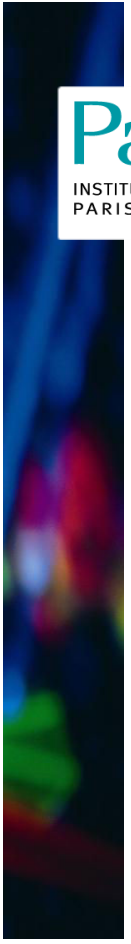
Waveform inversion based on wavefield decomposition

Fang Wang



ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

MINES ParisTech

École doctorale n° 398

Géosciences, Ressources Naturelles et Environnement

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

l'École nationale supérieure des mines de Paris

Spécialité "Dynamique et Ressources des Bassins Sédimentaires"

présentée et soutenue publiquement par

Fang WANG

le 16 Octobre 2015

Waveform inversion based on wavefield decomposition

L'inversion des formes d'ondes par décomposition des champs d'ondes

Directeur de thèse : **Hervé CHAURIS**
Co-encadrement de la thèse : **Daniela DONNO**

Jury

M. Stéphane OPERTO, Directeur de recherche, CNRS, Géoazur	Président
M. Paul SAVA, Professeur, Colorado School of Mines, Etats-Unis	Rapporteur
M. Gilles LAMBARE, Directeur de recherche, CGG, professeur associé IPGP	Rapporteur
M. Romain BROSSIER, Maître de conférences, Université Joseph Fourier Grenoble I	Examinateur
M. Hervé CHAURIS, Professeur, MINES ParisTech	Examinateur
Mme Daniela DONNO, Maître de conférences, MINES ParisTech	Examinatrice
M. François AUDEBERT, Chercheur senior, Total	Invité

MINES ParisTech
Centre de Géosciences
35 rue Saint-Honoré, 77300 Fontainebleau

T H È S E

Acknowledgements

I would like to thank all the people who have contributed to this thesis work.

First of all, I would like to thank my advisers Hervé Chauris and Daniela Donno. I thank them for their availability, patience, support and encouragement given along my stay in the geophysics group. I thank them also for having trusted in my ability to develop this research even though I had not much experience in geophysics before the thesis. I especially acknowledgement their patience with me and explain to me even the basic knowledge and formulations. This really helps me to start the thesis in the first year. The discussions with them have always been helpful for me. Thanks for their availability even on weekends to answer my questions or to help me revising reports and presentations. Their passion and persistence for the research also encourage me.

I would like to thank the members of my jury. Thanks to Paul Sava and Gilles Lambaré for having accepted to review my thesis in detail. Thanks to Stéphane Operto and Romain Brossier for examining my thesis.

I thank Total for financially and technically supporting my thesis. I especially acknowledge Henri Calandra from Total who has brought me to the area of geophysics and initiated this PhD project. I thank François Audebert from Total for providing the real data set used for the real application in this thesis and for his help to apply for the computational resources from Total. I thank Florent Ventimiglia from Total for kindly helping me with troublesome computer and account problems.

Many thanks to the members in the geophysics team of the Geosciences Center at MINES ParisTech. I especially thank Véronique Lachasse for her help to facilitate the administrative affairs since my first day at the office. Thanks to the Alexandrine Gesret and Mark Noble for their availability and advices.

I would like to thank all my officemates during the three and half years. Nidhal Belayouni, Sébastien Penz, Carlos Pérez, Elise Vi Nu Bas, Charles-Antoine Lameloise, Emmanuel Cocher, Sven Schilke, Yves-Marie Batany, Yubing Li and Hao Jiang for their friendship and support. Thanks Elise for offering me facilities when I arrived in Fontainebleau.

I also thank Dominique Vassiliadis from the Geosciences Center and Evelyne Peyrillou from the delegation to help me solving administrative matters regarding the visa, contract, residence cart, etc.

Finally, I would like to thank my family for always respecting and supporting my professional and life choices. My research would not have been possible without their support. Before continuing with the text of this research, I would like to especially thank my boyfriend Haisheng Wang who voluntarily came to France to accompany me and has always been at my side and supporting me during these years.

Abstract

The knowledge of the Earth subsurface poses economic, environmental, human and scientific issues. Seismic imaging is a procedure to image the Earth subsurface from the data observed at the surface. In the context of hydrocarbon exploration, seismic imaging techniques are widely used to characterise the first few kilometres of the Earth's interior. Full Waveform Inversion (FWI) is one of the efficient seismic imaging method. Recent advances in high performance computer make FWI feasible for large applications. In theory, FWI could reconstruct a high-resolution subsurface image provided that low frequency and wide angle/aperture/azimuth data are available. FWI is a data-fitting procedure and is resolved as an optimization problem. Depending on the frequency content of the data, the objective function of FWI may be highly nonlinear and has many local minima. If a data set mainly contains reflections, this problem particularly prevents the gradient-based methods from recovering the long wavelengths of the velocity model.

The model of the subsurface could be separated into two parts by scale separation. One part is the short-wavelength part which contains the singularities of the model and the other part is the long-wavelength part which is a smooth version of the model. In this thesis, I propose a variant of FWI based on the scale separation of these two parts to mitigate the nonlinearity of the problem.

In the first section, methodologies of the conventional FWI and the new proposed method are presented. The new method is a Reflection-based Waveform Inversion (RWI) method . It consists of decomposing the gradient of FWI into a short-wavelength part and a long-wavelength part and the inversion is performed in an alternating fashion between these two parts. The gradient decomposition is achieved by decomposing the wavefields into their one-way components. Different wavefield decomposition methods are also presented.

In the second section, we implement the FWI and the new method to several case studies. For numerical modeling, we use a finite-difference approach to resolve the acoustic wave equation with constant density in the time domain. The model update is based on the L-bfgs algorithm and the wavefield is decomposed using the 2D FFT-based method in the f-k domain. These case studies show the difficulties associated with FWI to recover the long-wavelength part of the velocity model when low frequency and large-offset data are absent, and the initial model is far from the true one. The new method shows its robustness in this case especially for constructing the long-wavelength model.

Keywords: Waveform inversion, wavefield decomposition, seismic imaging.

Résumé

La connaissance du sous-sol de la Terre pose les enjeux économiques, environnementaux, humains et scientifiques. L'imagerie sismique est une procédure à imager le sous-sol à partir des données observées à la surface. Dans le cadre de l'exploration des hydrocarbures, les méthodes d'imagerie sismique sont beaucoup utilisées pour caractériser les premiers kilomètres de l'intérieur de la Terre. Full Waveform Inversion (FWI) est l'une des méthodes d'imagerie sismique efficace. Les progrès récents en superordinateur rendent FWI possible pour les grandes applications. En théorie, FWI pourrait reconstruire une image du sous-sol à haute résolution, à condition que les données de basse fréquences et grand-angle/ouverture/azimut sont disponibles. FWI est une procédure d'ajuster des données, et est résolu comme un problème d'optimisation. En fonction du contenu en fréquence des données, la fonction objectif de FWI peut être fortement non linéaire et présente de nombreux minima locaux. Pour des données qui contiennent principalement des réflexions, ce problème empêche notamment les méthodes basées sur le gradient de retrouver les longues longueurs d'onde du modèle de vitesse.

Le modèle du sous-sol peut être séparé en deux parties par separation d'échelle. Une partie est la partie courte longueur d'onde qui contient les singularités du modèle. l'autre partie est la partie longue longueur d'onde qui est une version lisse du modèle. Dans cette thèse, nous proposons une variante de FWI basée sur la séparation d'échelle entre les deux parties pour atténuer la non-linéarité du problème.

Dans la première section, les méthodologies de la FWI classique et la nouvelle méthode proposée sont présentés. La nouvelle méthode est une méthode de l'inversion des formes d'ondes basées sur les réflexions. Il consiste à décomposer le gradient de FWI en une partie de courte longueur d'onde et une partie de longue longueur d'onde, et l'inversion est effectuée d'une manière alternée entre ces deux parties. La décomposition du gradient est obtenu par la décomposition des champs d'ondes en leurs composants unidirectionnels. Différentes méthodes de décomposition des champs d'ondes sont également présentées.

Dans la deuxième section, nous appliquons la FWI et la nouvelle méthode à plusieurs études de cas. Pour la modélisation numérique, nous utilisons une approche différence finie pour résoudre l'équation des ondes acoustiques avec une densité constante dans le domaine temporel. La mise à jour du modèle est basé sur l'algorithme L-BFGS et les champs d'ondes sont décomposés en utilisant la méthode basée sur FFT 2D dans le domaine f - k. Ces études montrent les difficultés liées à FWI pour récupérer la partie longue longueur d'onde du modèle de vitesse lorsque les données de basse fréquence et grands offsets sont absentes, et le modèle initial est loin du vrai modèle. La nouvelle méthode présente sa robustesse dans ce cas en particulier pour la construction du modèle de longue longueur d'onde.

Mots Clés: Inversion des formes d'ondes, décomposition des champs d'ondes, imagerie sismique.

Contents

List of figures

List of tables

Chapter 1

Introduction

Contents

Résumé du chapitre 1

L'imagerie sismique permet d'imager l'intérieur de la Terre, pour analyser les propriétés physiques du sous-sol. L'imagerie sismique pose des questions économiques, humaines, environnementales et scientifiques. Dans la géophysique d'exploration, cette technique est principalement utilisée pour la détection d'hydrocarbures. Avec le développement des ordinateurs de haute performance, l'imagerie sismique 3D est de plus en plus sophistiquée [Biondi, 2006], pour obtenir de meilleures résolutions. Le résultat de l'imagerie sismique 3D est similaire à une imagerie médicale. Mais l'imagerie sismique se prolonge plusieurs kilomètres ou plus dans la Terre. Et surtout, les sources et récepteurs ne sont qu'à la surface. Les méthodes d'imagerie sismique continuent à s'améliorer, et les ordinateurs les plus avancés maintenant permettent aux scientifiques de traiter les données en quelques jours plutôt qu'en mois, et accélèrent la découverte et la production finale du pétrole et du gaz.

Dans l'imagerie sismique, la propriété physique nous voulons imager est généralement le modèle de vitesse de la propagation des ondes, en utilisant les données sismiques enregistrées à la surface pendant l'acquisition terrestre ou marine. Les données enregistrées apportent des informations sur le sous-sol. La loi physique reliant le modèle de vitesse et les données est décrite par l'équation de propagation des ondes. Cette relation est habituellement non linéaire, et l'imagerie sismique est considérée comme un problème d'optimisation non linéaire.

Le modèle de vitesse que nous cherchons à caractériser pourrait être séparé en deux parties. Une première partie est la partie des grandes longueurs d'onde qui peut être considérée comme une version lisse du modèle, et qui est désignée comme le macro-modèle. L'autre partie est la partie des courtes longueurs d'onde qui contient toutes les singularités du modèle, et est appelée le modèle de réflectivité ou l'image migrée. La procédure standard dans l'imagerie sismique consiste à reconstruire les grandes longueurs d'onde en premier, suivie par la reconstruction des courtes longueurs d'onde. La qualité de la mise à jour des courtes longueurs d'onde dépend de l'exactitude du macro modèle.

Avec le développement des ordinateurs de haute performance, l'imagerie sismique 3D et 4D s'est développée. Elle fournit des images du sous-sol de haute résolution. Néanmoins, l'imagerie sismique présente plusieurs limitations communes. Par exemple, la résolution de l'imagerie sismique est limitée par la bande de fréquence de la source, de l'atténuation et de la géométrie d'acquisition. A cause de la bande limitée de la source et de l'acquisition finie, souvent il est difficile d'obtenir une résolution suffisante. De plus, les basses fréquences dans la source et les données de grands offsets et grandes ouvertures sont importantes pour la reconstruction des grandes longueurs d'ondes du modèle. Cependant, les données de grands offsets et grandes ouvertures présentent des défis de déploiement et financiers.

