the solutions by minimizing the heterogeneity and anisotropy of the subsurface and maximizing the diffractor focusing.

Inspired by their suggestion of Barbosa et al. (2008) for maximizing the diffractor focusing, here I propose regularization of the scatterer positions in the framework of Tikhonov regularization (Tikhonov, 1963) for the picks from pre-stack depth migrated domain.

All the local coherent events associated with a specific reflector in a CIG belong to the same CIP. Therefore, in adjoint slope tomography the desired inversion solution for the scatterer positions is the one that all the scatterers with the same CIP are positioned close to each other. This suggests regularizing the scatterer positions such that those with the same CIP are forced to coincide. This requires recognizing all the picked event with the same CIP during the picking in pre-stack depth migrated cube. This provides us with a powerful tool to control the positioning of the scatterers which are related to the same CIP.

If I index the CIP locations with i as $\mathbf{X}_i$, the position of scatterer $n_{s,r}$ associated with ith CIP can be considered as $\mathbf{x}_{i,n_{s,r}}$. By defining a suitable function, $\mathbf{L}$, which measures the distribution of the scatterers around their associated CIP, I can rewrite the adjoint slope tomographic misfit

function (5.13) as

$$
\begin{aligned}
C(\mathbf{m}) &= \frac{1}{2\sigma^2_{T_{s,r}}} \sum_{s=1}^{N_\mathbf{s}} \sum_{r=1}^{N^s_\mathbf{r}} \sum_{n_{s,r}=1}^{N^{s,r}_\mathbf{n}} (T_{s,r,n_{s,r}}(\mathbf{m}) - T^*_{s,r,n_{s,r}})^2 \\
&+ \frac{1}{2\sigma^2_{p_s}} \sum_{s=1}^{N_\mathbf{s}} \sum_{r=1}^{N^s_\mathbf{r}} \sum_{n_{s,r}=1}^{N^{s,r}_\mathbf{n}} (p_{s,n_{s,r}}(\mathbf{m}) - p^*_{s,n_{s,r}})^2 \\
&+ \frac{1}{2\sigma^2_{p_r}} \sum_{r=1}^{N_\mathbf{r}} \sum_{s=1}^{N^r_\mathbf{s}} \sum_{n_{s,r}=1}^{N^{s,r}_\mathbf{n}} (p_{r,n_{s,r}}(\mathbf{m}) - p^*_{r,n_{s,r}})^2 \qquad (5.1) \\
&+ \beta \sum_{s=1}^{N_\mathbf{s}} \sum_{r=1}^{N^s_\mathbf{r}} \sum_{n_{s,r}=1}^{N^{s,r}_\mathbf{n}} \mathbf{L}(\mathbf{X}_i, \mathbf{x}_{i,n_{s,r}}), \qquad (5.2)
\end{aligned}
$$

where β is the regularization parameter to be set. Since during the picking we do not have access to the true CIP positions (since the migration velocity is not the true velocity) one can consider the position of the local event with the shortest-offset as a floating position for the associated CIP (similar to PreSDM reflection tomography where a floating datum is defined for the reflectors, section 1.2.1) and define $\mathbf{L}$ as the l_2-norm distance function ($\mathbf{L}(\mathbf{X}_i, \mathbf{x}_{i,n_{s,r}}) = (\mathbf{X}_i - \mathbf{x}_{i,n_{s,r}})^2$) which measures the distance between the floating CIPs and their associated scatterers. During the inversion this floating CIP positions update and allows for converging toward the true positions.

This definition of the cost function modifies the expression of gradient with respect to the scatterer positions as follow

$$
\begin{aligned}
\nabla_{\mathbf{x}_{i,n_{s,r}}} C =& \frac{Q_{n_{s,r}}}{\partial \mathbf{x}_{n_{s,r}}} \Big(\mu_{s,r,n_{s,r}} (\mathbf{t}_s + \mathbf{t}_r) \Big) \\
&+ \frac{Q_{n_{s,r}}}{\partial \mathbf{x}_{i,n_{s,r}}} \Big(\frac{\xi_{s,n_{s,r}}}{2\Delta s} (\mathbf{t}_{s+1} - \mathbf{t}_{s-1}) \Big) \\
&+ \frac{Q_{n_{s,r}}}{\partial \mathbf{x}_{i,n_{s,r}}} \Big(\frac{\xi_{r,n_{s,r}}}{2\Delta r} (\mathbf{t}_{r+1} - \mathbf{t}_{r-1}) \Big) \\
&- 2\beta (\mathbf{X}_i - \mathbf{x}_{i,n_{s,r}}).
\end{aligned} \qquad (5.3)
$$

Moreover, the distance between the scatterers and their floating CIP can be a reliable tool to recognize the noisy scatterers and remove them from the inversion process; one can consider a threshold for the maximum distance and filter out the scatterers with greater distance than threshold. However, I should admit that finding the associated CIP for each local coherent is not straightforward and may arise the debate over interpretative picking and its difficulties.

5.2.2 First-arrival adjoint slope tomography

Current developments in seismic acquisitions provides large-offset data sets and recording refracted waves with large penetration depths. This promotes the idea of joint inversion of re-

flection and refracted information in seismic imaging techniques (Korenaga et al., 2000; Prieux et al., 2010; Zhou et al., 2015).

Our proposed slope tomographic approach is based on local coherent events associated with reflected waves and this can be extended to incorporate also the first-arrival local coherent events (which may include direct, diving and head waves events) to the inversion. The joint inversion of first-arrival and reflected local coherent events has been introduced by Prieux et al. (2010) in the framework of the classic stereotomography to build an initial velocity model for full waveform inversion. They considered that the first-arrival picks are linked to some scatterers with large scattering angles ($> 130°$). This allows them to use the same forward modelling engine for both reflection and first-arrival traveltime and slope calculations. The refracted waves may sample those parts of the subsurface with a deficit of reflection picks and compensate the lack of reflection paths for the subsurface parameter estimations. Moreover, long offset data with deep penetration for diving waves, provide more horizontal propagation paths for retrieving ϵ parameter in anisotropic media. Moreover, Gosselet et al. (2003, 2004); Gosselet and Bégat (2009) introduce the idea of using the slope of transmission arrivals, as well as their traveltime, for walk-away transmission stereotomography, borehole positioning and joint inversion of stereotomographic picks and picks form borehole data.

In this section, I develop first-arrival adjoint slope tomography for TTI media in which one can take advantage of traveltime and slopes of the first-arrivals in shot and receiver gathers to update the subsurface parameters (Fig. 5.1). The finial goal is to implement this tool in the anisotropic adjoint slope tomography and perform joint inversion of the first-arrival and the reflection picks. This extension for adjoint slope tomography is quite interesting because in the forward modelling by eikonal solvers the first-arrival traveltimes are calculated and there is no extra calculation in the forward modelling if I want to perform joint inversion of the first-arrival and reflection picks. I only need to solve some extra adjoint state equations which in the following I shall discuss them. Here, I only explain the formulation of first-arrival adjoint slope tomography for 2D TTI media; extension to the joint inversion scheme with reflection picks is straightforward.

The first arrival traveltime tomography (FATT) based on the eikonal solvers and the adjoint state method has been fully developed for isotropic and anisotropic media (Taillandier et al., 2009; Waheed et al., 2016). Here, I extend these methods such that account for both traveltime and slopes of the local events in the subsurface parameter reconstructions. Compared to the conventional FATT methods, here the extra constraint from the slopes may mitigate more the non-linearity of inverse problem.

Model and data space definition

In first-arrival adjoint slope tomography the data space $\mathbf{d}$ consists of

$$\mathbf{d} = (T_{s,r}, p_s, p_r)\,|_{s=1}^{N_\mathbf{s}}|_{r=1}^{N_\mathbf{r}^\mathbf{s}}, \tag{5.4}$$

where $T_{s,r}, p_s$ and p_r are the observed traveltime and slopes at the source and receiver positions for a pair of local first-arrival event with central source and receiver s and r, respectively. $N_\mathbf{s}$ is

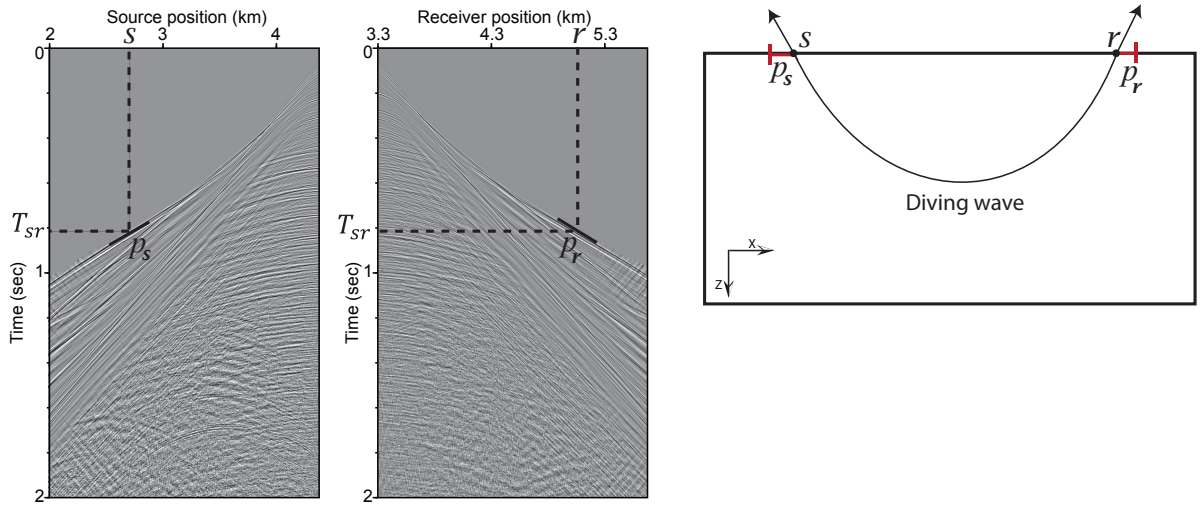

Figure 5.1 – Local coherent events of first-arrival phases in a shot and receiver gather. Here the labelled events are generated by diving waves. Each local coherent event is characterized with its central source-receiver (s, r), traveltime $(T_{s,r})$ and slopes at the source (p_s) and receiver (p_r) positions.

the number of distinct shot gathers and $N_{\mathbf{r}}^{\mathbf{s}}$ is the number of distinct receivers in the shot gather s. Note that here for each source-receiver pair there is only one pair of picks and there is no concept of scatterer in the physics of the wave propagations (Fig. 5.1).

Therefore, the model space includes only the velocity parameters

$$\mathbf{m} = \left(\{c_{i,m}\}|_{i=1}^{4}|_{m=1}^{M},\right)^{t}, \tag{5.5}$$

where $c_{i,m}$ for $i = 1 \rightarrow 4$ represent the cubic cardinal B-spline coefficients for the v_v, ϵ, δ and tilted angle θ, respectively.

Forward modelling

Through a similar approach that I used in the anisotropic slope tomography, I solve the 2D TTI eikonal equation from the sources and receiver positions

$$H(\mathbf{x}, \nabla t_s(\mathbf{x})) = 0, \tag{5.6}$$
$$t_s(\mathbf{x}_s) = 0 \tag{5.7}$$

and

$$H(\mathbf{x}, \nabla t_r(\mathbf{x})) = 0, \tag{5.8}$$
$$t_r(\mathbf{x}_r) = 0 \tag{5.9}$$

where $t_s(\mathbf{x})$ and $t_r(\mathbf{x})$ are the first-arrival traveltime maps originated from source s and receiver r, respectively, and H stands for the Hamiltonian representation of 2D TTI eikonal equation (eq. 3.84). After solving these equations for source s and receiver r, I calculate the first-arrival traveltime $T_{s,r}$ for the pair (s, r) through

$$\begin{aligned} T_{s,r} &= Q_{r_s}(\mathbf{t}_s) = Q_{s_r}(\mathbf{t}_r) \\ &= \frac{1}{2}(Q_{r_s}(\mathbf{t}_s) + Q_{s_r}(\mathbf{t}_r)), \end{aligned} \tag{5.10}$$

Here Q_{r_s} and Q_{s_r} are the sampling operators which extract the first-arrival traveltime at rth receiver of shot gather s, and at sth source of receiver gather r, respectively. The equality (5.10) is the direct consequence of applying the reciprocity rule between sources and receivers, and it is written in the form of mean value only for consistency between the formulation in first-arrival and reflection adjoin slope tomography. In order to calculate the slopes at the source and receiver positions I use the same approach I applied in the adjoint slope tomography (section 3.2.4) but with a slight modification. Considering the reciprocity between the source and receiver, I use the following finite difference operator to calculate the horizontal component of the slowness at the source and receiver

$$p_{s_r} = \frac{\partial t_s(\mathbf{x}_s)}{\partial x_s} \approx Q_{r_s}(\mathbf{t}_{s+1} - \mathbf{t}_{s-1})/2\Delta s, \tag{5.11}$$

$$p_{r_s} = \frac{\partial t_r(\mathbf{x}_r)}{\partial x_r} \approx Q_{s_r}(\mathbf{t}_{r+1} - \mathbf{t}_{r-1})/2\Delta r. \tag{5.12}$$

Here Δs and Δr are the source and receiver intervals, respectively. Since the finite difference operator is sampling the traveltime maps far from the singular points (i.e. source and receiver positions) and the subsurface model is smooth, this results in accurate slope calculations. Here the main difference with the slope estimation in section 3.2.4 is the sampling operator definition. In adjoint slope tomography the sampling operator samples the traveltime maps at the scatterer positions while here sampling is directly applied at the source and receiver positions. Note that, considering the reciprocity between the source and receiver, I can also calculate the slopes at the source and receiver positions directly by applying the finite difference operator on the $\mathbf{t}_r$ and $\mathbf{t}_s$ at the source s and receiver r positions, respectively.

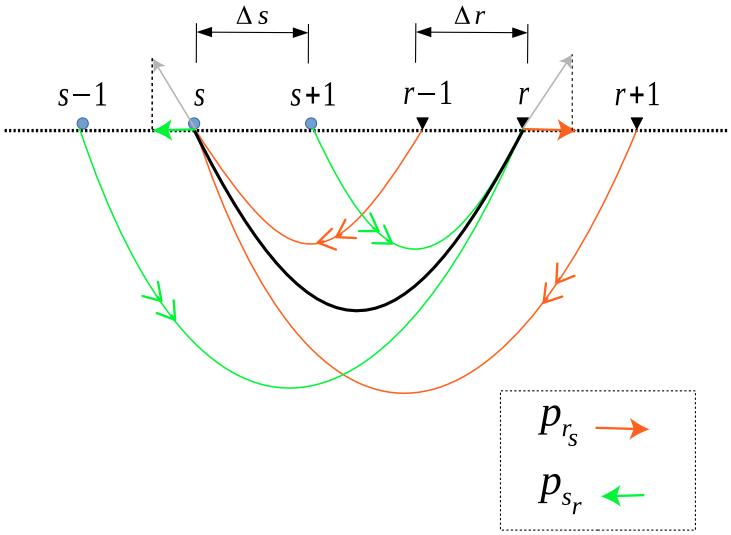

Figure 5.2 – Slope calculations for the first-arrival local coherent events. The horizontal component of the slowness vector at the source and receiver position, (p_{s_r}, p_{r_s}), are calculated using the traveltime maps emitted from the neighbouring sources and receivers (based on the reciprocity rule). The source/receiver intervals are Δr and Δs, respectively. The colors show each slope is calculated through which neighbouring traveltimes. The arrows on the rays designate the propagation directions.

Inverse problem

The l_2-norm misfit function between the observed and calculated data is as follow

$$\begin{aligned} C(\mathbf{m}) &= \frac{1}{2\sigma^2_{T_{s,r}}} \sum_{s=1}^{N_\mathbf{s}} \sum_{r=1}^{N^s_\mathbf{r}} (T_{s,r}(\mathbf{m}) - T^*_{s,r})^2 \\ &+ \frac{1}{2\sigma^2_{p_s}} \sum_{s=1}^{N_\mathbf{s}} \sum_{r=1}^{N^s_\mathbf{r}} (p_{s_r}(\mathbf{m}) - p^*_{s_r})^2 \\ &+ \frac{1}{2\sigma^2_{p_r}} \sum_{r=1}^{N_\mathbf{r}} \sum_{s=1}^{N^r_\mathbf{s}} (p_r(\mathbf{m}) - p^*_{r_s})^2, \end{aligned} \tag{5.13}$$

where $\sigma^2_{T_{s,r}}, \sigma^2_{p_s}$, and $\sigma^2_{p_r}$ are the elements of the covariance matrix, and the symbol $*$ designates the observables. The subsurface parameters can be updated iteratively through the L-BFGS method, similar to the model parameter updating in anisotropic adjoint slope tomography (eq. 3.93). I use the adjoint state method to calculate the gradient of the misfit function with respect to the subsurface parameters. Considering the procedure of gradient calculation in section 3.3.3 as a guide line, I define the Lagrangian as

$$
\begin{aligned}
\mathcal{L}(\mathbf{m},\mathbf{u},\bar{\mathbf{u}}) &= \mathcal{H}(\mathbf{u},\mathbf{m}) \\
&- \sum_{s=1}^{N_\mathbf{s}}\sum_{r=1}^{N_\mathbf{r}^s} \mu_{s,r}\ (T_{s,r} - (Q_{r_s}(\mathbf{t}_s) + Q_{s_r}(\mathbf{t}_r))/2) \\
&- \sum_{s=1}^{N_\mathbf{s}}\sum_{r=1}^{N_\mathbf{r}^s} \xi_{s_r}\ (p_{s_r} - Q_{r_s}(\mathbf{t}_{s+1} - \mathbf{t}_{s-1})/2\Delta s) \\
&- \sum_{r=1}^{N_\mathbf{r}}\sum_{s=1}^{N_\mathbf{s}^r} \xi_{r_s}\ (p_{r_s} - Q_{s_r}(\mathbf{t}_{r+1} - \mathbf{t}_{r-1})/2\Delta r) \\
&- \frac{1}{2}\sum_{s=1}^{N_\mathbf{s}} \Big\langle \lambda_s(\mathbf{x}) \mid H(\mathbf{x},\nabla t_s(\mathbf{x}) \Big\rangle_\Omega - \frac{1}{2}\sum_{r=1}^{N_\mathbf{r}} \Big\langle \lambda_r(\mathbf{x}) \mid H(\mathbf{x},\nabla t_r(\mathbf{x}) \Big\rangle_\Omega .
\end{aligned}
\tag{5.14}
$$

Here $\mathcal{H}(\mathbf{u},\mathbf{m})$ represent the misfit function, and the Lagrange multipliers or adjoint state variables and the associated state variables are gathered in vectors $\bar{\mathbf{u}} = (\mu_{s,r}, \xi_{s_r}, \xi_{r_s}, \boldsymbol{\lambda_s}, \boldsymbol{\lambda_r})$ and $\mathbf{u} = (T_{s,r}, p_{s_r}, p_{r_s}, \mathbf{t}_s, \mathbf{t}_r)$, respectively.

Zeroing the gradient of the Lagrangian with respect to the state variables results in the value of adjoint state variables as follow

$$
\begin{aligned}
\frac{\partial\mathcal{L}}{\partial T_{s,r}} = 0 \quad &\rightarrow \quad \mu_{s,r} = \frac{1}{\sigma^2_{T_{s,r}}}(T_{s,r} - T^*_{s,r}) = \frac{\Delta T_{s,r}}{\sigma^2_{T_{s,r}}}, \\
\frac{\partial\mathcal{L}}{\partial p_{s_r}} = 0 \quad &\rightarrow \quad \xi_{s_r} = \frac{1}{\sigma^2_{p_s}}(p_{s_r} - p^*_{s_r}) = \frac{\Delta p_{s_r}}{\sigma^2_{p_s}}, \\
\frac{\partial\mathcal{L}}{\partial p_{r_s}} = 0 \quad &\rightarrow \quad \xi_{r_s} = \frac{1}{\sigma^2_{p_r}}(p_{r_s} - p^*_{r_s}) = \frac{\Delta p_{r_s}}{\sigma^2_{p_r}},
\end{aligned}
$$

and gradient of the Lagrangian with respect to $\mathbf{t}_s$ is results in

$$
\begin{aligned}
\frac{\partial\mathcal{L}}{\partial\mathbf{t}_s} &= \frac{1}{2}\sum_{r=1}^{N_\mathbf{r}^s} Q^t_{s_r}\ \mu_{s,r} \\
&- \frac{1}{2\Delta s}\sum_{r=1}^{N_\mathbf{r}^{s+1}} Q^t_{r_{s+1}}\ \xi_{s+1_r} \\
&+ \frac{1}{2\Delta s}\sum_{r=1}^{N_\mathbf{r}^{s-1}} Q^t_{r_{s-1}}\ \xi_{s-1_r} \\
&+ (\nabla\cdot(\lambda_s(\mathbf{x})\,\mathcal{U}_s))_\Omega + (\lambda_s(\mathbf{x})\,\mathcal{U}_s.\mathbf{n})_\Gamma .
\end{aligned}
\tag{5.15}
$$

Zeroing this equation leads to the following state equation

$$(\nabla \cdot (\lambda_s(\mathbf{x})\,\mathcal{U}_s))_\Omega = \frac{1}{2}\sum_{r=1}^{N_\mathbf{r}^s} Q_{s_r}^t\ \mu_{s,r} - \frac{1}{2\Delta s}\Big(\sum_{r=1}^{N_\mathbf{r}^{s+1}} Q_{r_{s+1}}^t\ \xi_{s+1_r} + \sum_{r=1}^{N_\mathbf{r}^{s-1}} Q_{r_{s-1}}^t\ \xi_{s-1_r}\Big). \quad (5.16)$$

and in a similar way I derive the state equation which results in the value of λ_r

$$(\nabla \cdot (\lambda_r(\mathbf{x})\,\mathcal{U}_r))_\Omega = \frac{1}{2}\sum_{s=1}^{N_\mathbf{s}^r} Q_{r_s}^t\ \mu_{s,r} - \frac{1}{2\Delta r}\Big(\sum_{s=1}^{N_\mathbf{s}^{r+1}} Q_{s_{r+1}}^t\ \xi_{r+1_s} + \sum_{s=1}^{N_\mathbf{s}^{r-1}} Q_{s_{r-1}}^t\ \xi_{r-1_s}\Big). \quad (5.17)$$

These state equations can be solved through the FSM, the same approach I used to solve the adjoint equations (3.97)-(3.98). Interpretation of these equation is close to the adjoint slope tomography; For adjoint files λ_s and λ_r the residual of the slopes and the traveltimes are back-propagated from the receivers and sources toward the s and r positions, respectively. In our future investigation I shall demonstrate the rule of these residual back-propagations. After calculation of all the adjoint state variables, I can calculate the gradients through the following expressions and update the subsurface parameters in a similar way as the adjoint slope tomography,

$$\nabla_{v_v(\mathbf{x})} C = -\sum_{s=1}^{N_\mathbf{s}} \lambda_s(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_s(\mathbf{x}))}{2\,\partial v_v(\mathbf{x})} - \sum_{r=1}^{N_\mathbf{r}} \lambda_r(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_r(\mathbf{x}))}{2\,\partial v_v(\mathbf{x})}, \quad (5.18)$$

$$\nabla_{\epsilon(\mathbf{x})} C = -\sum_{s=1}^{N_\mathbf{s}} \lambda_s(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_s(\mathbf{x}))}{2\,\partial \epsilon(\mathbf{x})} - \sum_{r=1}^{N_\mathbf{r}} \lambda_r(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_r(\mathbf{x}))}{2\,\partial \epsilon(\mathbf{x})}, \quad (5.19)$$

$$\nabla_{\delta(\mathbf{x})} C = -\sum_{s=1}^{N_\mathbf{s}} \lambda_s(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_s(\mathbf{x}))}{2\,\partial \delta(\mathbf{x})} - \sum_{r=1}^{N_\mathbf{r}} \lambda_r(\mathbf{x}) \frac{2\,\partial H(\mathbf{x}, \nabla t_r(\mathbf{x}))}{2\,\partial \delta(\mathbf{x})}, \quad (5.20)$$

$$\nabla_{\theta(\mathbf{x})} C = -\sum_{s=1}^{N_\mathbf{s}} \lambda_s(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_s(\mathbf{x}))}{2\,\partial \theta(\mathbf{x})} - \sum_{r=1}^{N_\mathbf{r}} \lambda_r(\mathbf{x}) \frac{\partial H(\mathbf{x}, \nabla t_r(\mathbf{x}))}{2\,\partial \theta(\mathbf{x})}. \quad (5.21)$$

Extension to 3D

Chalard et al. (2000, 2002) introduced 3D classic stereotomography where the stereotomographic data set, theoretically, is build based on the traveltimes and the horizontal components of the slowness vector at the source and receiver positions in the inline and crossline directions. However, in many applications the large distance between the shot-lines makes the estimation of the slope at the source position in the crossline direction imprecise (Lambaré, 2008).

In extension of the adjoint slope tomography to 3D data sets, the first concerns is the calculation cost. For 2D data sets I stated that in our method the extra computational cost due to solving the eikonal PDE instead of a system of ODEs in ray tracing can be compensated by more limited number of simulations ($N_s + N_r$ versus N_{sct}). But, I can not contend that this rea-

soning is valid for 3D anisotropic data sets because of the computational cost of 3D version of our eikonal solver. However, this computational cost is totally affordable and it is not a barrier to 3D implementation of our method. Our formulation provides a straightforward scheme for parallel implementation of the algorithm such that I can dedicate each core for calculation of eikonal equation (or the adjoint state equation) of one source/receiver. For the current version of our slope tomography code (in Fortran 90) I use the message passing interface (MPI) for parallelism. This implementation is similar to the adjoint first-arrival traveltime tomography which Noble et al. (2010) investigated its performance for a 3D application.

However, from computational cost point of view, recent developments in resolution of eikonal equation (Han et al., 2017) and quartic equation solvers (Flocke, 2015) seems promising where one can solve the anisotropic eikonal equation while avoiding some iterations over the whole computational domain.

Parsimonious model space

Another interesting extension to the adjoint slope tomography is moving toward a parsimonious model space in which the scatterer point positions are not among the model parameters. For classic formulation this has been introduced by Chauris et al. (2002a) and then extended as non-linear slope tomographic methods Guillaume et al. (2008); Lambaré et al. (2008); Adler et al. (2008). By formulating the adjoint slope tomography based on the same framework as Guillaume et al. (2008) one can remove the scatterer position from the model space of adjoint slope tomography and avoid the leakage between the scatterer positions and the subsurface parameters. This may mitigate the ill-posedness of the inversion.

Implementation of the truncated Newton method

One can use the truncated Newton method as the optimization approach in the adjoint slope tomography. Compared to the L-BFGS method, this method provides a better approximation of inverse Hessian effect and can result in a better estimation of the subsurface parameters. In this method, one solves a Newton linear system through the matrix-free conjugate gradient algorithm to find the model update. This requires to consider an external loop for iterative update of the model parameters and an inner loop for model update calculation (see Métivier et al., 2013, for the application of truncated Newton method in FWI). Thanks to the better approximation of the Hessian, the computational cost of the inner loop can be compensated by a better convergence for the external loop.