L'inversion des forme d'ondes (en anglais, Full Wavefrom Inversion, FWI) est l'une des techniques principales de l'imagerie sismique. Nous pouvons citer [Fichtner, 2010] et [Virieux and Operto, 2009] pour une revue tutorielle de cette méthode. FWI définit un problème inverse non linéaire dans le domaine des données [Tarantola, 1984, Lailly, 1983], cherchant à minimiser les moindres carrés des différences entre les données observées et simulées. Le principe de FWI est d'utiliser tous les types d'ondes (ondes directes, réflexions, réfractions, et multiples) pour résoudre les différents paramètres du modèle (la vitesse, la densité, l'atténuation, ...). Les informations du champ d'ondes complet

(phase et l'amplitude) sont tous pris en considération en même temps, sans avoir besoin d'introduire explicitement le temps de trajet. FWI est capable de traiter des modèles complexes et de fournir des images de haute résolution. Dans l'industrie, FWI est principalement utilisée pour l'exploration pétrolière. Toutefois, elle est également appliquée à d'autres domaines, tels que l'électromagnétisme.

Les méthodes d'optimisation utilisées pour FWI pourraient être divisées en méthodes d'optimisation globales et méthodes d'optimisation locales. Les méthodes d'optimisation globales, telles que Monte-Carlo [Jin and Madariaga, 1994], les algorithmes génétique [Jin and Madariaga, 1993], permettent potentiellement d'explorer l'ensemble de l'espace pour trouver le minimum global de la fonction objective. Les méthodes d'optimisation locales, telles que les méthodes basées sur le gradient [Tarantola, 1984], commencent à partir d'un modèle initial, et calculent la direction de la mise à jour du modèle en utilisant le gradient de la fonction objective. Le gradient peut être calculé en utilisant la méthode de l'état-adjoint [Plessix, 2006], puis le modèle est mis à jour et des itérations sont répétées jusqu'à ce que la critère de convergence est remplie. Bien que les méthodes globales soient plus robustes pour résoudre les problèmes non linéaires, elles demandent un coût de calcul élevé, et la taille de l'espace des modèles et la taille des données à traiter sont au-delà de la capacité de calcul actuelle pour les applications réelles. Généralement, une méthode d'optimisation locale est utilisée pour la FWI et le problème d'optimisation est résolu de manière itérative pour compenser la non-linéarité [Tarantola, 1984].

FWI a été initialement appliquée en temps [Tarantola, 1984, Lailly, 1983]. La première application 2D de FWI en géophysique d'exploration a été mise en œuvre par [Gauthier et al., 1986]. [Tarantola, 1986] and [Mora, 1987] ont appliqué FWI en temps au cas élastique. FWI appliquée sur des données réelles est mis en œuvre par [Crase et al., 1990]. FWI en fréquence a été proposée depuis les années 90 [Pratt and Worthington, 1990, Pratt, 1999]. FWI multi-échelle a été étudiée dans le domaine temporel [Bunks et al., 1995] et dans le domaine fréquentiel [Sirgue and Pratt, 2004] en vue d'atténuer la non-linéarité du problème inverse. Cette approche consiste à d'abord inverser les données de plus basses fréquences et d'introduire progressivement des fréquences élevées dans l'inversion. Avec le développement de l'ordinateur haute performance, FWI est entré dans une nouvelle ère depuis 2009, lorsque les premières applications 3D de FWI ont été obtenus avec des données réelles [Plessix, 2009, Sirgue et al., 2010].

Un problème inverse est considéré comme le pendant du problème direct qui relie les paramètres de modèle avec les données que nous observons. Le problème direct de FWI simule la propagation des ondes dans un modèle donné. Les données réelles sont décrites plus précisément par la modélisation élastique plutôt que la modélisation acoustique, où la vitesse des ondes P, la vitesse des ondes S, la densité et l'atténuation sont tous considérées. Toutefois, il est encore une pratique courante d'appliquer la FWI acoustique sur des données réelles en considérant la propagation des ondes acoustiques. L'équation des ondes acoustiques pourrait être résolue dans le domaine temporel [Tarantola, 1984, Mora, 1987, Mora, 1989] ou dans le domaine fréquentiel [Pratt and Worthington, 1990, Sirgue and Pratt, 2004]. La résolution numérique de l'équation différentielle partielle pourrait être réalisée par des méthodes numériques, telles que la méthode des différences finies [Virieux, 1986, Moczo et al., 2004], éléments finis [Marfurt, 1984], etc. Lorsque la discrétisation spatiale est effectuée sur une grille régulière, la méthode des différences finies est généralement choisie car elle est rapide et facile à mettre en œuvre.

FWI est capable de potentiellement imager des structures complexes. Cependant, elle présente plusieurs limitations [Virieux and Operto, 2009]. Pour atténuer la nonlinéarité de FWI, de nombreuses formulations alternatives sont proposées, qui sont principalement liées aux choix de la représentation des données, aux choix de la norme des résidus des données, et aux choix du critère d'optimisation.

Lorsqu'on traite les données de sismique réflexion, l'utilisation des méthodes d'inversion basées sur les réflexion (en anglais Reflection Waveform Inversion, RWI) [Mora, 1989, Chavent et al., 1994, Xu et al., 2012, Biondi et al., 2012, Tang and Lee, 2013, Wang et al., 2013, Brossier et al., 2015, Zhou et al., 2015, Wang et al., 2015b] semblent prometteuses. Ces approches reposent sur la séparation explicite du modèle en un modèle de grandes longueurs d'onde et un modèle de réflectivité. Le modèle de réflectivité est obtenu en premier par la migration et ensuite se comporte comme les sources primaires en profondeur, ce qui permet de produire les transmissions entre les réflecteurs et la surface. Ces transmissions sont utiles pour la mise à jour du modèle de grandes longueurs d'ondes.

Des données de réflexions contiennent des informations de temps de trajet ainsi que l'amplitude, donc l'inversion de la forme d'onde des réflexions devrait en principe être capable de reconstruire à la fois les grandes longueurs d'ondes et les courtes longueurs d'ondes [Snieder et al., 1989, Hicks and Pratt, 2001]. Toutefois, en pratique, la qualité de l'inversion est dégradée à cause des données à bande limitée et offsets limités. Dans le domaine temporel, lorsqu'une grande gamme de fréquences est utilisée en même temps, ces deux composantes sont mélangées pendant la FWI. Ce couplage est un problème majeur pour la FWI [Snieder et al., 1989], puisque la mise à jour des grandes longueurs d'ondes a généralement une amplitude plus faible que la mise à jour des courtes longueurs d'onde. Par conséquent, lorsque ces deux composants sont mélangées ensemble, la mise à jour du modèle est dominée par les courtes longueurs d'ondes. Il est donc naturel d'inverser séparément ces deux parties.

Dans cette thèse, nous proposons une méthode d'inversion en deux étapes obtenue par la décomposition de la formule de gradient de FWI en une partie de grandes longueurs d'ondes et une partie de courtes longueurs d'ondes [Wang et al., 2013].

Dans le chapitre 2, nous passons en revue les notation basiques de la FWI, y compris la fonction objective et la résolution du problème direct en utilisant le schéma de différences finies. Le calcul du gradient de la fonction objective et les méthodes d'optimisation locales sont également détaillées. La résolution du gradient est analysée pour étudier la résolution de FWI.

Dans le chapitre 3, nous présentons la méthode d'inversion proposée, et montrons comment les deux étapes d'inversion séparés, soit une pour la mise à jour des courtes longueurs d'ondes et l'autre pour la mise à jour des grandes longueurs d'ondes, sont formulés. Nous illustrons cette approche en utilisant un modèle de couche 1D. Différentes méthodes de décomposition des champs d'ondes sont également introduites, l'une dans le domaine de Fourier, l'une en utilisant le vecteur de Poynting et la troisième dans le domaine des curvelets.

Dans le chapitre 4, nous appliquons la FWI, la FWI multi-échelle et la nouvelle méthode d'inversion proposée sur deux modèles synthétiques et comparons les résultats. Nous analysons aussi l'influence des paramètres clés pendant l'inversion sur le résultat d'inversion.

Dans le chapitre 5, nous appliquons la FWI et la nouvelle méthode d'inversion proposée

sur un jeu de données réelles, et présentons les pré-traitements effectués sur les données et les difficultés que nous rencontrons avec l'hypothèse acoustique de la propagation des ondes.

1.1 Seismic imaging

1.1.1 General principles

Seismic imaging allows to image the Earth's interior, to analyze the physical properties of the subsurface. It poses economic, human, environmental and scientific issues. In exploration geophysics, seismic imaging is mainly used for the detection of hydrocarbons. With the development of the high performance computers, sophisticated 3D seismic imaging is investigated [Biondi, 2006], yielding high-resolution images of the subsurface. The result of 3D seismic imaging is similar to an X-ray scan or medical imaging that extends several kilometres or more into the Earth. 4D seismic imaging [Zhang et al., 2007] allows imaging potential fluid migration paths within the reservoir, by inverting for time-lapse parameters. Methods for seismic imaging continues to improve, and more advanced computers now enable scientists to accelerate the data processing, speeding up the discovery and final production of oil and gas.

In seismic imaging, the physical property we want to image is usually the model of the wave propagation velocity, using the seismic data recorded during land or marine acquisitions. Seismic acquisitions use controlled seismic sources (explosive and vibroseis for land acquisitions or air guns for marine acquisitions). Then, seismic waves propagate in the subsurface and reflected or refracted waves are recorded by the seismic sensors placed at the surface (geophones for land acquisitions or hydrophones for marine acquisitions). The recorded data bring information of the subsurface. The physical law relating the velocity model and the data is described by the wave propagation equation. This relationship is usually nonlinear, and seismic imaging is considered as a nonlinear optimization problem.

1.1.2 Review of classic methods/scale separation

The wave-propagation velocity model that we would like to characterize could be divided into two parts. One part is the long-wavelength part which can be considered as a smooth version of the model, and is referred to as the macro or background model. The other part is the short-wavelength part which contains all the singularities of the model, and is referred to as reflectivity model or migrated image. The standard procedure in seismic imaging consists of recovering the long wavelengths first, followed by the reconstruction of the short wavelengths, as the quality of the short-wavelength update depends on the accuracy of the background model.

The background model could be recovered either by traveltime tomography [Bishop et al., 1985, Pratt and Chapman, 1992, Billette and Lambaré, 1998, Woodward et al., 2008] or by Migration Velocity Analysis (MVA) [Symes and Carazzone, 1991, Chauris and Noble, 2001, Sava and Biondi, 2004]. Traveltime tomography uses the traveltimes to estimate the wave propagation velocity. In industry, the ray-based tomography approaches involved migration, residual move-out picking, demigration and linear or non-linear velocity updates. Nonlinear slope tomography uses a local focusing criterion without using a pre-defined shape of the reflectors or of the RMO curves [Guillaume et al., 2008].Migration Velocity Analysis (MVA) is performed in the image domain and aims at building a relationship between migrated image perturbation and model perturbation. This approach a priori does not require picking.

The short wavelengths provide the fine structure of the subsurface model, allowing to localize reflectors in depth. The short wavelengths can be obtained by migration techniques, assuming the background velocity is correct. There are two major categories of migration methods: ray-based methods [Beylkin, 1985, Bleistein, 1987], which are based on the high frequency assumption, and wave-equation based methods [Baysal et al., 1983, Whitmore et al., 1983]. [Etgen et al., 2009] gives a comparison of different migration methods. The principle of migration is formulated by [Claerbout, 1971], and it consists of propagating the source signal and the recorded data into to the medium and cross correlate these two wavefields. The zero-lag cross correlation gives the locations of reflectors. There exist other imaging conditions, such as deconvolution-based imaging condition [Valenciano et al., 2003], source/receiver-normalized imaging condition [Kaelin et al., 2006], extended imaging condition [Sava and Fomel, 2006]. [Chattopadhyay and McMechan, 2008] and [Sava and Hill, 2009] give a summary of the imaging conditions. The classical correlation-based migration is qualitative, as it only provides a reflectivity image. Alternately, quantitative migration [Lambaré et al., 1992, Jin et al., 1992, Lameloise et al., 2015, Symes, 2015] allows imaging the values of the physical parameters. Recent developments have shown that for migration-based velocity analysis, quantitative migration is preferable as it provides a more accurate migration image.

1.1.3 Difficulties in seismic imaging

With the development of advanced computers, 3D and 4D seismic imaging are investigated, yielding more accurate subsurface images. Nevertheless, seismic imaging presents several common limitations:

- Insufficient resolution. The quality of seismic imaging resolution is limited by the source wavelet bandwidth, the wave propagation attenuation and the acquisition geometry. In the framework of the single-scattering approximation [Devaney, 1982, Miller et al., 1987], the wavenumber at a point of the model depends on the local wavelength and the scattering angle (Figure 1-1).

  According to Figure 1-1, high resolution can be obtained with small reflection angle and small wavelength, and small wavelength corresponds to small velocity and high frequency. However, during the wave propagation, waves suffer from attenuation effect. Apart from the geometric spreading, intrinsic attenuation is also presented due to the nature of some material. High frequency waves are particularly attenuated during the propagation, making difficult to image the deep part of the Earth. Besides, as we will show in Chapter 2, numerical resolution of the wave equation imposes the stability condition on the spatial sampling, which depends on the maximum frequency of the source wavelet. The use of a high frequency source requires finer spatial and temporal sampling, making the problem more computational extensive.

- Acquisition limits. The quality of the migrated image depends closely on the acquisition geometry and acquisition devices. [Mora, 1989] and [Pratt et al., 1996] show that the long wavelengths of the model could be retrieved through the use of long-offset (diving waves, refractions) and transmission data. If the data lacks of low frequencies, then long-offset and wide-aperture/azimuth acquisitions [Sirgue

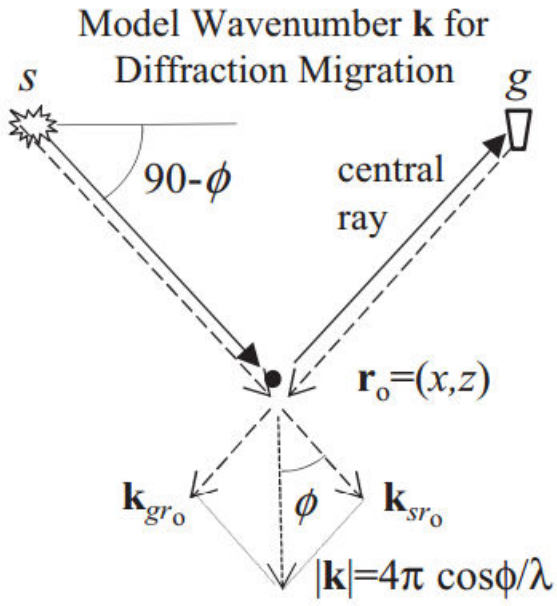


Figure 1-1: Wavenumber at a scattering point [Huang and Schuster, 2014].

et al., 2010, Shipp and Singh, 2002] are necessary in order to image deep parts of the model [Sirgue, 2006]. However, long-offset and wide-aperture/azimuth acquisitions usually present deployment and financial challenges.

- Requirement of good velocity model. [Versteeg, 1993] analysed the sensitivity of the depth migrated image to the velocity model and found that the accuracy of the depth image closely depends on the accuracy of the long wavelengths of the model, and, if an incorrect model is used, the migration image is unfocused and reflectors are mispositioned. In practice, the inversion of the background velocity model still remains a challenge.

1.2 Full waveform inversion

We now concentrate on a very popular method used for imaging the Earth. We refer to [Virieux and Operto, 2009] and [Fichtner, 2010] for recent reviews.

1.2.1 History

Full waveform inversion (FWI) is one technique for seismic imaging which develops rapidly. FWI defines a nonlinear inverse problem in the data-domain [Tarantola, 1984, Lailly, 1983], seeking to minimize the least-squares differences between observed and simulated data. The principle of FWI is to use all types of waves (direct waves, reflections, refractions, and multiples) to resolve different model parameters (velocity, density, attenuation). The information of the complete wavefield: traveltime, phase and amplitude are all taken into consideration at once. FWI is capable of dealing with complex models and of delivering high resolution images with a resolution of half the minimal wavelength [Sirgue and Pratt,

2004]. In industry, FWI is mainly used for the oil exploration; however, it could also be applied to other domains, such as electromagnetics and medical imaging.

The optimization methods used to solve the FWI nonlinear problem could be divided into global optimization methods and local optimization methods. Global optimization methods, such as Monte-Carlo [Jin and Madariaga, 1994], genetic [Jin and Madariaga, 1993] and simulated annealing algorithms [Sen and Stoffa, 1991], allow to potentially explore the whole model space to find the global minimum of the objective function. Local optimization methods, such as gradient-based methods [Tarantola, 1984], start from an initial model, and calculate the model update direction by computing the gradient of the objective function. The gradient could be calculated using the adjoint-state method [Plessix, 2006] and then the model is updated and iterations will be repeated until the convergence criterion is met. Although global methods are more robust to handle nonlinear problems, they require a high computational cost, and the size of the model space and the size of data to process is beyond current computation capacity for real data applications. Usually a local optimization method is used and the optimization problem is resolved iteratively to compensate the nonlinearity [Tarantola, 1984].

FWI was originally applied in the time domain [Tarantola, 1984, Lailly, 1983]. The first 2D application of FWI in exploration geophysics was implemented by [Gauthier et al., 1986]. [Tarantola, 1986] and [Mora, 1987] applied the time-domain FWI to the elastic case. FWI applied on real data set is implemented by [Crase et al., 1990]. The frequency-domain FWI was proposed since the '90s [Pratt and Worthington, 1990, Pratt, 1999]. The frequency-domain FWI is equivalent to the time-domain FWI when all frequencies are considered simultaneously [Pratt et al., 1998]. From a series of discrete frequencies, FWI in the frequency domain aims at retrieving model perturbations [Sirgue, 2006]. Rules have been designed to select frequencies [Sirgue, 2006]. Multiscale FWI was investigated in the time domain [Bunks et al., 1995] and in the frequency domain [Sirgue and Pratt, 2004] in order to mitigate the nonlinearity of the inverse problem. The method consists of starting by inverting the lower frequency data and of progressively introducing high frequencies in the inversion. Moreover, long-offset data could be used to retrieve first the long wavelengths of the model [Shipp and Singh, 2002, Sirgue, 2006].

With the development of high performance computing, FWI entered in a new era since 2009, when the first 3D FWI applications were successfully obtained with real data [Plessix, 2009, Sirgue et al., 2010]. The redundancy of the 3D acquisitions, together with the availability of long-offset and wide-azimuth data allow reducing the non linearity of the problem.

1.2.2 Physics of wave propagation

An inverse problem is considered as the "inverse" of the forward problem which relates the model parameters to the data we observe. The forward problem in FWI simulates the wave propagation in a given model.

Among the recorded seismic waves, we can distinguish between P waves and S waves. P waves, also called pressure waves or primary waves, are characterized by particle motion in the same direction as wave propagation. S waves, also called shear waves or secondary waves, are characterized by particle motion in the plane perpendicular to the direction of wave propagation. P waves and S waves are all body waves but surface waves could also

be present in real data. Real data are more accurately described by elastic rather than acoustic modeling, where P-velocity, S-velocity, density and attenuation are all considered. However it is still common practice to apply acoustic FWI on real data. By considering the acoustic wave propagation, S waves are neglected and only the P waves are modeled. The acoustic wave equation could be resolved in the time domain [Tarantola, 1984, Mora, 1987, Mora, 1989] or in the frequency domain [Pratt and Worthington, 1990, Sirgue and Pratt, 2004]. The numerical resolution of the partial differential wave equation could be achieved by numerical methods, such as finite difference [Virieux, 1986, Moczo et al., 2004], finite element [Marfurt, 1984], etc. When the spatial discretization is done using a regular grid, the finite-difference method is usually chosen because it is fast and easy to implement.

1.2.3 Current limitations of FWI

FWI is capable of potentially imaging complex structures. However, it presents several limitations [Virieux and Operto, 2009]:

- Local minima. The objective function of FWI is highly nonlinear with respect to the model parameters and has many local minima [Bunks et al., 1995]. These local minima prevent gradient-based techniques from finding the global minimum if the initial model is far from the global solution. This is known as the cycle-skipping problem. A phase mismatch of less than one half of the wavelength is required between the exact model and the initial model, to avoid falling into a local minimum.

- Lack of low frequency data. [Claerbout, 1982] and [Jannane et al., 1989] first demonstrate that reflectivity model is linearly derived from the reflected waves in the data, while the large-scale velocity model does not linearly depend on the data. It seems that the data are not sensitive to middle range scale wavelengths. To avoid the non-linearity, multiscale FWI is needed to reconstruct the model, starting from low frequencies and progressively adding higher frequencies. When the low frequencies are missed in the data, which is often the case with real data, the FWI usually fail to retrieve the long-wavelength part of the model, and only acts as a migration mode.

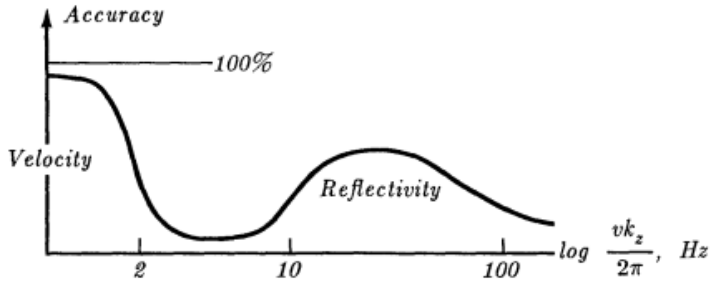


Figure 1-2: Separation of scales between the frequency content of the velocity model and of the seismic data [Claerbout, 1982].

- Multi-parameter inversion. The principle of FWI is to use all types of waves (direct waves, reflections, refractions, and multiples) to resolve different model parameters

(velocity, density, attenuation). Since the '90s, FWI has been mainly used for imaging the P-wave velocity model by considering the acoustic wave propagation. This approximation is quite valid when FWI is mainly based on direct arrivals (diving waves, transmissions), as these waves are less sensitive to density perturbations. But for reflected data, which are sensitive to density perturbations, the P-wave velocity model does not always allow the reliable interpretation of the complex structure of the subsurface, as the density model, the S-wave velocity model and the attenuation factor all have a great influence on the amplitude of reflection data. [Mulder and Plessix, 2008] study the effect of acoustic inversion of elastic data through a marine data set and conclude that variable-density acoustic inversion of marine data can have some value if the subsurface is not too complex and the target is not too deep. For complex models, multi-parameter inversion should be considered. However simultaneous inversion of several parameters is still challenging. The main obstacle is that different parameters have coupling effects and different orders of magnitude, making the inversion poorly conditioned and more nonlinear. In [Operto et al., 2013, Prieux et al., 2013, Malinowski et al., 2007], the authors propose guidelines to choose suitable strategies for multi-parameter FWI.