Chapter 6

Appendix

6.1 Appendix A: State equation gradient with respect to traveltime

Here we discuss calculate of the gradient of two last terms in Lagrangian (3.94) with respect to state variables $\mathbf{t}_s$ and $\mathbf{t}_r$. We only derive the gradient with respect to $\mathbf{t}_s$ and the other is similar.

Taking derivative of Lagrangian (3.94) with respect to the state variable $\mathbf{t}_s$ involves calculation of

$$\frac{\partial \langle \lambda_s(\mathbf{x}) | H(\mathbf{x}, \nabla t_s(\mathbf{x})) \rangle_\Omega}{\partial t_s(\mathbf{x})}. \tag{6.1}$$

For the sake of clarity and brevity we drop the spatial dependency of variables $(A, C, E, \mathcal{R}, \mathbf{B}, t_s, \lambda_s)$ and we use the continues form of inner product. Considering Hamiltonian (3.86),

$$\begin{aligned}
\frac{\partial \langle \lambda_s(\mathbf{x}) | H(\mathbf{x}, \nabla t_s(\mathbf{x})) \rangle_\Omega}{\partial t_s(\mathbf{x})} &= \int_\Omega \lambda_s \frac{\partial H(\mathbf{x}, \nabla t_s(\mathbf{x}))}{\partial t_s(\mathbf{x})} d\mathbf{x} \\
&= \int_\Omega \lambda_s \Big[2A(\mathcal{R}\nabla t_s)_x \frac{\partial (\mathcal{R}\nabla t_s)_x}{\partial \mathbf{t}_s} + 2C(\mathcal{R}\nabla t_s)_z \frac{\partial (\mathcal{R}\nabla t_s)_z}{\partial t_s} + \\
&\qquad 2E(\mathcal{R}\nabla t_s)_x (\mathcal{R}\nabla t_s)_z \Big((\mathcal{R}\nabla t_s)_x \frac{\partial (\mathcal{R}\nabla t_s)_z}{\partial t_s} + (\mathcal{R}\nabla t_s)_z \frac{\partial (\mathcal{R}\nabla t_s)_x}{\partial t_s} \Big) \Big] d\mathbf{x} \\
&= 2 \int_\Omega \lambda_s \, \mathbf{B} \, \frac{\partial (\mathcal{R}\nabla t_s)}{\partial t_s} d\mathbf{x} \\
&= 2 \int_\Omega \lambda_s (\mathbf{B}\mathcal{R}). \frac{\partial (\nabla t_s)^t}{\partial t_s} d\mathbf{x},
\end{aligned} \tag{6.2}$$

where superscript t denotes the transpose operator $\mathbf{B}_{1\times 2}$ is

$$\begin{aligned}
B_1 &= A(\mathcal{R}\nabla t_s)_x + E(\mathcal{R}\nabla t_s)_x ((\mathcal{R}\nabla t_s)_z)^2, \\
B_2 &= C(\mathcal{R}\nabla t_s)_z + E(\mathcal{R}\nabla t_s)_z ((\mathcal{R}\nabla t_s)_x)^2.
\end{aligned} \tag{6.3}$$

The resulted vector of matricial product $\mathbf{B}\mathcal{R}$ is the group velocity vector map $\mathcal{U}_s$ for the waves

which are propagating from source s. By using the integration by parts [1] we can conclude

$$\int_{\Omega} \lambda_s(\mathbf{B}\mathcal{R}).\frac{\partial(\nabla t_s)^t}{\partial t_s}d\mathbf{x} = \int_{\Omega} \lambda_s \mathcal{U}_s \cdot \frac{\partial(\nabla t_s)^t}{\partial t_s}d\mathbf{x} = \int_{\Gamma} \lambda_s \mathcal{U}_s.\overrightarrow{n}\, d\mathbf{x} + \int_{\Omega} \nabla \cdot (\lambda_s \mathcal{U}_s)d\mathbf{x}. \tag{6.5}$$

Here vector $\overrightarrow{n}$ is the normal to the boundary Γ. Therefore, eq. (6.2) can be summarized as

$$\frac{\partial\langle\lambda_s(\mathbf{x})|H(\mathbf{x},\nabla t_s(\mathbf{x}))\rangle_{\Omega}}{\partial t_s(\mathbf{x})} = 2(\nabla \cdot (\lambda_s(\mathbf{x})\,\mathcal{U}_s))_{\Omega} + 2(\lambda_s(\mathbf{x})\,\mathcal{U}_s.\mathbf{n})_{\Gamma}. \tag{6.6}$$

Therefore the gradient of the Lagrangian (3.94) with respect to the state variable $\mathbf{t}_s$ reads

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\mathbf{t}_s} &= \sum_{r=1}^{N_{\mathbf{r}}^s}\sum_{n_{s,r}=1}^{N_{\mathbf{n}}^{s,r}} Q^t_{n_{s,r}}\mu_{s,r,n_{s,r}} \\ &- \frac{1}{2\Delta s}\sum_{r=1}^{N_{\mathbf{r}}^{s+1}}\sum_{n_{s+1,r}=1}^{N_{\mathbf{n}}^{s+1,r}} Q^t_{n_{s+1,r}}\xi_{s+1,n_{s+1,r}} \\ &+ \frac{1}{2\Delta s}\sum_{r=1}^{N_{\mathbf{r}}^{s-1}}\sum_{n_{s-1,r}=1}^{N_{\mathbf{n}}^{s-1,r}} Q^t_{n_{s-1,r}}\xi_{s-1,n_{s-1,r}} \\ &+ \sum_{s=1}^{N_{\mathbf{s}}}(\nabla \cdot (\lambda_s(\mathbf{x})\,\mathcal{U}_s))_{\Omega} + \sum_{s=1}^{N_{\mathbf{s}}}(\lambda_s(\mathbf{x})\,\mathcal{U}_s.\mathbf{n})_{\Gamma}.\end{aligned} \tag{6.7}$$

In a similar way we can calculate the gradient with respect to $\mathbf{t}_r$.

6.2 Appendix B: Calculation of state equations (3.97) and (3.98)

In order to solve the equations (3.97) and (3.98) we use a similar approach used by Leung and Qian (2006) and Taillandier et al. (2009). Here we solve only the equation (3.97) and the other can be solved in a similar way.

We rewrite the equation (3.97) as

$$\frac{\partial(a\lambda)}{\partial x} + \frac{\partial(b\lambda)}{\partial z} = D, \tag{6.8}$$

1. For vectorial filed $\overrightarrow{\Psi}$ and scalar field φ, the integration by part reads

$$\int_{\Omega} \varphi\nabla\cdot\overrightarrow{\Psi} + \int_{\Omega}\overrightarrow{\Psi}\cdot\nabla\varphi = \int_{\Gamma}\varphi\overrightarrow{\Psi}\cdot\overrightarrow{n}, \tag{6.4}$$

where Ω is the domain which in φ and $\overrightarrow{\Psi}$ are defined, and Γ in the boundary of this domain. $\overrightarrow{n}$ is the normal to this boundary.

where λ is equivalent to the λ_s and the a and b are the group velocity in horizontal and vertical direction, respectively. The right hand side term D stands for the three source terms in the right hand side of eq. (3.97).

Leung and Qian (2006) considered a computational cell at (x_i, z_j), while i and j represent the indexes for conservative discretization of subsurface parameters on a Cartesian grid (Fig. 6.1). The importance of a conservative finite difference scheme lies in the fact that the source terms in the right hand side of eqs (3.97) and (3.98) are defined on the grid points and we need to respect the conservation property of equations (3.97) and (3.98). Therefore, we can discretize the eq. (6.8) as

$$\frac{1}{\Delta x}(a_{i+\frac{1}{2},j}\lambda_{i+\frac{1}{2},j} - a_{i-\frac{1}{2},j}\lambda_{i-\frac{1}{2},j}) + \frac{1}{\Delta z}(b_{i,j+\frac{1}{2}}\lambda_{i,j+\frac{1}{2}} - b_{i,j-\frac{1}{2}}\lambda_{i,j-\frac{1}{2}}) = D. \tag{6.9}$$

In order to find the value of λ on the grid interfaces, $\lambda_{i\pm\frac{1}{2},j}$ and $\lambda_{i,j\pm\frac{1}{2}}$, we need to consider the characteristics propagation. For instance, when $a_{i+\frac{1}{2},j} > 0$ the characteristic which determine the value of λ goes from left hand side of the interface to the right hand side and this suggests the value of $\lambda_{i+\frac{1}{2},j} = \lambda_{i,j}$, otherwise $\lambda_{i+\frac{1}{2},j} = \lambda_{i+1,j}$ (Leung and Qian, 2006). Accordingly, we can rewrite eq. (6.9) as

$$\begin{aligned}&\frac{1}{\Delta x}(a^+_{i+\frac{1}{2},j}\lambda_{i,j} + a^-_{i+\frac{1}{2},j}\lambda_{i+1,j}) - (a^+_{i-\frac{1}{2},j}\lambda_{i-1,j} + a^-_{i-\frac{1}{2},j}\lambda_{i,j})\\ +&\frac{1}{\Delta z}(b^+_{i,j+\frac{1}{2}}\lambda_{i,j} + b^-_{i,j+\frac{1}{2}}\lambda_{i,j+1}) - (b^+_{i,j-\frac{1}{2}}\lambda_{i,j-1} + b^-_{i,j-\frac{1}{2}}\lambda_{i,j}) = D,\end{aligned} \tag{6.10}$$

where

$$\begin{aligned}a^\pm_{i+\frac{1}{2},j} &= \frac{a_{i+\frac{1}{2},j} \pm |a_{i+\frac{1}{2},j}|}{2}, \qquad a^\pm_{i-\frac{1}{2},j} = \frac{a_{i-\frac{1}{2},j} \pm |a_{i-\frac{1}{2},j}|}{2}\\ b^\pm_{i,j+\frac{1}{2}} &= \frac{b_{i,j+\frac{1}{2}} \pm |b_{i,j+\frac{1}{2}}|}{2}, \qquad b^\pm_{i,j-\frac{1}{2}} = \frac{b_{i,j-\frac{1}{2}} \pm |b_{i,j-\frac{1}{2}}|}{2}.\end{aligned} \tag{6.11}$$

So, we conclude

$$\begin{aligned}\Big(\frac{a^+_{i+\frac{1}{2},j} - a^-_{i-\frac{1}{2},j}}{\Delta x} + \frac{b^+_{i,j+\frac{1}{2}} - b^-_{i,j-\frac{1}{2}}}{\Delta z}\Big)\lambda_{i,j} = &\frac{a^+_{i-\frac{1}{2},j}\lambda_{i-1,j} - a^-_{i+\frac{1}{2},j}\lambda_{i+1,j}}{\Delta x}\\ &+ \frac{b^+_{i,j-\frac{1}{2}}\lambda_{i,j-1} - b^-_{i,j+\frac{1}{2}}\lambda_{i,j+1}}{\Delta z} + D.\end{aligned} \tag{6.12}$$

By using the fast sweeping technique, the algorithm ensures coverage of all the characteristics. But before that we need to calculate the value of a and b on the interfaces.

Compared to the adjoint first arrival traveltime tomography and isotropic adjoint slope tomography, where the group velocity is same as phase velocity, here we face more complexity in calculation of a and b on the interfaces. In the isotopic case, $a = \nabla_x \mathbf{t}$ and $b = \nabla_z \mathbf{t}$ which result in a straightforward calculation of a and b on interfaces: for example, $a_{i+\frac{1}{2},j} = (T_{i+1,j} - T_{i,j})/\Delta x$, where $T_{i,j}$ is the value of the traveltime map t on grid point (x_i, z_j) and Δx is the grid interval in x direction.

But for TTI media the horizontal and vertical components of the group velocity depend on

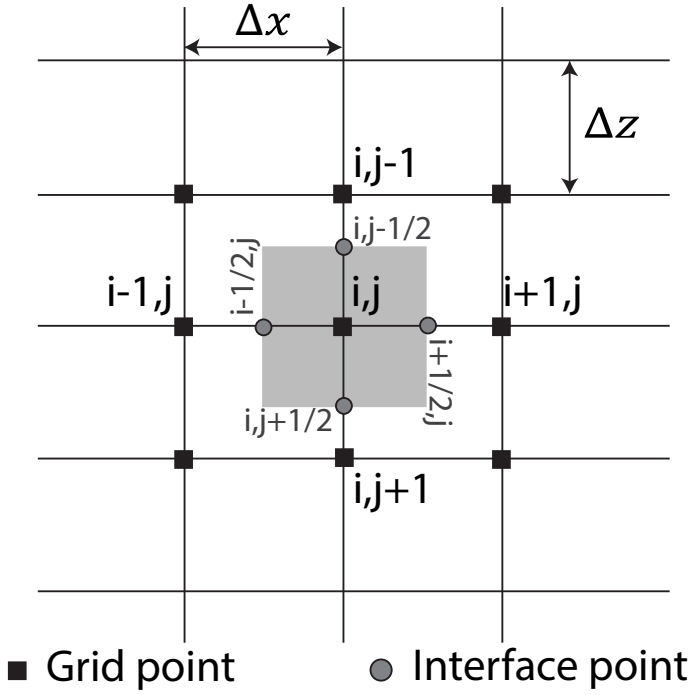

Figure 6.1 – Cartesian grid is discretized by using the conservative finite difference scheme.

both $\nabla_z \mathbf{t}$ and $\nabla_x \mathbf{t}$. This implies calculation of the $\partial_z \mathbf{t}$ on horizontal interfaces and $\nabla_x \mathbf{t}$ on vertical interfaces.