1.2.4 Alternative formulations

To mitigate the non-linearity of FWI, many alternatives formulations are proposed, which are mainly related with:

- Choice of the representations of the data, such as logarithm of the complex value of the data [Shin and Min, 2006], separation of data amplitude and phase [Shin and Min, 2006, Bozdağ et al., 2011], envelope of the data [Chi et al., 2014, Wu et al., 2014, Bozdağ et al., 2011], normalized integration of the data [Donno et al., 2013], energy flux of the data [Causse, 2002]. Among these possible alternatives, the Laplace-domain FWI [Shin and Cha, 2008] has been shown to be effective to build a smooth velocity model. The authors show that by transforming the wave in the Laplace domain, the objective function of the ℓ_2 norm of the logarithmic wavefield appears to be more convex and artificial frequencies smaller than the frequencies in the source wavelet are created.

- Choice of data residual norm. The ℓ_2 norm is based on the assumption of Gaussian distribution of data uncertainties [Tarantola, 2005]. It may be not valid for all cases. Moreover, it is sensitive to large errors [Tarantola, 2005]. ℓ_1 norm [Crase et al., 1990], which is not based on Gaussian assumption, has been proven to be more robust in the presence of noise in the data [Brossier et al., 2010]. Besides, as ℓ_1 norm ignores the amplitude in the residual, the inversion is less sensitive to the large data error [Virieux and Operto, 2009]. We can also cite the Huber norm [Guitton and Symes, 2003, Ha et al., 2009], the sech norm and the Cauchy criterion norm [Crase et al., 1990, Amundsen, 1991]. These alternative norms could be considered as intermediate between ℓ_1 norm and ℓ_2 norm.

- Choice of minimization criterion. The conventional criterion measures the least-squares data misfit which could be quite nonlinear with respect to model. Other

criterion are proposed to mitigate the nonlinearity of the problem, such as data differential based criterion [Chauris and Plessix, 2012] or data correlation based criterion [van Leeuwen and Mulder, 2008]. [van Leeuwen and Mulder, 2008] show that compared to the least-squares functional, most of the weighted norms of the correlation function have a large basin of attraction and respond smoothly to a perturbation of the true velocity model.

1.3 Reflection-based waveform inversion

1.3.1 Principles

When dealing with reflection data, the use of Reflection-based Waveform Inversion (RWI) [Mora, 1989, Chavent et al., 1994, Xu et al., 2012, Biondi et al., 2012, Tang and Lee, 2013, Wang et al., 2013, Brossier et al., 2015, Zhou et al., 2015, Wang et al., 2015b] appears promising. The method relies on the explicit separation of the model into a background model and a reflectivity model. The reflectivity model is inverted first by migration and then serves as primary sources in depth, which allows computing the transmission wavepaths from the reflectors to the surface. These transmission wavepaths are useful for the update of the background model. A new reflectivity model must be generated by migration according to the background velocity model update at each nonlinear iteration and the two-step workflow is iterated until a fixed convergence criterion is met [Brossier et al., 2015].

Migration Based Traveltime Tomography (MBTT) [Chavent et al., 1994] is based on a combination of migration and modelling. It combines the least-squares migration and the multiscale FWI to mitigate the nonlinearity. [Tang and Lee, 2013] use the wavefield decomposition to separate the migration part and the tomography part in the gradient. Their idea is to mix these two parts together and set different weights to enhance the tomography part. [Tang and Lee, 2013] determine, at each iteration, the weights by solving an optimization problem, which is not a trivial task. The advantage is that, it is easier to compute the misfit function at each iteration and then to control the convergence. [Brossier et al., 2015] propose a reflection-based full waveform inversion. They use a correlation-based misfit function instead of the classic least-squares data misfit function to avoid cycle-skipping. Besides, the inversion is performed in the pseudo-time domain instead of depth domain to avoid recomputing the reflectivity at each iteration of the inversion. [Zhou et al., 2015] propose to introduce diving waves along with reflections to improve the reconstruction of the shallow parts of the model, which in turn improves the imaging of the deeper parts. [Alkhalifah and Wu, 2015] combine FWI and Migration Velocity Analysis (MVA) to generate a new objective function.

1.3.2 Limitations of RWI

In recent years, RWI has shown its robustness compared to conventional FWI. However, this method is not mature yet, and several limitations could be listed. Firstly, as RWI is based on the reflection data, other waves, such as diving waves, refractions, which are useful for the determination of the long-wavelength model, are neglected. [Zhou et al., 2015] propose to combine the reflections with the diving waves to enhance the long-wavelength

components of the model. [Wang et al., 2015b] propose to combine the reflections with refractions. In [Wang et al., 2013] and [Tang and Lee, 2013] the proposed methods do not use reflections explicitly and could a priori incorporate diving waves and refractions automatically. Further investigations on this point are needed.

Secondly, during the inversion, at each nonlinear iteration a new reflectivity must be recomputed, causing a slow convergence rate. [Brossier et al., 2015] propose to perform the migration in the pseudo-time domain as the zero-offset traveltime is preserved during the inversion, to avoid the re-computation of the reflectivity model at each iteration.

Thirdly, the nonlinear step is sensitive to the reflectivity model obtained in the migration step. Thus building a true-amplitude reflectivity is important. The reflectivity is usually obtained by migration methods. However, as the migration uses the adjoint of forward-modeling operator instead of the inverse operator, a single iteration does not necessarily preserve the true amplitude of migration images. Alternatively Least-Squares Migration (LSM) can be used. LSM is an iterative migration method resolving the linearized inversion problem. It has been shown to improve amplitude information and to better focus migrated images. Several authors show that the LSM provides more reliable results than conventional migration images [Clapp, 2005, Valenciano, 2008, Zhang et al., 2005].

1.4 Objective and outline of the thesis

Seismic reflection data contain traveltime information as well as reflection-amplitude information, therefore waveform inversion of reflections should in principle be able to recover both the long wavelengths and the short wavelengths of the model [Snieder et al., 1989, Hicks and Pratt, 2001]. However, in practice, the quality of the inversion is degraded due to the band-limited and offset-limited data. In the time domain, when a large frequency range is used at once, these two components are coupled during FWI. This coupling is a major problem for FWI [Snieder et al., 1989], as the long-wavelength update usually has a smaller amplitude than the short-wavelength update. Therefore, when these two components are mixed together, the velocity update is mainly driven by the short-wavelength update. It is therefore natural to try to invert separately the long-wavelength and the short-wavelength components of the velocity model.

Following [Xu et al., 2012] and [Zhou et al., 1995], we propose a two-step inversion workflow achieved by decomposing the gradient formula of FWI into the long wavelength part and short wavelength part [Wang et al., 2013]. From the literature, the FWI gradient is computed as the cross-correlation of the forward propagated source wavefield and the back-propagated residual wavefield [Tarantola, 1984, Lailly, 1983, Plessix, 2006]. With two-way modeling, both down-going and up-going components are present in the propagating wavefields. After decomposition of the forward and back-propagated wavefields into their down- and up-going components, the correlation of the two initial wavefields actually gives four terms. The back-scattered correlations provide the positions of the reflectors (short-wavelength components), while the forward-scattered correlations give information along the propagation paths (long-wavelength components). For Reverse-Time Migration (RTM) imaging, the forward-scattered correlations are usually removed because they create long-wavelength artifacts in the final migrated image [Liu et al., 2011, Yoon and

Marfurt, 2006]. But for velocity analysis, these correlations help in updating the long wavelength components of the velocity model [Diaz and Sava, 2012, Wang et al., 2013, Brossier et al., 2015].

In Chapter 2, I review the basics of FWI, including the objective function, and the resolution of the direct problem using the finite-difference scheme. The calculation of the gradient of the objective function and the available local optimization methods are also detailed.

In Chapter 3, I outline the proposed inversion procedure, and show how the two separate inversion steps, either for short-wavelength or long-wavelength update, are formulated. I illustrate the method using a 1D layer model. Different wavefield decomposition techniques are also introduced, one in the Fourier domain, one using the Poynting vector and an alternative in the curvelet domain.

In Chapter 4, I apply FWI, multiscale FWI and the proposed inversion method on two synthetic models and compare the results. I also analyze the influence of key parameters during the inversion.

In Chapter 5, I apply FWI and the proposed inversion method on the Brunei real dataset, and present the preprocessings on the data and the difficulties we meet with the acoustic assumption.

Publications:

- Wang, F., Chauris, H., Donno, D., and Calandra, H. "Taking advantage of wave field decomposition in full waveform inversion." 75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013. 2013
- Wang, F., Chauris, H., Donno, D., Audebert, F., and Calandra, H. "Full waveform inversion based on wavefield decomposition." SEG summer research workshop. 2015.
- Wang, F., Chauris, H., Donno, D., Audebert, F., and Calandra, H. "Full waveform inversion based on wavefield decomposition." Geophysics. Submitted.

Chapter 2

Full waveform inversion

Contents

Résumé du chapitre 2

Un problème inverse vise à reconstruire le modèle du sous-sol à partir d'un ensemble de données observées. L'imagerie sismique définit un problème inverse en géophysique. En entrée, les données sismiques sont enregistrées aux positions des récepteurs. Les solutions que nous recherchons sont les propriétés physiques du sous-sol. Le problème directe consiste à simuler les données sismiques dans un modèle connu. L'inversion des formes d'ondes est une procédure pour ajuster les données. Le principe de FWI est d'utiliser tous les types d'ondes (ondes directes, réflexions, réfractions, multiples, ...) pour résoudre les différents paramètres du modèle (la vitesse, la densité, l'atténuation). Le champ d'onde n'a pas besoin d'être décomposé en temps de trajet et amplitudes par exemple.

FWI a été proposée par [Lailly, 1983] et [Tarantola, 1984]. Cette méthode est largement utilisée dans la sismologie globale et dans l'exploration pétrolière. Parmi les très grand nombre de références, nous pouvons citer [Fichtner, 2010] et [Virieux and Operto, 2009] pour une revue tutorielle de cette méthode.

Dans ce chapitre, nous introduisons les formulations classiques de FWI, y compris la fonction objective, la résolution du problème direct en utilisant le schéma des différences finies. Le calcul du gradient de la fonction objective et les méthodes d'optimisations locales sont également détaillées. L'analyse de résolution du gradient est mise en œuvre pour étudier l'influence de la configuration d'inversion sur la résolution du gradient.

La fonction objective de la FWI vise à minimiser l'écart entre les données enregistrées et les données calculées [Tarantola, 1984]. FWI est généralement considérée comme un problème d'optimisation. Vue l'échelle du modèle, souvent les méthodes d'optimisation locales sont utilisées.

Le problème direct de FWI consiste à modéliser la propagation des données sismiques dans un modèle donné. Les données observées sont obtenues par la résolution de l'équation des ondes. Dans le cadre de cette thèse, nous nous limitons à l'équation des ondes acoustiques. L'équation des ondes acoustiques peut être résolue dans le domaine temporel [Tarantola, 1984, Mora, 1987, Mora, 1989] ou fréquentiel [Pratt and Worthington, 1990, Sirgue and Pratt, 2004]. Elle peut être résolue analytiquement ou numériquement. Les méthodes numériques, par exemple, différences finies [Virieux, 1986, Crase et al., 1990, Moczo et al., 2004], éléments finis [Marfurt, 1984, Choi et al., 2008], sont généralement utilisées. Dans cette thèse, nous utilisons la méthode des différences finies, d'ordre 4 en temps et d'ordre 8 en espace. La méthode des différences finies est un moyen naturel pour résoudre des équations aux dérivées partielles car elle est rapide et facile à mettre en œuvre. Elle ne repose pas sur l'hypothèse haute fréquence. Elle est donc plus générale que les méthodes basées sur les rais et fournit une implémentation relativement efficace par rapport aux méthodes des éléments finis. La méthode des différences finies estime directement chaque terme différentiel en utilisant le développement de Taylor sur une grille régulière. En plus, nous utilisons les conditions absorbantes aux bords du modèle pour simuler un milieu infini.

Pour le problème d'inversion de données réelles, la signature de la source est généralement inconnue et doit être estimée en même temps que les paramètres du modèle. Dans cette thèse, nous utilisons la méthode proposée par [Pratt, 1999] pour estimer la source.

La mise à jour du modèle par FWI est obtenue avec le gradient et le Hessien de la fonction objective. Le gradient est la dérivée de premier ordre de la fonction objective

par rapport aux paramètres du modèle. Le gradient relie la perturbation des données à la perturbation du modèle. La mise à jour du modèle est basée sur le gradient. En fait, la direction inverse du gradient donne la plus grande pente de la fonction objective. Le gradient peut être calculé efficacement avec la méthode état adjoint [Plessix, 2006], et est obtenu comme la corrélation entre le champ direct et le champ rétropropagé des résidus. Dans ce sens, le gradient de FWI est similaire au principe d'imagerie proposée par [Claerbout, 1971].

Le Hessien est la dérivée du deuxième ordre de la fonction objective par rapport au modèle. L'inverse du Hessien permet de corriger le facteur d'atténuation géométrique et illumination inégale, et de déconvoluer le gradient causé par la bande limitée des données [Pratt et al., 1998, Brossier et al., 2009].

La résolution de l'inversion dépend directement de la résolution du gradient. La résolution du gradient est fonction des différents types d'ondes et des dispositifs de l'acquisition. Concrètement, les ondes directes, les ondes plongeantes, les ondes transmises et les multiples donnent une résolution plus basse que les ondes de première réflexion [Huang and Schuster, 2014], donc fournissent un gradient plus basse fréquence. Pour l'acquisition, l'inversion avec une source plus basses fréquences [Claerbout, 1985, Jannane et al., 1989] et des données de plus grands offsets [Mora, 1989, Pratt et al., 1996] fournissent un modèle plus basse fréquence. D'où viennent les stratégies d'inversion multi-échelles [Bunks et al., 1995, Sirgue, 2003] et la stratégie de multi-étapes qui commence par les grands offsets pour d'abord retrouver les grandes longueurs d'ondes du modèle.

Selon les méthodes utilisées pour calculer l'inverse du Hessien, les méthodes de mise à jour du modèle de vitesse peuvent être divisées en plusieurs catégories. Les méthodes basées sur le gradient supposent que le Hessien est un diagonal ne dépendant que d'un scalaire. Elles donnent une solution simple. La méthode du gradient conjugué combine la direction actuelle et la direction du gradient précédente pour accélérer la convergence. La méthode de Newton calcule le Hessien complet, tandis que le Gauss-Newton ne considère que le premier terme du Hessien. La méthode de quasi-Newton [Nocedal, 1980] repose sur l'estimation du Hessien, permettant de profiter de l'information contenue dans le Hessien aux itérations précdentes, sans payer le prix d'une vraie itération du Newton complet. L-BFGS est une variante de BFGS. Elle garde en mémoire seulement un nombre limité d'itérations, généralement entre 3 à 15 itérations, étant moins exigeante en mémoire que BFGS. L'algorithme L-BFGS sera utilisé dans les applications présentées dans les chapitres 4 et 5 de cette thèse.

Le succès des méthodes d'optimisation locales dépend de l'exactitude du modèle initial. Comme les fonctions objectives peuvent présenter des minima locaux, le modèle initial doit être situé proche du minimum global pour assurer la convergence . Si le déphasage entre les données calculées et les données observées ne dépasse pas la moitié de la période du signal, les méthodes d'optimisation ajusteront ces deux données sans ambiguïté de phase. Sinon, les méthodes d'optimisation ajusteront ces deux données avec un déphasage d'une ou plusieurs périodes, provoquant la convergence vers un minimum local. Ceci est le phénomène de "cycle-skipping". Le modèle initial pour la FWI peut être obtenu par tomographie du temps de trajet ou par l'analyse de vitesse par migration. Pour atténuer la non linéarité, la méthode multi-échelle est souvent utilisée.

2.1 Introduction

An inverse problem aims at reconstructing the underlying model from a set of observed data. Seismic imaging defines an inverse problem in geophysics. The input is the recorded seismic data at receivers positions, and the solution we search for are the physical properties of the subsurface. The forward problem consists of simulating the seismic data in a known model. Full Waveform Inversion (FWI) is an inverse problem and is defined as a data-fitting procedure. The principle of FWI is to use all types of waves (direct waves, reflections, refractions, and multiples) to resolve different model parameters (velocity, density, attenuation). The information of the complete wavefield: travel time, phase and amplitude are all taken into consideration. Full waveform inversion was pioneered by [Lailly, 1983] and [Tarantola, 1984]. This method is widely used in global seismology and in oil exploration. Among the extremely large number of references, we refer to [Fichtner, 2010] and [Virieux and Operto, 2009] for a tutorial review of the method.

In this chapter, I review the classic Full Waveform Inversion (FWI) formulation, including the objective function, the resolution of the direct problem using the finite-difference scheme. The derivation of the gradient of the objective function and the local optimization methods are also detailed. The resolution analysis of the gradient is carried out to study the influence of inversion setup on the gradient resolution.

2.2 Objective function

FWI is stated as a non-linear optimization problem. [Tarantola, 2005] demonstrates that, by assuming Gaussian distribution of the data uncertainties, the maximum of the probability density function of the model is achieved by minimizing the least-squares functional of data residual. Based on this probabilistic study, the objective function of FWI is formulated, aiming at minimizing the differences between simulated and observed data with the ℓ_2 norm. The objective function $J(m)$ in the time domain is therefore written as [Tarantola, 1984]:

$$J(m) = \frac{1}{2} \sum_{\mathbf{s},\mathbf{r}} \int_0^T dt \left[d^{cal}(m)(\mathbf{s},\mathbf{r},t) - d^{obs}(\mathbf{s},\mathbf{r},t) \right]^2, \tag{2.1}$$

where m is the model parameter, $\mathbf{s} = (s_x, s_z)$ and $\mathbf{r} = (r_x, r_z)$ are the shot and receiver positions, T is the recorded time length, t is the time variable, d^{cal} are the simulated data, and d^{obs} are the observed data, both recorded at the surface. The least-squares differences are summed over all the sources $\mathbf{s}$ and all the receivers $\mathbf{r}$.

To simplify, equation (2.1) can be re-written in vectorial notation as:

$$J(m) = \frac{1}{2} \Delta d^t \cdot \Delta d, \tag{2.2}$$

with $\Delta d = d^{cal} - d^{obs}$ denoting the data residual, and $(.)^t$ denoting the transpose operator.

In the time domain, the seismic data are represented by the seismograms of dimension n_t x n_r, where n_t is the number of recorded time samples and n_r is the number of seismic traces within a shot gather. Figure 2-1 displays an example of a seismic shot gather.

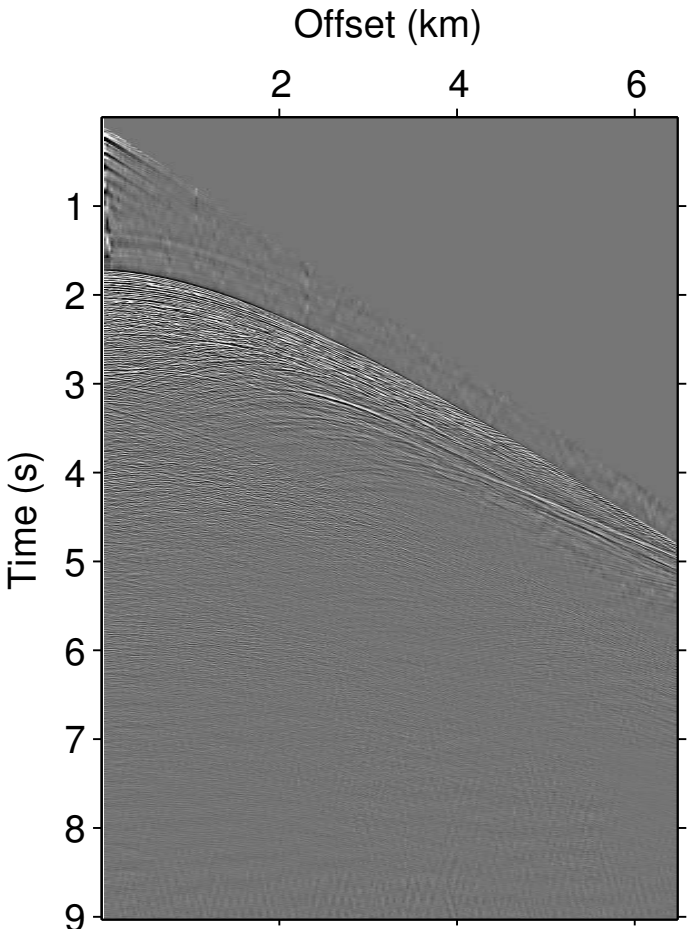


Figure 2-1: An example of seismic data. We can observe the direct wave and the reflected waves.

The objective of the inversion is to minimize the functional $J(m)$ in equation (2.2) by modifying the model m. In order to ease the extraction of information from the data and to improve the algorithm convergence, a weight W is usually applied to the data. The weighted least-squares objective function is defined as:

$$J(m) = \frac{1}{2}\Delta d^t W^t W \Delta d. \tag{2.3}$$

Commonly used weights W are data selection functions that, for example, give emphasis to long/short offsets, select events around the first arrival, etc...

2.3 Forward problem

2.3.1 Wave equation

The forward problem in FWI consists of modeling the seismic data propagation in a given model m. The simulated data d^{cal} is obtained by solving the wave equation. The physical law relating the model parameters and the data is the visco-elastic wave equation. For the scope of this thesis, we limit ourselves to the simplest case, by considering the constant density acoustic approximation of the wave propagation, which reads:

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} - \Delta u = S, \tag{2.4}$$

where c is the velocity, u is the wavefield, and S is the source term. u at the surface provides d^{cal}. In the acoustic approximation, u is generally the pressure. We discuss in Chapter 4 the effect of variable density.

2.3.2 Numerical resolution

The acoustic wave equation can be resolved in the time domain [Tarantola, 1984, Mora, 1987, Mora, 1989] or in the frequency domain [Pratt and Worthington, 1990, Sirgue and Pratt, 2004]. It can be resolved analytically or numerically. In the time domain, the numerical resolution of the partial differential wave equation could be achieved by numerical methods, such as finite difference [Virieux, 1986, Crase et al., 1990, Moczo et al., 2004], finite element [Marfurt, 1984, Choi et al., 2008], discontinuous Galerkin [Cockburn et al., 2000, Dumbser and Käser, 2006]. [Virieux et al., 2009] and [Plessix, 2007, Plessix, 2009] present some analyses of the complexity of different numerical modeling methods.