Fig. 6.2 represents the finite difference stencils we use to calculate spatial gradient of travel-time maps on the vertical and horizontal interfaces, I_z and I_x, respectively. These stencils lead us toward the approximations of $\nabla_x \mathbf{t}$ and $\nabla_z \mathbf{t}$ on the horizontal interfaces as

$$\begin{aligned} \nabla_x t &= \frac{t(G_4) - t(G_3)}{\Delta x}, \\ \nabla_z t &= \frac{(t(G_6) + t(G_5)) - (t(G_2) + t(G_1))}{4\Delta z}, \end{aligned} \tag{6.13}$$

and on vertical interfaces as

$$\begin{aligned} \nabla_z t &= \frac{(t(G_6) + t(G_3)) - (t(G_4) + t(G_1))}{4\Delta x}, \\ \nabla_x t &= \frac{t(G_5) - t(G_2)}{\Delta z}. \end{aligned} \tag{6.14}$$

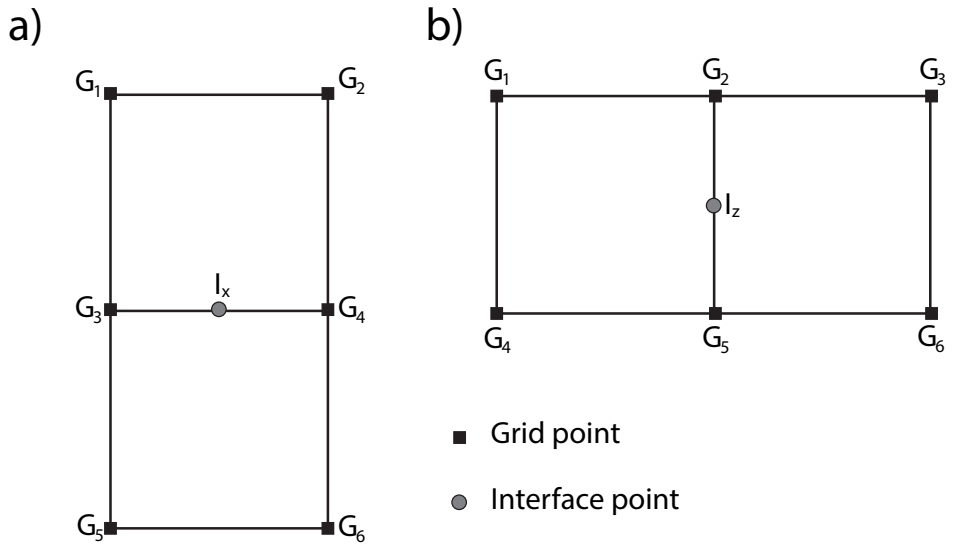

Figure 6.2 – The stencils which are used to extract the gradient of traveltimes on the vertical and horizontal interfaces.

6.3 Appendix C: Effect of source/receiver intervals on slope estimation

Our estimation of slopes (horizontal component of the slowness vector) at the source and receiver positions is based on a finite difference operator which samples the neighbouring traveltime maps of the main source and receiver at the position of scatterers, eqs (3.90)-(3.91). In order to assess the precision of these estimations, we consider the BP TTI model (Fig. 3.32) with a regular distribution of sources and receivers on the surface. We choose one source-receiver pair on the surface at $x = 7$ and $x = 10$ km, respectively. We calculate the slopes at these source and receiver position for a scatterers at $(x, z) = (15, 7)$ km while the source/receiver intervals is 50 m. We consider these slope values as the reference values and by increasing the source intervals and receiver intervals we calculate the deviations of recalculated slopes from their reference values. Fig. 6.3 shows the absolute errors in percentage for the variation of source/receiver intervals from 50 m to 500 m. Because of the subsurface model heterogeneities, the error pattern for slopes at the source and receiver is slightly different but the absolute error for both slopes even for 500 m interval is negligible. The main reason behind these observation relates to the smoothness of the subsurface models and consequently the smoothness of traveltime maps, where there is no caustics. As we discussed for the Marmousi example (section 3.2.5), in the presence of complex structures in the subsurface model and caustics in traveltime maps, our estimation of two-way traveltimes and slopes is erroneous.

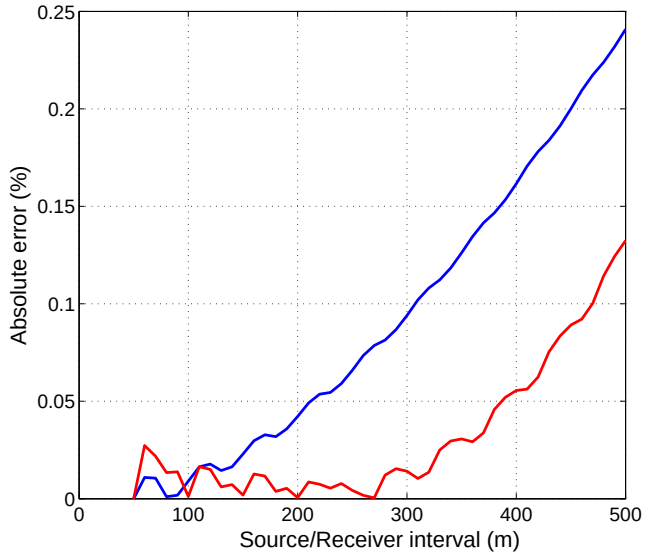

Figure 6.3 – Effect of source/receiver intervals on slope estimations. Diagram show the absolute error in percentage for slope estimation at the source/receiver position respect to the source/receiver intervals for BP TTI model (Fig. 3.32). Blue curve is for a source at $x = 10$ km and the red curve for a receiver at $x = 7$ km. The corresponding scatterer is at $(x, z) = (15, 7)$ km.

6.4 Appendix D: Gradient validations

In this section, we validate the adjoint state method solution for the calculated gradients with respect to the subsurface parameters against the finite difference method. We consider the VTI subsurface models introduced in Fig. 3.22 as the true models. We place a single scatterer at $(x, z) = (10, 2.3)$ km which is sampled by a source-receiver pair with 4 km offset. The source-receiver pair are at the depth of 500 m and the observables are built through the our forward modelling engine. In the following the position of the scatterer is fixed.

In gradient calculation with the finite difference method, we set the model parameter under question $\mathbf{m}$ equal to its true background value $\mathbf{m}_0$ (i.e. the associated true subsurface model without anomaly), while other subsurface parameters are fixed to their true values. This results in an initial value for the misfit function, C_0. By perturbing the initial model $\mathbf{m}_0$ with δm, at each grid point, we calculate the value of misfit function $C_i(\delta m)$. Here "i" designates the grid point. Then, we derive the value of misfit function gradient with respect to the parameter $\mathbf{m}$ at each grid point by

$$\nabla_{m_i} C = \frac{C_i(\delta m) - C_0}{\delta m}. \tag{6.15}$$

Also, we calculate the gradient for this initial model through the adjoint state method. Figs 6.4-6.6 show the direct comparison between the calculated gradient with finite difference method and the adjoint state method for the subsurface parameters v_v, ϵ, δ. Two solutions are close together and only at the shallow part because of the singular points for the source and receiver positions, there are some mismatch between two solutions.

We repeat the previous validation process while the value for the background tilted angle is 10 degree anti-clockwise. This means the model is TTI. Figs 6.7-6.9 show the gradient and the direct comparisons between two methods. The non-zero titled angle alters the shape of the gradients. Here, the sensitivity to the vertical and horizontal propagations should be defined based on the tilted symmetry axis rotation. For example in this TTI model, the highest sensitivity of ϵ gradient is for the propagations which are perpendicular to the symmetry axis. The effect of non zero tilted angle on the gradient of v_v is emerged as a small trail on the bottom of left side adjoint tube (Figs 6.7). The amplitude of ϵ gradient at the bottom of right side adjoint tube, where the propagation is perpendicular to the symmetry axis, is maximum (Fig. 6.8). On the other hand, at this region the δ gradient is minimum (Fig. 6.9) which implies the sensitivity to the vertical propagation should be almost zero. For all the gradients in TTI media the solution of finite difference method and the adjoint state method are close together.

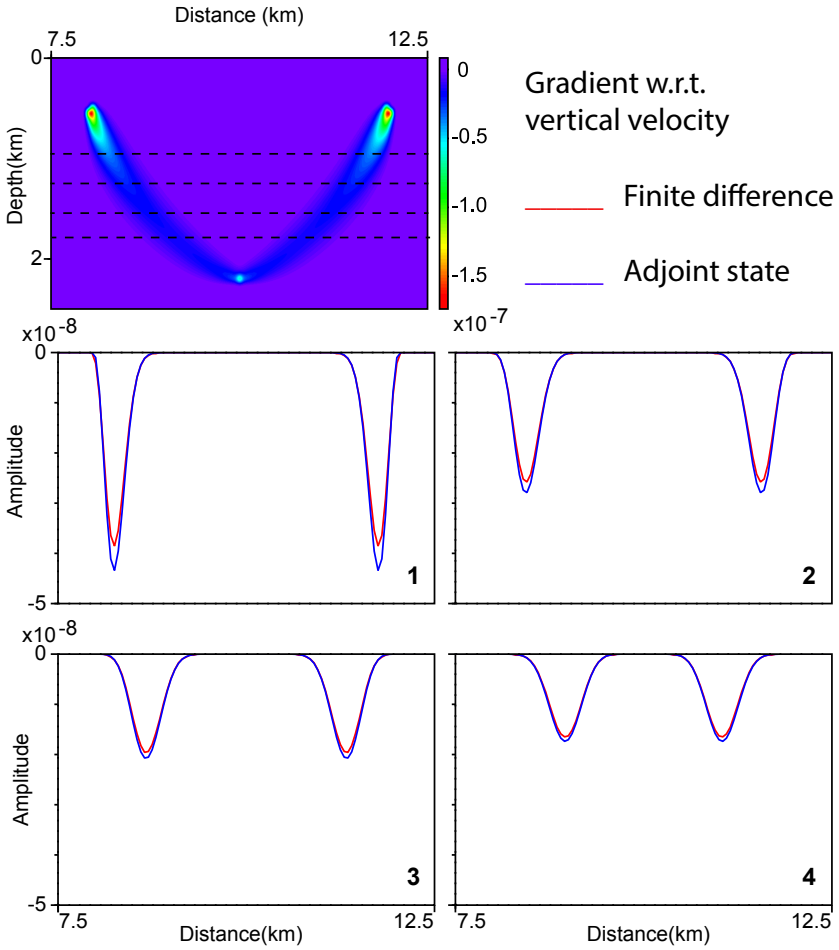

Figure 6.4 – Gradient validation in a VTI medium. Gradient with respect to v_v for one scatterer (from the example in Fig. 3.22). Diagrams show the direct comparison of finite difference method and the adjoint state method solutions across the dashed lines. Diagrams are numbered respectively from top to bottom.

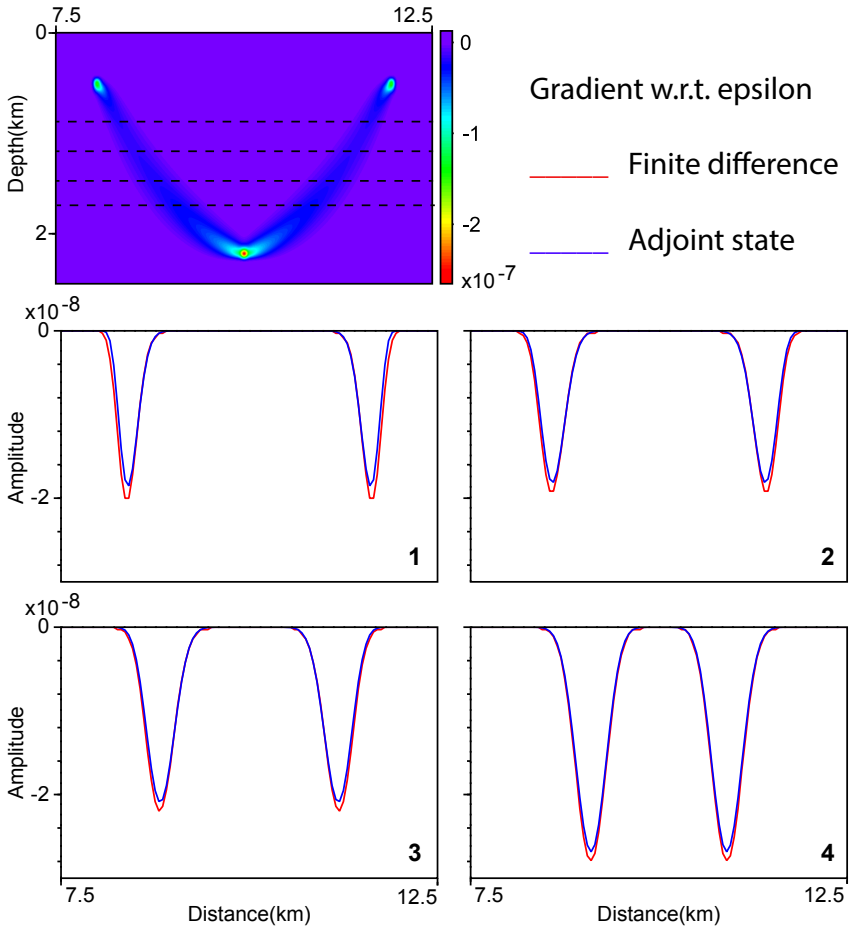

Figure 6.5 – Gradient validation in a VTI medium. Gradient with respect to ϵ for one scatterer (from the example in Fig. 3.22). Diagrams show the direct comparison of finite difference method and the adjoint state method solutions across the dashed lines. Diagrams are numbered respectively from top to bottom.

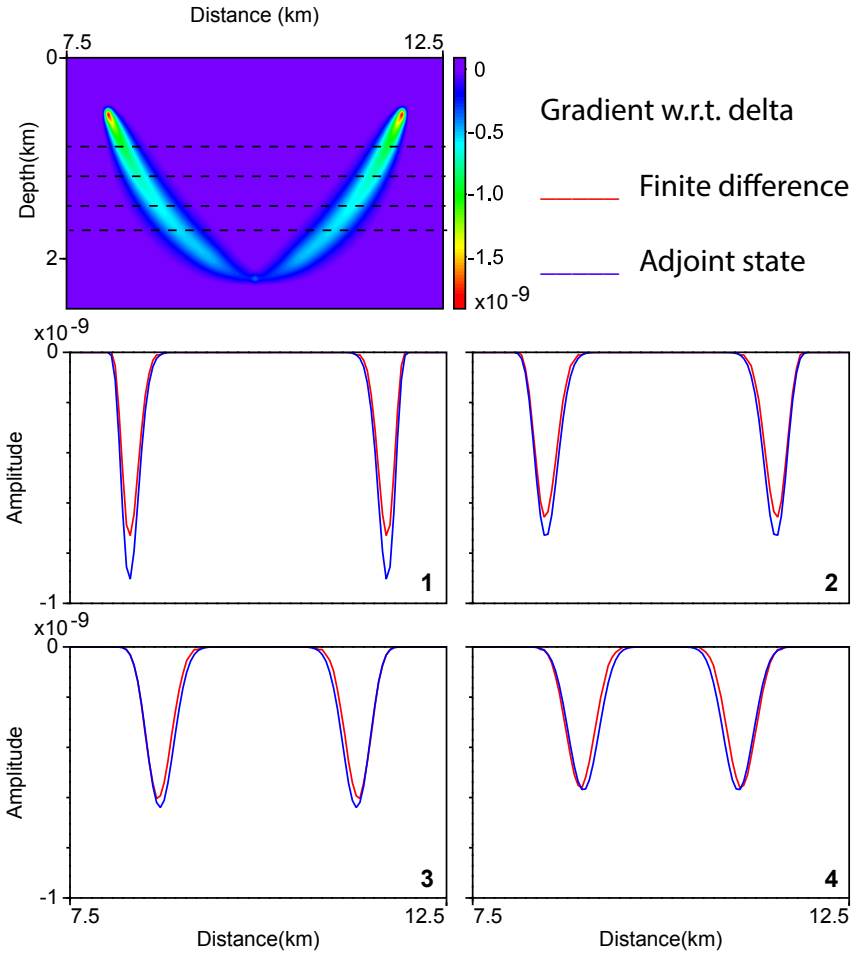

Figure 6.6 – Gradient validation in a VTI medium. Gradient with respect to δ for one scatterer (from the example in Fig. 3.22). Diagrams show the direct comparison of finite difference method and the adjoint state method solutions across the dashed lines. Diagrams are numbered respectively from top to bottom.

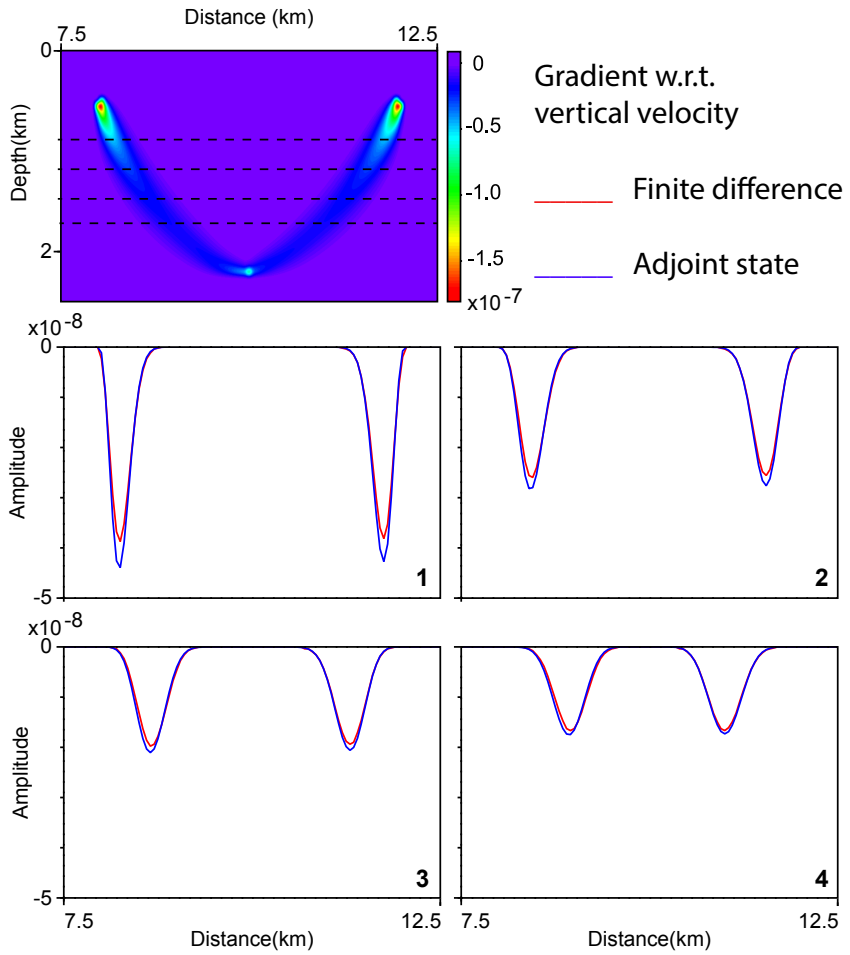

Figure 6.7 – Gradient validation in a TTI medium. Gradient with respect to v_v for one scatterer (from the example in Fig. 3.22 with non-zero titled angle). Diagrams show the direct comparison of finite difference method and the adjoint state method solutions across the dashed lines. Diagrams are numbered respectively from top to bottom.

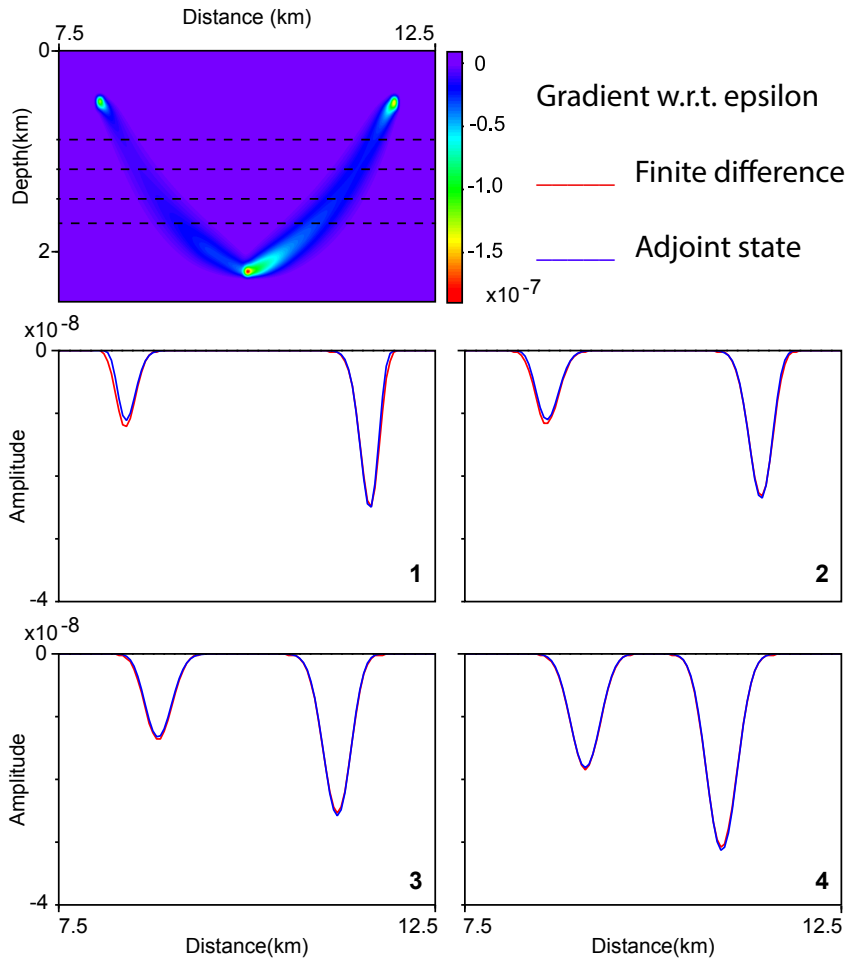

Figure 6.8 – Gradient validation in a TTI medium. Gradient with respect to ϵ for one scatterer (from the example in Fig. 3.22 with non-zero titled angle). Diagrams show the direct comparison of finite difference method and the adjoin state method solutions across the dashed lines. Diagrams are numbered respectively from top to bottom.

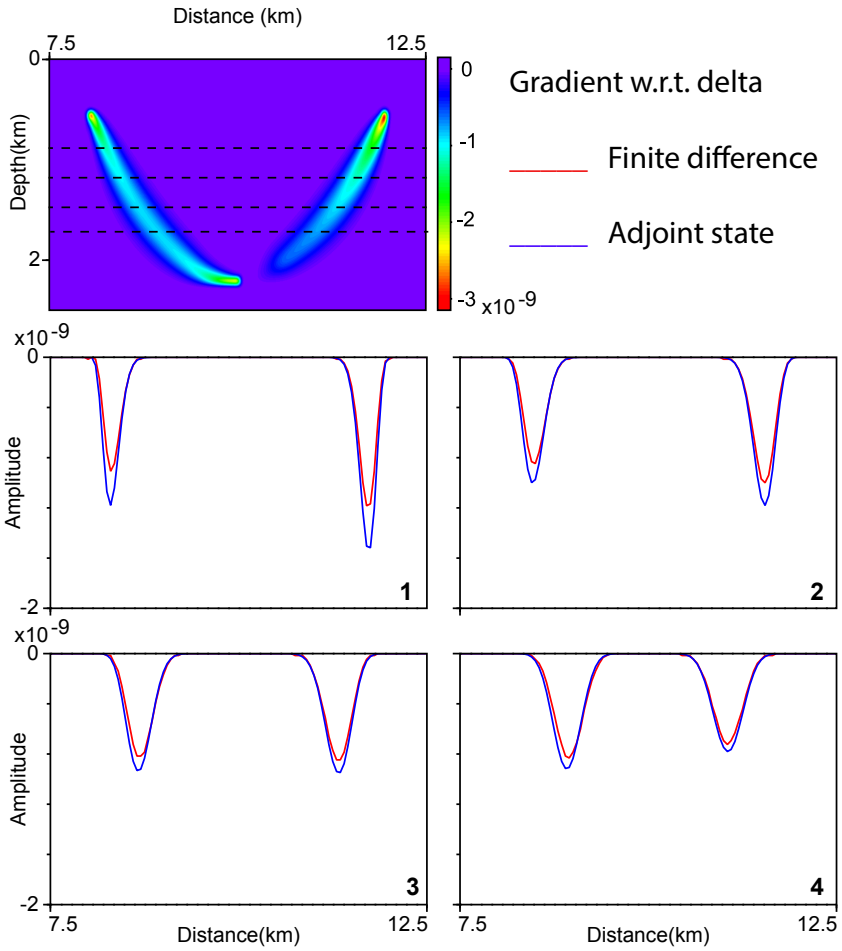

Figure 6.9 – Gradient validation in a TTI medium. Gradient with respect to δ for one scatterer (from the example in Fig. 3.22 with non-zero titled angle). Diagrams show the direct comparison of finite difference method and the adjoin state method solutions across the dashed lines. Diagrams are numbered respectively from top to bottom.

Bibliography

Adler, F., Baina, R., Soudani, M. A., Cardon, P., and Richard, J. B. (2008). Nonlinear 3d tomographic least-squares inversion of residual moveout in kirchhoff prestack-depth-migration common-image gathers. *GEOPHYSICS*, 73(5):VE13–VE23.

Afnimar and Koketsu, K. (2000). Finite difference traveltime calculation for head waves travelling along an irregular interface. *Geophysical Journal International*, 143:729–734.

Akcelik, V., Biros, G., and Ghattas, O. (2002). Parallel multiscale gauss-newton-krylov methods for inverse wave propagation. IEEE Computer Society Press.

Akçelik, V. (2002). *Multiscale Newton-Krylov methods for inverse acoustic wave propagation*. PhD thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania.

Aki, K., Christoffersson, A., and Husebye, E. S. (1977). Determination of the three-dimensional seismic structure of the lithosphere. *Journal of Geophysical Research*, 82(2):277–296.

Aki, K. and Richards, P. (1980). *Quantitative Seismology: Theory and Methods*. W. H. Freeman & Co, San Francisco.

Al-Yahya, K. (1989). Velocity analysis by iterative profile migration. *Geophysics*, 54(6):718–729.

Alerini, M., Lambaré, G., Baina, R., Podvin, P., and Bégat, S. L. (2007). Two-dimensional PP/PS-stereotomography: P- and S-waves velocities estimation from OBC data. *Geophysical Journal International*, 170:725–736.

Alkhalifah, T. (2000). An acoustic wave equation for anisotropic media. *Geophysics*, 65:1239–1250.

Alkhalifah, T. (2003). An acoustic wave equation for orthorhombic anisotropy. *Geophysics*, 68:1169–1172.