The finite-difference method is a natural way to solve partial differential equations. It is not based on high frequency assumption, thus it is more general than ray-based methods [Cerveny et al., 1977], and it provides a relatively efficient implementation compared to finite-element approaches. The method directly estimates each differential term using the Taylor expansion on a regular grid. Suppose that the 2D model dimension is m_{model} x n_{model}, the spatial samplings are Δx and Δz, the recording time is t, the time step is dt and the number of time step is n_t, then the 2^{nd} order Taylor expansion of a wavefield u at the n^{th} time step provides the estimation of the 2^{nd} order temporal derivative of u as follows:

$$\frac{\partial^2 u(x_{ij}, t^n)}{\partial t^2} \simeq \frac{u_{ij}^{n+1} - 2u_{ij}^n + u_{ij}^{n-1}}{\Delta t^2}, \tag{2.5}$$

where i, j are the spatial coordinates, and t^n is the temporal coordinate. Similarly, we get the 2^{nd} order spatial derivative of u:

$$\Delta u(x_{ij}, t^n) \simeq \frac{(u_{i+1j}^n + u_{ij+1}^n) - 4u_{ij}^n + (u_{i-1j}^n + u_{ij-1}^n)}{\Delta h^2}, \tag{2.6}$$

by setting $\Delta h = \Delta x = \Delta z$.

Inserting equation (2.6) and equation (2.5) into equation (2.4) gives the explicit resolution of u at time step $n+1$ and spatial position (i, j) using values of u at previous iterations.

$$u_{ij}^{n+1} = 2u_{ij}^n - u_{ij}^{n-1} + \frac{c^2 \Delta t^2}{\Delta h^2}(u_{i+1j}^n + u_{ij+1}^n + u_{i-1j}^n + u_{ij-1}^n - 4u_{ij}^n) + c^2 \Delta t^2 S. \tag{2.7}$$

In this work we use the scheme with 4^{th} order in time and 8^{th} order in space for the wave propagation modeling. It is based on the same principle as the above formulations. This wave propagation modeling operator is auto-adjoint as explained later.

In this conventional finite-difference implementation, 5 coefficients (including the central position and the 4 neighboring positions on the standard Cartesian coordinate system) are considered for 2D modeling and 7 coefficients (including the central position and the 6 neighboring positions on the standard Cartesian coordinate system) are considered for 3D modeling. [Operto et al., 2007] propose to use not only the neighboring positions on the standard Cartesian coordinate system, but also those on the 45° rotated Cartesian coordinate system. For this implementation, 9 coefficients are considered for 2D modeling and 27 coefficients are considered for 3D modeling.

[Virieux, 1986] first introduced the finite-difference scheme on a staggered grid for modeling the seismic rupture problem. Note that for constant density, velocity and pressure are defined at the same position while for variable density, the density is known at half-integer positions, and the velocity is known at integer positions. This staggered grid scheme resolves the problems with instabilities in models with high velocity contrasts and with grid dispersion in media with high Poisson ratio, at least for the conventional grid finite difference scheme.

The spatial and temporal sampling should be chosen carefully to avoid numerical dispersion and to ensure the stability of the solution. The spatial sampling Δh is defined such that

$$\Delta h \leq \frac{\lambda_{min}}{n_\lambda} = \frac{V_{min}}{f_{max}} \frac{1}{n_\lambda}, \tag{2.8}$$

where λ_{min} corresponds to the smallest wavelength, which depends on the maximum frequency f_{max} and the minimum velocity V_{min}. n_λ denotes the number of grid points per wavelength. The number of points per wavelength depends on the spatial and temporal order in the Taylor expansion, usually 5.5 points per wavelength is chosen for order 8 in space and order 4 in time.

The temporal sampling Δt is defined by the Courant-Friedrichs-Lewy (CFL) condition

to ensure the stability of the forward propagation:

$$\Delta t V_{max} \sqrt{\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta z)^2}} \leq 1, \tag{2.9}$$

where V_{max} is the maximum velocity, and Δx and Δz are the spatial discretizations.

The spatial discretization of the finite difference method is achieved on a regular grid, making it simple to implement. Nevertheless, the utilization of regular grid also gives rise to certain limitations as the spatial sampling is fixed and not adapted to local physical properties, thus the method is not as flexible as numerical methods using irregular grids, such as finite elements and discontinuous Galerkin. [Marfurt, 1984] compares the accuracy of finite-difference and finite-element for wave propagation modeling and observes that finite-element scheme better models curvi-linear interfaces and thin beds than finite-difference scheme. [Lombard and Piraux, 2004] propose the extension of the finite-difference method with floating interfaces on regular grids. However, for the scope of this work, we consider the classical finite-difference implementation.

2.3.3 Boundary conditions

Equation (2.4) defines the wave propagation in an infinite domain. However, because of the limited acquisition extension and of the limited computational capacity, the models we consider are usually limited in a finite domain. If we do not define any boundary condition at the borders of the model, waves will be artificially reflected from the borders of the model. To attenuate these parasite waves, spatial boundary conditions must be applied.

Absorbing boundary conditions are widely used to simulate an infinite medium. Ideally, absorbing boundary conditions should let pass the wavefields coming out of the model but prevent wavefields coming back into the model. Some types of absorbing boundary conditions rely on the modification of local attenuation properties by adding additional layers around the model [Cerjan et al., 1985]. The additional layers work as sponges. However, these sponges layers do not always work perfectly. Waves impinging with normal incidence are optimally absorbed by the sponges, however parasite reflections would appear for other incident angles [Brossier et al., 2009].

[Berenger, 1994] introduced a highly effective absorbing boundary condition based on the Perfectly Matched Layer (PML) for electromagnetics. In the continuous limit, it is proven that a PML interface between a regular medium and the fictitious perfectly matched medium completely absorbs incident waves from the regular medium, regardless of their incidence angle.

To be realistic, free surface boundary condition is also required, which describes the free surface of the Earth, thus simulating reflections from the Earth-air interface. However, in this thesis, I do not define the free surface, and I consider absorbing boundary layers all around the model.

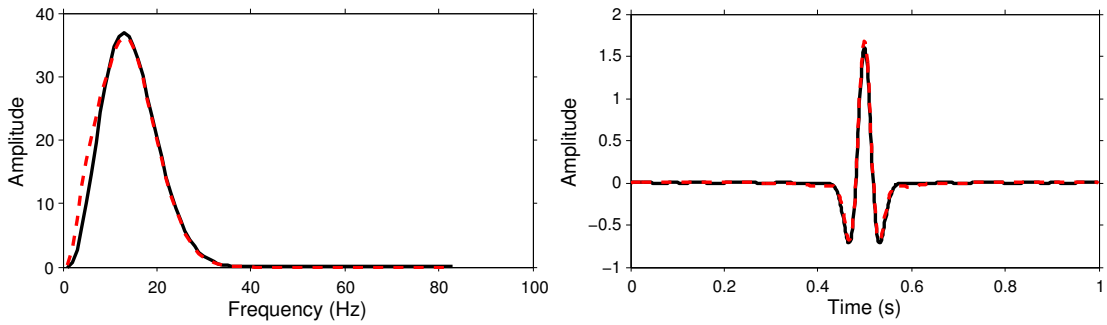


Figure 2-2: The original Ricker wavelet (black) and the estimated source (red) in the frequency domain (left) and in the time domain (right).

2.4 Source signature estimation

For a real data inversion problem, the source signature is usually unknown and has to be estimated along with the model parameters. [Tarantola, 1984] estimates the source using the adjoint-state method along with the other model parameters. The source wavelet is resolved iteratively using the gradient and the Hessian of the objective function with respect to the source wavelet.

[Pratt, 1999] propose a source signature estimation method for each discrete frequency during the iterative inversion procedure, by resolving a linear inverse problem as the relationship between the wavefield and the source is linear (refer to equation (2.4)). The solution of the linear inverse problem is given by:

$$S = \frac{u^t d^*}{u^t u^*}, \tag{2.10}$$

where $u = d^{cal}$ is the simulated wavefield in the actual velocity model and $d = d^{obs}$ is the observed wavefield. The source signature S could be estimated during the inversion procedure once the forward wavefield is modeled. In principle, the source should be estimated at each iteration of the inversion process. For the synthetic examples shown in Chapter 4, the source is defined as a given Ricker wavelet, while for the real data test in Chapter 5, the source wavelet is estimated according to equation (2.10).

We have validated the implementation of the method described in equation (2.10) on a 2D synthetic example. A Ricker wavelet with central frequency of 12 Hz is used for modeling the observed data. The initial guess of the source wavelet is a Dirac wavelet with frequency bandwidth between 0 Hz and 40 Hz. For each frequency between 0 Hz and 40 Hz, we estimate the source wavelet using equation (2.10). Figure 2-2 displays the initial Ricker wavelet and the estimated wavelet in the frequency domain and in the time domain. We note that the reconstructed source wavelet is very close to the original one. The velocity is correct at the surface in this case.

Some authors [Lee and Kim, 2003, Zhou and Greenhalgh, 2003] propose to normalize the wavefield to make the inversion independent of the source function, such that the source estimation step is not needed for these methods.

2.5 Linearization of the inverse problem

FWI is considered as a non-linear optimization problem. Usually a local optimization method is used. If we assume that the data is locally approximately linear with the model and that the initial model m_0 is close to the true one, then the objective function is locally approximately quadratic in the neighbourhood of the starting model m_0, and we search for the minimum of the objective function in the neighbourhood of m_0, with

$$m = m_0 + \Delta m. \tag{2.11}$$

Suppose the model perturbation vector Δm is relatively small compared to model vector m, then the second-order Taylor expansion gives:

$$J(m) = J(m_0+\Delta m) = J(m_0)+\nabla_m J(m_0)\Delta m+\frac{1}{2}\Delta m^t \frac{\partial(\nabla_m J)}{\partial m}(m_0)\Delta m+O(\Delta m^3). \tag{2.12}$$

We denote the Hessian matrix as

$$H = \frac{\partial(\nabla_m J)}{\partial m}, \tag{2.13}$$

and we express the derivative of J with respect to m using equation (2.12) with the locally linear assumption:

$$\nabla_m J(m) = \nabla_m J(m_0) + H(m_0)\Delta m. \tag{2.14}$$

The minimum of the objective function is reached at m if the gradient at this location is zero. By setting $\nabla_m J(m)$ to zero we have:

$$\Delta m = -H^{-1}\nabla_m J(m_0). \tag{2.15}$$

The first term H^{-1} is the pseudo-inverse of the Hessian matrix. The second derivative term is the gradient vector of the objective function. By substituting equation (2.15) into equation (2.11), we get the Newton solution that allows updating the initial model [Tarantola, 1984, Virieux and Operto, 2009].

If the objective function is quadratic, leading to the linear inverse problem case, the convergence of the problem could be achieved in one iteration using equation (2.15). If the objective function is not exactly quadratic, the Newton method will not converge in one iteration, and the problem must be resolved iteratively to compensate the problem non-linearity. In this case, the model update is given by:

$$m^{k+1} = m^k - H^{-1}(m^k)\nabla_m J(m^k). \tag{2.16}$$

2.6 Gradient and Hessian

2.6.1 Gradient

Equation (2.16) allows the calculation of model perturbation at each iteration after the computation of the gradient and the Hessian matrix. The gradient is a vector with the same dimension as the model. It relates the data perturbation with the model perturbation. It

is the first-order derivative of the objective function J with respect to the model parameter m.

$$G(m) = \nabla_m J(m) = \left[\frac{\partial d^{cal}(m)}{\partial m)}\right]^t (d^{cal}(m) - d^{obs}) = F(m)^t \Delta d, \tag{2.17}$$

where F is the Fréchet derivative matrix. If we consider the linear direct problem, F represents the Born modeling operator, with $\Delta d = F\Delta m$ [Tarantola, 2005].

As shown in equation (2.15), the resolution of local optimization problems is based on the gradient. In fact, the opposite direction of the gradient gives the largest slope of the objective function. [Pratt et al., 1998] gives a physical interpretation of the gradient vector by illustrating the partial derivative wavefield F using a point diffraction model. They conclude that the partial derivative wavefield represents forward propagation using virtual sources at the diffracting locations. The gradient is the correlation of the partial derivative wavefield and the residual vector. This correlation tests the similarity between these two wavefields and allows getting the diffracting points which give rise to the data residuals. These diffracting points are initially not present in the model.

[Sirgue, 2003] gives an interpretation of the gradient vector by performing the Singular Value Decomposition (SVD) on the Fréchet derivative matrix. He expresses the gradient vector in the model eigenvector basis and concludes that the gradient vector is the sum of the projected components of true model perturbation along the model eigenvectors, multiplied by the model eigenvalues. Thus, the gradient direction is the direction of the true model perturbation in a stretched model space.

Instead of calculating the Fréchet derivative F, which is not trivial, the gradient can be efficiently calculated with the adjoint-state method [Plessix, 2006], by using the augmented functional and Lagrange multipliers. With the adjoint-state method, and with the fact that the modeling operator is auto-adjoint, the computation of the gradient of the objective function J is simplified and is given by:

$$G = \nabla_m J(m) = -\frac{2}{m^3} \sum_{\mathbf{s}} \int_0^T dt Q(\mathbf{s}, t) \frac{\partial^2 P(\mathbf{s}, t)}{\partial t^2} = P \otimes Q. \tag{2.18}$$

From equation (2.18), we can observe that the construction of the gradient requires two wavefields, the forward propagated wavefield P, resolving the direct wave equation (2.4), and the back-propagated wavefield Q, satisfying the adjoint wave equation:

$$\frac{1}{m^2} \frac{\partial^2 Q}{\partial t^2} - \Delta Q = d^{cal} - d^{obs}, \tag{2.19}$$

with the data residual $d^{cal} - d^{obs}$ as the source term. This equation is back propagated with Q ($t_{max} = 0$). The gradient G is computed as the cross-correlation of P and Q. This formula is similar to the migration imaging principle of [Claerbout, 1971]. The gradient obtained after the first iteration of the local optimization looks like a migrated image obtained by Reverse-Time Migration (RTM).

2.6.2 Hessian and preconditioning

The Hessian is a matrix of dimensions n_{model} x n_{model}. It is the second-order derivative of the objective function J with respect to the model parameter m:

$$H(m) = \frac{\partial^2 J(m)}{\partial m^2} = F^t(m)F(m) + \frac{\partial F(m)}{\partial m}\Delta d. \tag{2.20}$$

From equation (2.15) we observe that the Hessian relates the gradient with the true model perturbation. The gradient only does not give a correct model update direction as the computed gradient may suffer from geometrical spreading and characteristics from the source and acquisition setup. The inverse of Hessian in equation (2.15) helps to correct the geometrical spreading, uneven illumination and deconvolves the gradient [Pratt et al., 1998, Brossier et al., 2009].

Referring to equation (2.20), we can explicitly separate the Hessian into two terms. Of the two terms, the first term is straightforward to compute if the Fréchet derivatives F are available, while the second term is much more difficult to compute. [Métivier et al., 2013] propose a matrix-free method to compute the full Hessian term using 2D adjoint-state method and inner linear optimization. If the full inverse of Hessian is computed, it yields the Full-Newton method. If only the first term is considered, it is the approximate Hessian and it yields the Gauss-Newton method [Pratt et al., 1998].

The diagonal terms of the approximate Hessian of the Gauss-Newton solution contain the zero-lag autocorrelations and therefore represent the square of the amplitude of the partial-derivative wavefield. Applying the inverse of these diagonal terms to the gradient partly compensates the geometric spreading effect. In the framework of surface seismic experiments, this scaling provides a good balance between shallow and deep perturbations. The off-diagonal terms of the approximate Hessian are computed by correlation between partial-derivative wavefields associated with different model parameters.

For 1D media, the approximate Hessian is a band-diagonal matrix, and the numerical bandwidth decreases as the frequency increases. For a finite range of frequencies, however, the Hessian is no longer diagonal and not even diagonally dominant [Chavent and Plessix, 1999, Pratt et al., 1998]. The off-diagonal elements of the approximate Hessian account for the limited-bandwidth effects related to the source term. Applying its inverse to the gradient can be interpreted as a deconvolution of the gradient from these limited-bandwidth effects. This deconvolution of gradient enables to focus and sharpen the FWI images [Pratt et al., 1998].

The explicit resolution of the inverse of the approximate Hessian matrix is difficult, as the matrix size is too large to be directly used in practical applications. Sometimes, it is even impossible, as the Hessian may be not positive and requires regularization. Some authors propose to consider only the diagonal terms of the approximated Hessian [Shin et al., 2001, Chavent and Plessix, 1999]. But as the gradient of the least-squares functional multiplied by a diagonal matrix cannot perfectly deconvolve the band-limited effect, iterations are needed to gradually construct the inverse of the approximate Hessian [Lambaré et al., 1992, Nemeth et al., 1999]. Some approximations of the diagonal of the approximate Hessian are also proposed to construct amplitude-preserving migration images [Shin et al., 2001, Plessix and Mulder, 2004].

[Virieux and Operto, 2009] illustrates the scaling effect performed by the diagonal

elements of the approximate Hessian and the deconvolution effect performed by the off-diagonal elements of the approximate Hessian. They show that the inversion image is sharper when using the approximate Hessian instead of setting it as a scalar.

It is also possible to precondition the gradient without using the Hessian, but by exploiting some a priori information. For example, weighting the data residuals in time or in depth linearly is an appropriate way to compensate the geometrical spreading.

2.7 Resolution analysis

2.7.1 Gradient formulation

The gradient of the objective function of FWI is calculated as the cross-correlation of the forward wavefield P and back-propagated wavefield Q. The gradient formulation (equation (2.18)) can also be written in the frequency domain using Green's function as [Sirgue and Pratt, 2004]:

$$G = -\omega^2 \sum_{\mathbf{s}} \sum_{\mathbf{r}} Re(G_0^*(\mathbf{x},\mathbf{s}) G_0^*(\mathbf{x},\mathbf{r}) \Delta d(\mathbf{r},\mathbf{s})), \tag{2.21}$$

where $G_0(\mathbf{x},\mathbf{s})$ and $G_0(\mathbf{x},\mathbf{r})$ are the Green's functions for the source and receiver wavefield, and Δd is the data residual at the receiver positions.

As the velocity update of FWI is based on the gradient, the resolution analysis of the gradient is necessary to study the influence of different parameters on the spectrum of the gradient, and to find the reason why FWI fails to recover the long wavelengths of the model.

2.7.2 Resolution analysis of different waves

[Huang and Schuster, 2014] derive formulas of resolution limits for different types of waves, including diving waves, primary reflections, multiple reflections, and diffractions. The resolution analysis was initially realized to study seismic migration. However we can benefit from it for the spectral analysis of waveform inversion. Here, I give a summary of the analysis and the conclusions. Readers could refer to [Huang and Schuster, 2014] for a detailed derivation.

Resolution analysis of a scattering point

In the framework of the inverse-scattering and diffraction tomography [Devaney, 1982, Miller et al., 1987], the spatial frequency vector $\mathbf{k}$ sampled in the image at a diffraction point is related to the wavelength λ and the half-scattering angle θ. We denote the directions of the the source and the receiver plane wave by $\mathbf{d_s}$ and $\mathbf{d_r}$. With the assumption of the far field, and by replacing the Green's function with the plane-wave approximation, for a single frequency we have:

$$\begin{aligned} G_0(\mathbf{x},\mathbf{s}) &\approx exp(ik_0 \mathbf{d_s} \cdot \mathbf{x}), \\ G_0(\mathbf{x},\mathbf{r}) &\approx exp(ik_0 \mathbf{d_r} \cdot \mathbf{x}). \end{aligned} \tag{2.22}$$

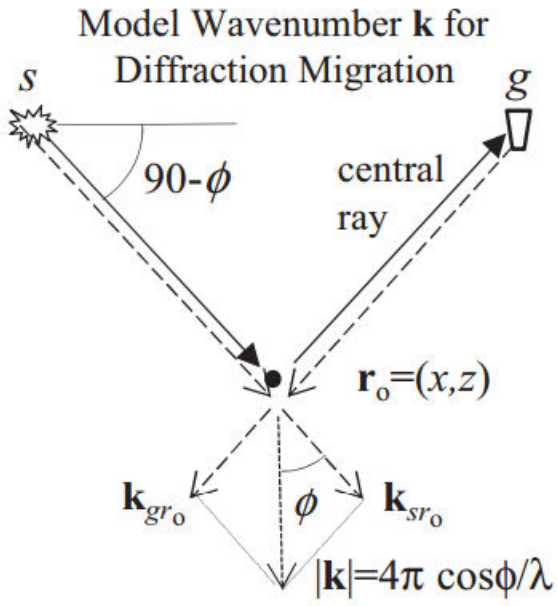


Figure 2-3: Wavenumber illumination. One source-receiver pair, one scattering point and one frequency in the data provides one wavenumber in the image space [Huang and Schuster, 2014].

Inserting equation (2.22) into equation (2.21), yields

$$G = -\omega^2 \sum_{\mathbf{s}} \sum_{\mathbf{r}} Re(exp(-ik_0(\mathbf{d_s} + \mathbf{d_r}))\Delta d(\mathbf{r}, \mathbf{s}), \tag{2.23}$$

The coverage of the wavenumber $\mathbf{k}$ of the gradient at the scattering point (Figure 2-3) is defined according to:

$$\mathbf{k} = k_0(\mathbf{d_s} + \mathbf{d_r}) = \frac{2\omega}{c}\cos\theta\mathbf{n} = 4\pi\frac{\cos\theta}{\lambda}\mathbf{n}, \tag{2.24}$$

where c is the velocity and ω is the angular frequency. According to equation (2.24), the resolution is optimal for zero-offset data, and is equal to $\frac{\lambda}{2}$. For FWI, low frequency (small ω) and long offset (large θ) data provide long-wavelength velocity update, while high frequency (large ω) and short offset (small θ) data provide short-wavelength velocity update. The resolution analysis of other types of waves is also based on this single-scattering resolution analysis.

Resolution analysis of primary reflections

For primary reflections, in the far-field approximation, the horizontal resolution limit Δx and the vertical resolution limit Δz (Figure 2-4e) are defined according to:

$$\begin{cases} \Delta x = \frac{Z\lambda}{4X}, \\ \Delta z = \frac{\lambda}{4}, \end{cases} \tag{2.25}$$

where Z is the reflector depth, X is the source-receiver offset, and λ is the wavelength.