Alkhalifah, T. (2011). Scanning anisotropy parameters in complex media. *Geophysics*, 76(2):U13–U22.

Alkhalifah, T. and Larner, K. (1994). Migration error in transversely isotropic media. *Geophysics*, 59(9):1405–1418.

Alkhalifah, T. and Plessix, R. (2014). A recipe for practical full-waveform inversion in anisotropic media: An analytical parameter resolution study. *Geophysics*, 79(3):R91–R101.

Alkhalifah, T. and Tsvankin, I. (1995). Velocity analysis for transversely isotropic media. *Geophysics*, 60:1550–1566.

Arora, J. S., Elwakeil, O. A., Chahande, A. I., and Hsieh, C. C. (1995). Global optimization methods for engineering applications: A review. *Structural and Multidisciplinary Optimization*, 9(3-4):137–159.

Aster, R. C., Borchers, B., and Thurber, C. H. (2004). *Parameter Estimation and Inverse Problems*. Academic Press.

Audebert, F., Pettenati, A., and Dirks, V. (2006). Tti anisotropic depth migration-which tilt estimate should we use? In *68th EAGE Conference and Exhibition incorporating SPE EUROPEC 2006*.

Bakulin, A., Woodward, M., Nichols, D., Osypov, K., and Zdraveva, O. (2010). Localized anisotropic tomography with well information in vti media. *Geophysics*, 75(5):D37–D45.

Barbosa, B., Costa, J., Gomes, E., and Schleicher, J. (2008). Resolution analysis for stereotomography in media with elliptic and anelliptic anisotropy. *Geophysics*, 73(4):R49–R58.

Barbosa, B. S. S., Costa, J. C., Gomes, E. . N. S., and Schleicher, J. (2006). Sensitivity analysis for stereotomography in elliptic and anelliptic media. pages 3413–3417. SEG.

Begat, S. L., Petersen, S. A., and Jahren, L. (2008). Anisotropic reflection and transmission joint tomography-methods and interests. EAGE.

Behera, L. and Tsvankin, I. (2009). Migration velocity analysis for tilted transversely isotropic media. *Geophysical prospecting*, 57:13–26.

Belayouni, N. (2013). *Nouveaux algorithmes efficaces de modélisation 2D et 3D: Temps des premières arrivées, angles à la source et amplitudes*. PhD thesis, Ecole nationale supérieure des mines de Paris.

Benamou, J. D., Luo, S., and Zhao, H.-K. (2010). A compact upwind second order cheme for the eikonal equation. *Journal of Computational Mathematics*, 28:489–516.

Berkhout, A. (2014). An outlook on the future of seismic imaging, Part II: Full-Wavefield Migration. *Geophysical Prospecting*, 62:931–949.

Berkhout, A. J. (1985). *Seismic migration: imaging of acoustic energy by wave field extrapolation; A. theoretical aspects*. Elsevier.

Billette, F. (1998). *Estimation de macro-modèles de vitesse en sismique réflexion par stéréotomographie*. PhD thesis, Université Paris VII.

Billette, F. and Lambaré, G. (1998). Velocity macro-model estimation from seismic reflection data by stereotomography. *Geophysical Journal International*, 135(2):671–680.

Billette, F., Le Bégat, S., Podvin, P., and Lambaré, G. (2003). Practical aspects and applications of 2D stereotomography. *Geophysics*, 68:1008–1021.

Biondi, B. (1992). Velocity estimation by beam stack. *Geophysics*, 57:1034–1047.

Biondi, B. and Almomin, A. (2013). Tomographic full waveform inversion (TFWI) by combining FWI and wave-equation migration velocity analysis. *The Leading Edge*, September, special section: full waveform inversion:1074–1080.

Biondi, B. L. (2006). *3D seismic imaging*. Society of Exploration Geophysicists.

Bishop, T. N., Bube, K. P., Cutler, R. T., Langan, R. T., Love, P. L., Resnick, J. R., Shuey, R. T., and Spinder, D. A. (1985). Tomographic determination of velocity and depth in laterally varying media. *Geophysics*, 50:903–923.

Bourgeois, A., Bourget, M., Lailly, P., Poulet, M., Ricarte, P., and Versteeg, R. (1991). Marmousi, model and data. In *The Marmousi Experience*, pages 5–16. Eur. Ass. Expl. Geophys.

Bozdağ, E., Trampert, J., and Tromp, J. (2011). Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. *Geophysical Journal International*, 185(2):845–870.

Brossier, R., Operto, S., and Virieux, J. (2015). Velocity model building from seismic reflection data by full waveform inversion. *Geophysical Prospecting*, 63:354–367.

Bube, K. P., Langan, R. T., and Nemeth, T. (2005). Analysis of the spectral hole in velocity versus depth resolution for reflection traveltimes with limited aperture. *Geophysics*, 70(3):U37–U45.

Bulant, P. (1996). Two-point ray tracing in 3-D. *Pure and Applied Geophysics*, 148:421–447.

Byrd, R. H., Lu, P., and Nocedal, J. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific and Statistical Computing*, 16:1190–1208.

Carcione, J., Kosloff, D., and Kosloff, R. (1988). Wave-propagation simulation in an elastic anisotropic (transversely isotropic) solid. *Quarterly Journal of Mechanics and Applied Mathematics*, 41(3):319–345.

Castellanos, C., Métivier, L., Operto, S., Brossier, R., and Virieux, J. (2015). Fast full waveform inversion with source encoding and second-order optimization methods. *Geophysical Journal International*, 200(2):720–744.

Castle, R. J. (1994). A theory of normal moveout. *GEOPHYSICS*, 59(6):983–999.

Červený, V. (2001). *Seismic Ray Theory*. Cambridge University Press, Cambridge.

Červený, V., Klimeš, V., and Pšenčik, I. (1988). Complete seismic-ray tracing in three-dimensional structures. In Doornbos, D. J., editor, *Seismological algorithms*, pages 89–168. Academic Press, New York.

Červený, V., Molotkov, I. A., and Pšenčik, I. (1977). *Ray Method in Seismology*. Charles University Press, Praha.

Chalard, E., Podvin, P., Bégat, S. L., Berthet, P., and David, B. (2002). 3-D stereotomographic inversion on a real dataset. In *Expanded Abstracts*, pages 946–948.

Chalard, E., Podvin, P., Lambaré, G., and Audebert, F. (2000). Principles of 3-D stereotomography. In *Extended Abstracts, 62nd annual meeting (29 May-2 June 2000, Glasgow)*, pages P–144. Eur. Ass. Geosc. Eng.

Chapman, C. (2004). *Fundamentals of seismic waves propagation*. Cambridge University Press, Cambridge, England.

Chauris, H. (2000). *Analyse de vitesse par migration pour l'imagerie des structures complexes en sismique réflexion*. PhD thesis, Ecole des Mines de Paris.

Chauris, H. and Noble, M. (2001). 2D velocity macro model estimation from seismic reflection data by local Differential Semblance Optimization: applications on synthetic and real data. *Geophysical Journal International*, 144:14–26.

Chauris, H., Noble, M., Lambaré, G., and Podvin, P. (2002a). Migration velocity analysis from locally coherent events in 2-D laterally heterogeneous media, Part I: Theoretical aspects. *Geophysics*, 67(4):1202–1212.

Chauris, H., Noble, M., Lambaré, G., and Podvin, P. (2002b). Migration velocity analysis from locally coherent events in 2-D laterally heterogeneous media, Part II: Applications on synthetic and real data. *Geophysics*, 67(4):1213–1224.

Chavent, G. (1974). Identification of parameter distributed systems. In Goodson, R. and Polis, M., editors, *Identification of function parameters in partial differential equations*, pages 31–48. American Society of Mechanical Engineers, New York.

Chavent, G. (2009). *Nonlinear least squares for inverse problems*. Springer Dordrecht Heidelberg London New York.

Cheng, X., Jiao, K., Sun, D., and Vigh, D. (2016). Multiparameter estimation with acoustic vertical transverse isotropic full-waveform inversion of surface seismic data. *Interpretation*, 4(4):SU1–SU16.

Claerbout, J. (1985). *Imaging the Earth's interior*. Blackwell Scientific Publication.

Claerbout, J. F. (1976). *Fundamentals of Geophysical Data Processing*. McGraw-Hill Book Co.

Clapp, R., Biondi, B., and Claerbout, J. (2004). Incorporating geologic information into reflection tomography. *Geophysics*, 69:533–546.

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

Costa, J. C., da Silva, F. J., Gomes, E. N., Schleicher, J., Melo, L. A., and Amazonas, D. (2008). Regularization in slope tomography. *Geophysics*, 73(5):VE39–VE47.

Crampin, S. (1981). A review of wave motion in anisotropic and cracked elastic-media. *Wave motion*, 3(4):343–391.

Datta, D. and Sen, M. K. (2016). Estimating starting models for full waveform inversion using a global optimization method. *Geophysics*, 81(4):R211–R223.

de Boor, C. (1978). *A practical guide to splines*. New York, Springer-Verlag.

Dixon, L. and Szego, G. (1978). *Toward global optimization, In: Wirgin, A. (Ed.), Wavefield Inversion*. Amsterdam: North-Holland.

Duveneck, E. and Bakker, P. M. (2011). Stable P-wave modeling for reverse-time migration in tilted TI media. *Geophysics*, 76(2):S65–S75.

Eaton, D. (1993). Finite difference traveltime calculation for anisotropic media. *Geophysical Journal International*, 114:273–280.

Etgen, J., Gray, S. H., and Zhang, Y. (2009). An overview of depth imaging in exploration geophysics. *Geophysics*, 74(6):WCA5–WCA17.

Farra, V. (1989). Ray perturbation theory for heterogeneous hexagonal anisotropic media. *Geophysical Journal International*, 99:723–737.

Farra, V. (1992). Bending method revisited: a hamiltonian approach. *Geophysical Journal International*, 109:138–150.

Farra, V. and Madariaga, R. (1987). Seismic waveform modeling in heterogeneous media by ray perturbation theory. *Journal of Geophysical Research*, 92:2697–2712.

Farra, V. and Madariaga, R. (1988). Non-linear reflection tomography. *Geophysical Journal*, 95:135–147.

Fichtner, A. and Trampert, J. (2011). Resolution analysis in full waveform inversion. *Geophysical Journal International*, 187:1604–1624.

Flocke, N. (2015). Algorithm 954: An accurate and efficient cubic and quartic equation solver for physical applications. *ACM Transactions on Mathematical Software (TOMS)*, 41(4):30.

Floudas, C. A. (2013). *Deterministic Global Optimization: Theory, Methods and Applications*. Springer Science - Business Media.

Fomel, S. (2007). Shaping regularization in geophysical estimation problems. *Geophysics*, 72:R29–R36.

Fomel, S., Luo, S., and Zhao, H.-K. (2009). Fast sweeping method for the factored eikonal equation. *Journal of Computational Physics*, 228:6440–6455.

Forgues, E. and Lambaré, G. (1997). Parameterization study for acoustic and elastic ray+born inversion. *Journal of Seismic Exploration*, 6:253–278.

Gardner, G. H., French, W. S., and Matzuk, T. (1974). Elements of migration and velocity analysis. *Geophysics*, 39:811–825.

Gauthier, O., Virieux, J., and Tarantola, A. (1986). Two-dimensional nonlinear inversion of seismic waveforms: numerical results. *Geophysics*, 51(7):1387–1403.

Ghatak, A. (2005). *Optics*. McGraw-Hill.

Gholami, Y. (2012). *Two-dimensional seismic imaging of anisotropic media by full waveform inversion*. PhD thesis, Université Sophia-Antipolis.

Gholami, Y., Brossier, R., Operto, S., Prieux, V., Ribodetti, A., and Virieux, J. (2013a). Which parametrization is suitable for acoustic VTI full waveform inversion? - Part 2: application to Valhall. *Geophysics*, 78(2):R107–R124.

Gholami, Y., Brossier, R., Operto, S., Ribodetti, A., and Virieux, J. (2013b). Which parametrization is suitable for acoustic VTI full waveform inversion? - Part 1: sensitivity and trade-off analysis. *Geophysics*, 78(2):R81–R105.

Goldberg, D. E. and Kuo, C. H. (1987). Genetic algorithm in pipeline optimization. *J. Computation in Civil Engrg*, 1:128–141.

Gosselet, A. and Bégat, S. L. (2009). Combining borehole and surface seismic data for velocity field estimation through slope tomography. *Geophysical Journal International*, 176(3):897–908.

Gosselet, A., Bégat, S. L., and Petersen, S. A. (2003). 2d walk-away transmission stereotomography. In *SEG Technical Program Expanded Abstracts 2003*, pages 678–681. Society of Exploration Geophysicists.

Gosselet, A., Bégat, S. L., and Petersen, S. A. (2004). Slope transmission tomography for borehole positioning. In *SEG Technical Program Expanded Abstracts 2004*, pages 2331–2334. Society of Exploration Geophysicists.

Gray, S. H. and May, W. P. (1994). Kirchhoff migration using eikonal equation traveltimes. *Geophysics*, 54:810–817.

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

Grechka, V., Zhang, L., and Rector III, J. W. (2004). Shear waves in acoustic anisotropic media. *Geophysics*, 69:576–582.

Grossmann, I. E. (2013). *Global Optimization in Engineering Design*. Springer Science - Business Media.

Guillaume, P., Lambaré, G., Leblanc, O., Mitouard, P., Moigne, J. L., Montel, P., Prescott, T., Siliqi, R., Vidal, N., Zhang, X., and Zimine, S. (2008). Kinematic invariants: an efficient and flexible approach for velocity model building. In *Advanced velocity model building techniques for depth imaging, SEG workshop*.

Hajela, P. (1990). An approach to the nonconvex optimization problem. *AIAA Journal*, 28(7):1205–1210.

Han, S., Zhang, W., and Zhang, J. (2015). A high-order fast sweeping method for calculating quasi-p traveltimes in 2d tti media. In *Expanded Abstracts, 77^{th} Annual EAGE Meeting (Madrid)*.

Han, S., Zhang, W., and Zhang, J. (2017). Calculating qP-wave travel times in 2D TTI media by high-order fast sweeping methods with a numerical quartic equation solver. *Geophysical Journal International*, 210(3):1560–1569.

He, W. and Plessix, R. (2017). Analysis of different parameterisations of waveform inversion of compressional body waves in an elastic transverse isotropic earth with a vertical axis of symmetry. *Geophysical Prospecting*, 65:1004–1024.

Helbig, K. (1994). *Foundations of anisotropy for exploration seismics*. Handbook of geophysical exploration: Seismic exploration. Pergamon.

Hicks, G. J. (2002). Arbitrary source and receiver positioning in finite-difference schemes using Kaiser windowed sinc functions. *Geophysics*, 67:156–166.

Hole, D. and Zelt, B. (1995). 3-D finite difference reflection traveltimes. *Geophysical Journal International*, 121:427–434.

Hole, J. A. (1992). Nonlinear high-resolution three-dimensional seismic travel time tomography. *Journal of Geophysical Research*, 97:6553–6562.

Jannane, M., Beydoun, W., Crase, E., Cao, D., Koren, Z., Landa, E., Mendes, M., Pica, A., Noble, M., Roeth, G., Singh, S., Snieder, R., Tarantola, A., and Trezeguet, D. (1989). Wavelengths of Earth structures that can be resolved from seismic reflection data. *Geophysics*, 54(7):906–910.

Jiang, F. and Zhou, H. (2010). A strategy to estimate anisotropic parameters by error analysis of traveltime inversion. pages 323–327. SEG.

Jones, I. F. (2010). An introduction to velocity model building: Eage. Technical report.

Kao, C. Y., Osher, S., and Qian, J. (2004). Lax-friedrichs sweeping schemes for static hamilton-jacobi equations. *Journal of Computational Physics*, 196:367–391.

Keller, H. B. and Perozzi, P. J. (1983). Fast seismic ray tracing. *SIAM Journal on Applied Mathematics*, 43:981–992.

Kelley, C. (1995). *Iterative Methods for Linear and Nonlinear Equations*. SIAM.

Kim, S. (2001). The most-energetic traveltime of seismic waves. *Applied Mathematics Letters*, 14:313–319.

Kim, S. and Cook, R. (1999). 3d traveltime computation using second order eno scheme. *Geophysics*, 64:1867–1876.

Kiyashchenko, D., Kashtan, B., and Plessix, R.-E. (2004). Anisotropic migration weight for amplitude-preserving migration and sensitivity analysis. *Geophysical Journal International*, 157:753–763.

Knott, C. G. (1899). Reflection and refraction of elastic waves with seismological applications. *Philosophical Magazine*, 48:64–97.

Korenaga, J., Holbrook, W. S., Kent, G. M., Kelemen, P. B., Detrick, R. S., Larsen, H. C., Hopper, J. R., and Dahl-Jensen, T. (2000). Crustal structure of the southeast greenland margin from joint refraction and reflection seismic tomography. *Journal of Geophysical Research*, 105:21591–21614.

Lailly, P. (1984). The seismic inverse problem as a sequence of before stack migrations. In Bednar, R. and Weglein, editors, *Conference on Inverse Scattering, SIAM, Philadelphia*, pages 206–220. Soc. Ind. appl. Math.

Lambaré, G. (2008). Stereotomography. *Geophysics*, 73(5):VE25–VE34.

Lambaré, G., Alerini, M., Baina, R., and Podvin, P. (2004). Stereotomography: a semi-automatic approach for velocity macromodel estimation. *Geophysical Prospecting*, 52:671–681.

Lambaré, G., Deladerriére, N., Traonmilin, Y., Touré, J. P., Moigne, J. L., and Herrmann, P. (2009). Non-linear tomography for time imaging. EAGE.

Lambaré, G., Hermann, P., Touré, J., Capar, L., Guillaume, P., Bousquié, N., Grenié, D., and Zimine, S. (2007). From time to depth imaging: an accurate workflow. pages 3014–3017. SEG.

Lambaré, G., Lucio, P. S., and Hanyga, A. (1996). Two-dimensional multivalued traveltime and amplitude maps by uniform sampling of ray field. *Geophysical Journal International*, 125:584–598.

Lambaré, G., Touré, J. P., Moigne, J. L., Zimine, S., Herrmann, P., Carbonara, S., and Federici, F. (2008). Velocity update for prestack time migration. SEG.

Lavaud, B., Baina, R., and Landa, E. (2004). Post-stack stereotomography: A robust strategy. page C22. EAGE.

Le Bouteiller, P., Benjemaa, M., Chauris, H., Métivier, L., Tavakoli, B., Noble, M., and Virieux, J. (2017). Discontinuous Galerkin method for TTI Eikonal equation. In *Proceedings of the* 79^{th} *EAGE Conference & Exhibition*. EAGE.

Lee, W., Jennings, P., Kisslinger, C., and (eds), H. K. (2002). *International Handbook of Earthquake and Engineering Seismology*. Academic Press.

Lelièvre, P. G., Farquharson, C. G., and Hurich, C. A. (2011). Computing first-arrival seismic traveltimes on unstructured 3-D tetrahedral grids using the fast marching method. *Geophysical Journal International*, 184:885–896.

Leung, S. and Qian, J. (2006). An adjoint state method for three-dimensional transmission traveltime tomography using first-arrivals. *Communications in Mathematical Sciences*, 4(1):249–266.

Li, Y. and Biondi, B. (2011). Migration velocity analysis for anisotropic model. pages 201–206. SEG.

Li, Y., Shen, P., and Perkins, C. (2012). Vti migration velocity analysis using rtm. pages 1–6. SEG.

Lines, L. R. and Treitel, S. (1984). Tutorial: a review of least-squares inversion and its application to geophysical problems. *Geophysical Prospecting*, 32:159–186.

Lucio, P. S., Lambaré, G., and Hanyga, A. (1996). 3D multivalued travel time and amplitude maps. *Pure and Applied Geophysics*, 148:113–136.

Luo, J. and Wu, R.-S. (2015). Seismic envelope inversion: reduction of local minima and noise resistance. *Geophysical Prospecting*, 63(3):597–614.

Luo, S. and Qian, J. (2012). Fast sweeping method for factored anisotropic eikonal equations: multipicative and additive factors. *Journal of Scientific Computing*, 52:360–382.

Luo, S. and Sava, P. (2011). A deconvolution-based objective function for wave-equation inversion. *SEG Technical Program Expanded Abstracts*, 30(1):2788–2792.

Ma, X. and Alkhalifah, T. (2013). qp-wave traveltime computation by an iterative approach. In *Expanded Abstracts*. EAGE.

Menke, W. (2012). *Geophysical data analysis: Discrete inverse theory*, volume 45. Academic press.

Messud, J., Lambaré, G., Guillaume, P., and Rohel, C. (2015). Non-linear slope tomography for orthorhombic pre-stack time migration. EAGE.

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

Métivier, L. and Brossier, R. (2016). The SEISCOPE optimization toolbox: A large-scale nonlinear optimization library based on reverse communication. *Geophysics*, 81(2):F11–F25.

Métivier, L., Brossier, R., Labbé, S., Operto, S., and Virieux, J. (2014b). A robust absorbing layer for anisotropic seismic wave modeling. *Journal of Computational Physics*, 279:218–240.

Métivier, L., Brossier, R., Labbé, S., Operto, S., and Virieux, J. (2014c). Smart: dissipative absorbing layer technique for general elastodynamics equations. application as s-waves filter in acoustic ti media. *Seismic Technology*, 11(4):14.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J. (2016a). Increasing the robustness and applicability of full waveform inversion: an optimal transport distance strategy. *The Leading Edge*, 35(12):1060–1067.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J. (2016b). Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion. *Geophysical Journal International*, 205:345–377.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J. (2016c). An optimal transport approach for seismic tomography: Application to 3D full waveform inversion. *Inverse Problems*, 32(11):115008.

Métivier, L., Brossier, R., Virieux, J., and Operto, S. (2013). Full Waveform Inversion and the truncated Newton method. *SIAM Journal On Scientific Computing*, 35(2):B401–B437.

Montel, J., Guillaume, P., Lambaré, G., and Leblanc, O. (2010). Non-linear slope tomography-extension to maz and waz. In *72nd EAGE Conference and Exhibition incorporating SPE EUROPEC 2010*.

Mora, P. R. (1987). Nonlinear two-dimensional elastic inversion of multi-offset seismic data. *Geophysics*, 52:1211–1228.

Mora, P. R. (1988). Elastic wavefield inversion of reflection and transmission data. *Geophysics*, 53:750–759.

Mora, P. R. (1989). Inversion = migration + tomography. *Geophysics*, 54(12):1575–1586.

Nag, S., Alerini, M., Duveneck, E., and Ursin, B. (2006). 2-D stereotomography for anisotropic media. In *Expanded Abstracts*, pages 3305–3308.

Nag, S., Alerini, M., and Ursin, B. (2010). PP/PS anisotropic stereotomography. *Geophysical Journal International*, 181(1):427–452.

Neckludov, D., Baina, R., and Landa, E. (2006). Residual stereotomographic inversion. *Geophysics*, 71:E35–E39.

Nguyen, S., Baina, R., Alerini, M., and Lambaré, G. (2008). Stereotomography assisted by migration of attributes. *Geophysical Prospecting*, 56:613–625.

Noble, M., Gesret, A., and Belayouni, N. (2012). A new 3d eikonal solver for accurate traveltimes, takeoff angles and amplitudes. In *Extended abstract, 74th EAGE Conference & Exhibition incorporating SPE EUROPEC (Copenhagen)*.

Noble, M., Gesret, A., and Belayouni, N. (2014). Accurate 3-d finite difference computation of travel time in strongly hetrogeneous media. *Geophysical Journal International*, 199:1572–1585.

Noble, M., Thierry, P., Taillandier, C., and Calandra, H. (2010). High-performance 3D first-arrival traveltime tomography. *The Leading Edge*, 29(1):86–93.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, 2nd edition.

Nolet, G. (1987). *Seismic tomography with applications in global seismology and exploration geophysics*. D. Reidel publishing Company.

Operto, S., Brossier, R., Gholami, Y., Métivier, L., Prieux, V., Ribodetti, A., and Virieux, J. (2013). A guided tour of multiparameter full waveform inversion for multicomponent data: from theory to practice. *The Leading Edge*, Special section Full Waveform Inversion(September):1040–1054.

Operto, S., Lambaré, G., Podvin, P., and Thierry, P. (2003). 3-D ray-Born migration/inversion. part 2: application to the SEG/EAGE overthrust experiment. *Geophysics*, 68(4):1357–1370.

Operto, S., Virieux, J., Ribodetti, A., and Anderson, J. E. (2009). Finite-difference frequency-domain modeling of visco-acoustic wave propagation in two-dimensional TTI media. *Geophysics*, 74 (5):T75–T95.

Operto, S., Xu, S., and Lambaré, G. (2000). Can we image quantitatively complex models with rays? *Geophysics*, 65(4):1223–1238.

Paige, C. C. and Saunders, M. A. (1982). LSQR: an algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical software*, 8(1):43–71.

Pica, A. (1997). Fast and accurate finite difference solution of the 3D eikonal equation parameterized in celerity. In *Expanded Abstracts*, pages 1774–1777. Society of Exploration Geophysics.

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

Plessix, R. E. and Cao, Q. (2011). A parametrization study for surface seismic full waveform inversion in an acoustic vertical transversely isotropic medium. *Geophysical Journal International*, 185:539–556.

Podvin, P. and Lecomte, I. (1991). Finite difference computation of traveltimes in very contrasted velocity model : a massively parallel approach and its associated tools. *Geophysical Journal International*, 105:271–284.

Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, part I : theory and verification in a physical scale model. *Geophysics*, 64:888–901.

Pratt, R. G., Plessix, R. E., and Mulder, W. A. (2001). Seismic waveform tomography: the effect of layering and anisotropy. In *Presented at the 63th EAGE Conference & Exhibition, Amsterdam, EAGE.*, page P092.

Pratt, R. G., Sirgue, L., Hornby, B., and Wolfe, J. (2008). Cross-well waveform tomography in fine-layered sediments - meeting the challenges of anisotropy. In *70^{th} Annual EAGE Conference & Exhibition, Roma*, page F020.

Prieux, V., Brossier, R., Gholami, Y., Operto, S., Virieux, J., Barkved, O., and Kommedal, J. (2011). On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study. *Geophysical Journal International*, 187:1495–1515.

Prieux, V., Lambaré, G., Operto, S., and Virieux, J. (2013). Building starting model for full waveform inversion from wide-aperture data by stereotomography. *Geophysical Prospecting*, 61(Issue supplement: 60 year anniversary issue):109–137.

Prieux, V., Operto, S., Lambaré, G., and Virieux, J. (2010). Building starting model for full waveform inversion from wide-aperture data by stereotomography. *SEG Technical Program Expanded Abstracts*, 29(1):988–992.

Qian, J., Cheng, L.-T., and Osher, S. (2003). A level set based eulerian approach for anisotropic wave propagations. *Wave Motion*, 37(4):365–379.

Qian, J. and Leung, S. (2004). A level set based Eulerian method for paraxial multivalued traveltimes. *Journal of Computational Physics*, 197:711–736.

Qian, J. and Symes, W. W. (2001). Paraxial eikonal solvers for anisotropic quasi-p travel times. *Journal of Computational Physics*, 173:256–278.

Qian, J., Zhang, Y.-T., and Zhao, H.-K. (2007). A fast sweeping method for static convex Hamilton–Jacobi equations. *Journal of Scientific Computing*, 31:237–271.

Rawlinson, N., Pozgay, S., and Fishwick, S. (2010). Seismic tomography: a window into deep earth. *Physics of the Earth and Planetary Interiors*, 178(3):101–135.

Riabinkin, L. A. (1957). *Fundamentals of resolving power of Controlled Directional Reception (CDR) of seismic waves*, volume 14, pages 36–60. Society of Exploration Geophysics. Translated and paraphrased from Prikladnaya, 16, 3-36.

Riabinkin, L. A., Napalkov, I. V., Znamenskii, V. V., Voskresenskii, I. N., and Rapoport, M. B. (1962). Theory and practice of the CDR seismic method. *Transaction of the Gubkin Institute of Petrochemical and Gas Production (Moscow)*, 39.

Rieber, F. (1936). A new reflection system with controlled directional sensitivity. *Geophysics*, 1:97–106.

Rusmanugroho, H., Modrak, R., and Tromp, J. (2017). Anisotropic full-waveform inversion with tilt-angle recovery. *Geophysics*, 82(3):R135–R151.

Saenger, E. H., Gold, N., and Shapiro, S. A. (2000). Modeling the propagation of elastic waves using a modified finite-difference grid. *Wave motion*, 31:77–92.

Sajeva, A., M.Aleardi, Stucchi, E., Bienati, N., and Mazzotti, A. (2016). Estimation of acoustic macro models using a genetic full-waveform inversion: Applications to the marmousi model. *Geophysics*, 81(4):R173–R184.

Santos, L. T., Schleicher, J. ., Tygel, M., and Hubral, P. (2000). Modeling, migration, and demigration. *The Leading edge*, 19(7):712–715.

Sarkar, D. and Tsvankin, I. (2004). Migration velocity analysis in factorized vti media. *Geophysics*, 69:708–718.

Sava, P. and Biondi, B. (2004). Wave-equation migration velocity analysis. i. theory. *Geophysical Prospecting*, 52(6):593–606.

Sen, M. K. and Stoffa, P. L. (2013). *Global Optimization Methods in Geophysical Inversion (second edition)*. Elsevier Science Publishing Co.

Sexton, P. and Williamson, P. (1998). 3d anisotropic velocity estimation by model-based inversion of pre-stack traveltimes. pages 1855–1858. SEG.

Shah, H. (2007). The 2007 bp anisotropic velocity-analysis benchmark. In *Expanded Abstracts*. EAGE workshop.

Sheriff, R. E. and Geldart, L. P. (1995). *Exploration seismology, second edition*. Cambridge University Press.

Stork, C. (1992). Reflection tomography in the post migrated domain. *Geophysics*, 57:680–692.

Sword, C. H. (1986). Tomographic determination of interval velocities from picked reflection seismic data. In *Expanded Abstracts*, pages 657–660. Society of Exploration Geophysics.

Sword, C. H. (1987). *Tomographic determination of interval velocities from reflection seismic data: The method of controlled directional reception*. PhD thesis, Stanford University.

Symes, W. W. (1998). High frequency asymptotics, differential semblance and velocity estimation. In *Expanded Abstracts*, pages 1616–1619. Society of Exploration Geophysics.

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

Symes, W. W. and Carazzone, J. J. (1991). Velocity inversion by differential semblance optimization. *Geophysics*, 56:654–663.

Taillandier, C., Noble, M., Chauris, H., and Calandra, H. (2009). First-arrival travel time tomography based on the adjoint state method. *Geophysics*, 74(6):WCB1–WCB10.

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

Tarantola, A. (1984b). Linearized inversion of seismic reflection data. *Geophysical Prospecting*, 32:998–1015.

Tarantola, A. (1987). *Inverse problem theory: methods for data fitting and model parameter estimation*. Elsevier, New York.

Tavakoli F., B., Operto, S., Ribodetti, A., and Virieux, J. (2017a). Anisotropic slope tomography based on eikonal solver and adjoint-state method. In *79th EAGE Conference and Exhibition 2017*.

Tavakoli F., B., Operto, S., Ribodetti, A., and Virieux, J. (2017b). Slope tomography based on eikonal solvers and the adjoint-state method. *Geophysical Journal International*, 209(3):1629–1647.

Tavakoli F., B., Ribodetti, A., Operto, S., and Virieux, J. (2016). Adjoint stereotomography. In *SEG Technical Program Expanded Abstracts 2016*, pages 5269–5273.

Tavakoli F., B., Ribodetti, A., Virieux, J., and Operto, S. (2015). An iterative factored eikonal solver for TTI media. In *SEG technical program expanded abstracts 2015*, volume 687, pages 3576–3581.

Telford, W. M., Geldart, L. P., and Sheriff, R. E. (1990). *Applied Geophysics - Second Edition*. Cambridge University Press.

Thierry, P., Operto, S., and Lambaré, G. (1999). Fast 2D ray-Born inversion/migration in complex media. *Geophysics*, 64(1):162–181.

Thomsen, L. A. (1986). Weak elastic anisotropy. *Geophysics*, 51:1954–1966.

Tikhonov, A. N. (1963). Resolution of ill-posed problems and the regularization method (in russian, french translation, mir, moscow, 1976). *Dokl. Akad. Nauk SSSR*, 151:501–504.

Tsvankin, I. (1995). Normal moveout from dipping reflectors in anisotropic media. *Geophysics*, 60:268–284.

Tsvankin, I. (1997). Anisotropic parameters and p-wave velocity for orthorhombic media. *Geophysics*, 62:1292–1309.

Tsvankin, I. (2001a). Seismic signature and analysis of reflection data in anisotropic media. In Helbig, K. and Treitel, S., editors, *Seismic Exploration*, volume 29 of *Handbook of Geophysical Exploration*. Pergamon.

Tsvankin, I. (2001b). Seismic signature and analysis of reflection data in anisotropic media. In Helbig, K. and Treitel, S., editors, *Seismic Exploration*, volume 29 of *Handbook of Geophysical Exploration*. Elsevier, 2 edition.

Tsvankin, I. and Thomsen, L. (1994). Nonhyperbolic reflection moveout in anisotropic media. *Geophysics*, 59:1290–1304.

Tsvankin, I. and Thomsen, L. (1995). Inversion of reflction traveltime for transverse isotropy. *Geophysics*, 60(4):1095–1107.

Um, J. and Thurber, C. (1987). A fast algorithm for two-point seismic ray tracing. *Bulletin of the Seismological Society of America*, 77:972–986.

van Leeuwen, T. and Herrmann, F. (2015). A penalty method for pde-constrained optimization in inverse problems. *Inverse Problems*, 32(1):015007.

van Leeuwen, T. and Herrmann, F. J. (2013). Mitigating local minima in full-waveform inversion by expanding the search space. *Geophysical Journal International*, doi:10.1093/gji/ggt258.

van Leeuwen, T. and Mulder, W. (2008). Velocity analysis based on data correlation. *Geophysical Prospecting*, 56(6):791–803.

Vidale, D. (1988a). Finite-difference calculation of travel time. *Bulletin of the Seismological Society of America*, 78:2062–2076.

Vidale, D. (1988b). Finite-difference calculation of travel time. *Bulletin of the Seismological Society of America*, 78:2062–2076.

Vidale, J. (1990). Finite-difference calculation of travel times in three dimensions. *Geophysics*, 55(5):521–526.

Vinje, V., Iversen, E., Astebøl, K., and Gjøystdal, H. (1996). Estimation of multivalued arrivals in 3D models using wavefront construction-part I. *Geophysical Prospecting*, pages 819–842.

Vinje, V., Iversen, E., and Gjøystdal, H. (1992). Traveltime and amplitude estimation using wavefront construction. In *Extended Abstracts*, pages 504–505. Eur. Ass. Expl. Geophys.

Vinje, V., Iversen, E., and Gjøystdal, H. (1993). Traveltime and amplitude estimation using wavefront construction. *Geophysics*, pages 1157–1166.

Virieux, J. (1989). Perturbed ray tracing on an heterogeneous sphere. *Geophysical Research Letters*, 16:405–408.

Virieux, J. (1991). Fast and accurate ray tracing by hamiltonian perturbation. *Journal of Geophysical Research*, 96(B1):579–594.

Virieux, J. and Farra, V. (1991). Ray tracing in 3D complex isotropic media: an analysis of the problem. *Geophysics*, 16:2057–2069.

Virieux, J., Farra, V., and Madariaga, R. (1988). Ray tracing for earthquake location in laterally heterogeneous media. *Journal Geophysical Research*, 93(B6):1116–1120.

Virieux, J. and Lambaré, G. (2015). Theory and observations: Body waves, ray methods, and finite-frequency effects. *Treatise on Geophysics (Second Edition)*, 1:169–200.

Virieux, J. and Operto, S. (2009). An overview of full waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26.

Waheed, U. B., Flagg, G., and Yarman, C. E. (2016). First-arrival traveltime tomography for anisotropic media using the adjoint-state method. *Geophysics*, 81(4):R147–R155.

Waheed, U. B., Yarman, C. E., and Flagg, G. (2014). An iterative fast sweeping based eikonal solver for tilted orthorhombic media. In *Expanded Abstracts*, pages 480–485. Society of Exploration Geophysics.

Waheed, U. B., Yarman, C. E., and Flagg, G. (2015). An iterative, fast-sweeping-based eikonal solver for 3D tilted anisotropic media. *Geophysics*, 80:C49–C58.

Wang, F., Donno, D., Chauris, H., Calandra, H., and Audebert, F. (2016). Waveform inversion based on wavefield decomposition. *Geophysics*, 81(6):R457–R470.

Wang, X. and Tsvankin, I. (2013). Ray-based gridded tomography for tilted transversely isotropic media. *Geophysics*, 78(1):C11–C23.

Warner, M. and Guasch, L. (2016). Adaptive waveform inversion: Theory. *GEOPHYSICS*, 81(6):R429–R445.

Warner, M., Ratcliffe, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N., Vinje, V., Stekl, I., Guasch, L., Win, C., Conroy, G., and Bertrand, A. (2013). Anisotropic 3D full-waveform inversion. *Geophysics*, 78(2):R59–R80.

Weibull, W. W. and Arntsen, B. (2014). Anisotropic migration velocity analysis using reverse-time migration. *Geophysics*, 79(1):R13–R25.

Wellington, P. (2016). *Efficient 1D, 2D and 3D Geostatistical constraints and their application to Full Waveform Inversion*. PhD thesis, Univ. Grenoble Alpes.

Wiechert, E. and Zoeppritz, K. (1907). Über erdbebenwellen. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, pages 415–549.

Williamson, P. (1991). A guide to the limits of resolution imposed by scattering in ray tomography. *Geophysics*, 56:202–207.

Woodward, M. J., Nichols, D., Zdraveva, O., Whitfield, P., and Johns, T. (2008). A decade of tomography. *Geophysics*, 73(5):VE5–VE11.

Wu, Z. and Alkhalifah, T. (2015). Simultaneous inversion of the background velocity and the perturbation in full-waveform inversion. *Geophysics*, 80(6).

Xu, S., Wang, D., Chen, F., Lambaré, G., and Zhang, Y. (2012). Inversion on reflected seismic wave. *SEG Technical Program Expanded Abstracts 2012*, pages 1–7.

Yilmaz, O. (2001). *Seismic data analysis*. Society of Exploration Geophysicists: processing, inversion and interpretation of seismic data.

Yuan, Y. O., Simons, F. J., and Tromp, J. (2016). Double-difference adjoint seismic tomography. *Geophysical Journal International*, 206(3):1599–1618.

Zhang, L., Rector III, J. W., and Hoversten, G. M. (2005). Eikonal solver in the celerity domain. *Geophysical Journal International*, 162:1–8.

Zhang, Y.-T., Zhao, H.-K., and Qian, J. (2006). High order fast sweeping methods for static Hamilton–Jacobi equations. *Journal of Scientific Computing*, 29:25–56.

Zhao, H. (2005). A fast sweeping method for eikonal equations. *Mathematics of computation*, 74:603–627.

Zhou, B. and Greenhalgh, S. (2008). Non-linear traveltime inversion for 3-d seismic tomography in strongly anisotropic media. *Geophysics*, 172:383–394.

Zhou, C., Jiao, J., Lin, S., Sherwood, J., and Brandsberg-Dahl, S. (2011). Multiparameter joint tomography for tti model building. *Geophysics*, 76(5):WB183–WB190.

Zhou, H., Pham, D., and Gray, S. (2004). Tomographic velocity analysis in strongly anisotropic media. pages 2347–2350. SEG.

Zhou, W. (2016). *Velocity Model Building by Full Waveform Inversion of Early Arrivals and Reflections and Case Study with Gas Cloud Effect*. PhD thesis, PhD thesis: Université de Grenoble.

Zhou, W., Brossier, R., Operto, S., and Virieux, J. (2015). Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation. *Geophysical Journal International*, 202(3):1535–1554.

Zhu, H. and Fomel, S. (2016). Building good starting models for full-waveform inversion using adaptive matching filtering misfit. *Geophysics*, 81(5):U61–U72.

Zoeppritz, K. and Geiger, L. (1907). Über erdbebenwellen ii. *Laufzeitkurven. Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse*, pages 529–549.