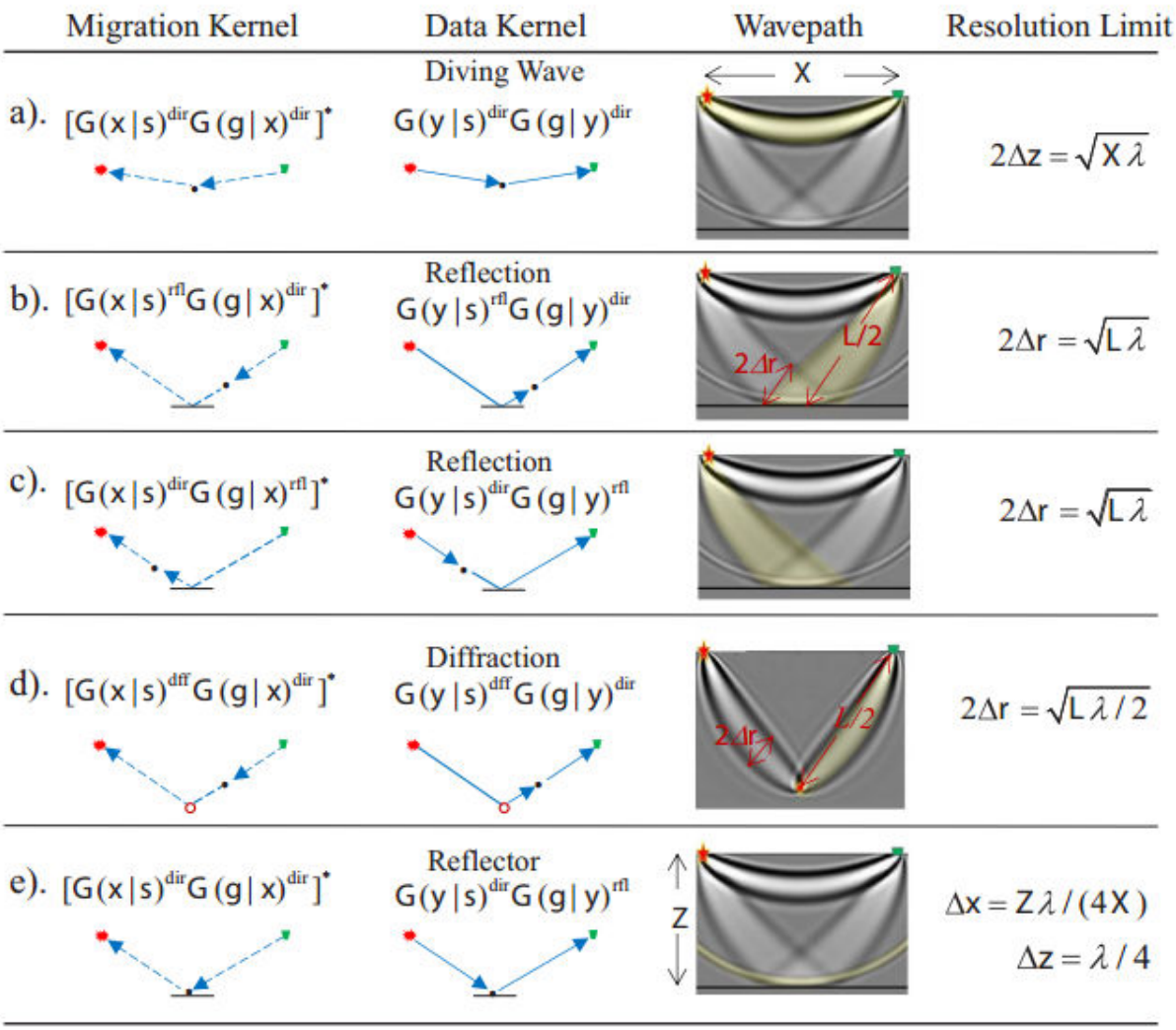


Figure 2-4: Approximate resolution limits for (a) diving waves, (b) and (c) reflection-related transmission, (d) diffraction-related transmission and (e) reflection migration [Huang and Schuster, 2014].

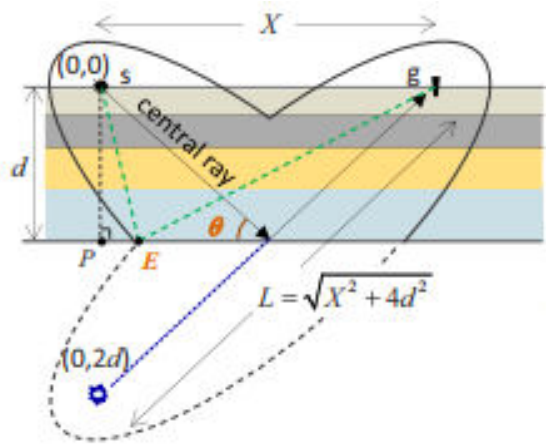


Figure 2-5: First Fresnel zone for the specular reflection [Huang and Schuster, 2014].

Resolution analysis of diving waves

For diving waves, the vertical limitation Δz is equal to half width of the first Fresnel zone at the point middle-way between the source and the receiver (Figure 2-4a). It depends on the travel length X and the wavelength λ:

$$2\Delta z = \sqrt{X\lambda}, \tag{2.26}$$

Resolution analysis of reflection-related transmission

The reflection-related transmission provides the rabbit-ear wavepaths shown in Figure 2-4b and 2-4c. The resolution analysis is similar to that of the diving waves. It is calculated using the mirror-reflection scheme (Figure 2-5) as if the receiver is located at the mirror position (0, 2d) in a homogeneous model. The resolution limit $2\Delta r$ perpendicular to the ray at the midpoint is equal to the width of the first Fresnel zone, given by:

$$2\Delta r = \sqrt{L\lambda}, \tag{2.27}$$

where L is the travel length and λ is the wavelength.

Resolution analysis of diffraction-related transmission

The diffraction-related transmission is similar to the reflection-based transmission, except that the diffraction point is considered as the virtual receiver and the secondary source. The propagation distance is halved compared to the reflection transmission, leading to a narrower wavepath with the same travel length L and wavelength λ. The resolution limit is given by (Figure 2-4d):

$$2\Delta r = \sqrt{L\lambda/2}. \tag{2.28}$$

Resolution analysis of multiple-related transmission

The multiple-related transmission simulates the repeated wavepath between the source and the reflector and between internal reflectors. The resolution limit is computed according

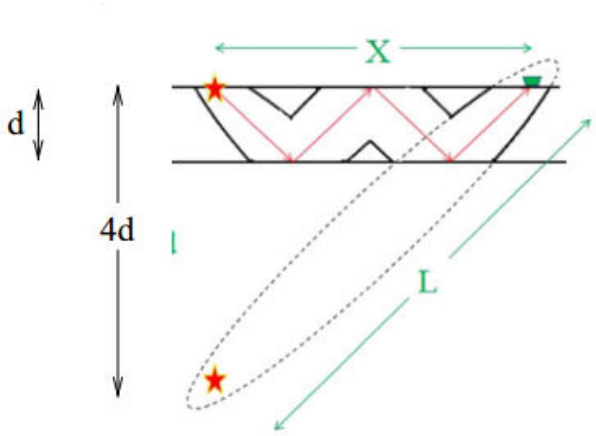


Figure 2-6: Wavepath for the first-order free-surface multiple [Huang and Schuster, 2014].

to the mirror-reflection scheme (Figure 2-6 and 2-7):

$$2\Delta r = \sqrt{L\lambda}, \tag{2.29}$$

where L is the travel length and λ is the wavelength.

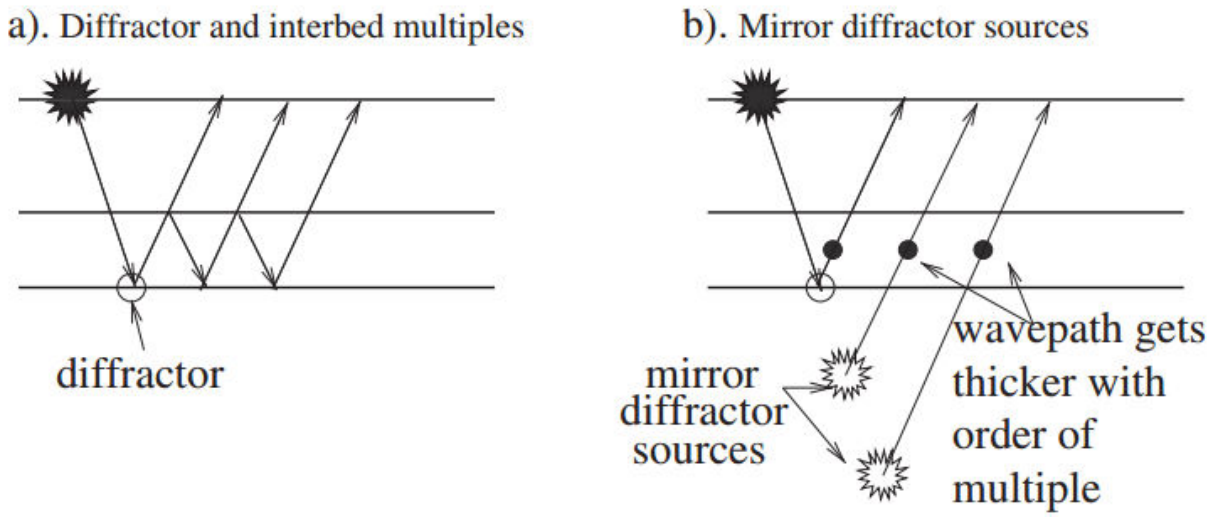


Figure 2-7: a) Ray diagram for interbed multiples generated by a diffractor in a thin layer and b) the associated mirror sources diagram [Huang and Schuster, 2014].

From the previous analysis, we conclude that the resolution limit of different waves mainly depends on the scattering angle, the wavelength, and the wave propagation distance. These parameters are function of the seismic acquisition setup. For example, long offsets are favorable for recording wide-angle and long-propagation distance data.

2.7.3 Resolution analysis and acquisition setup

In this section, we show the influence of the source bandwidth and the data offset range on the wavenumber components of the FWI gradient through a 1D layer model example.

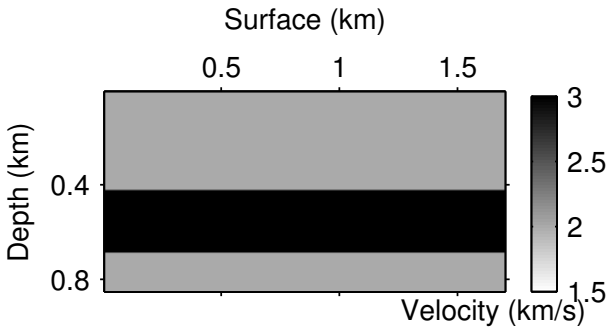


Figure 2-8: 1D exact velocity model.

The true model is shown in Figure 2-8. The model dimensions are 1.7 km by 0.85 km. The model consists of a strong velocity contrast, and the velocities for the top layer, middle layer and bottom layer are 2 km/s, 3 km/s and 2 km/s, respectively. The initial model is a homogeneous model at 2 km/s. Sources and receivers are evenly distributed at the surface, with the source spacing being 17 m and receiver spacing being 8.5 m.

Resolution analysis and source bandwidth

We perform conventional FWI in the time domain for the 1D layer model. The initial model is homogeneous at 2 km/s. We analyze four FWI tests with different source wavelets. These four source wavelets are all Ricker wavelets, but the central frequencies for the four tests are respectively 4 Hz, 8 Hz, 12 Hz and 16 Hz (Figure 2-9). Here, low frequencies are present in these source wavelets. The inversion is performed in the time domain with all frequencies mixed together. Figure 2-10 shows the gradients calculated at the first iteration of FWI for each test and the wavenumbers of the vertical slice taken in the middle of each gradient model. We observe that the spectrum of the gradient is closely related to the bandwidth of the source wavelet and the frequency bandwidth in the gradient increases with the increase of the frequency bandwidth in the source wavelet. This result corresponds to equation 2.24 and to the conclusion that for the same acquisition geometry, lower frequency data provide longer-wavelength velocity update, while higher frequency data provide shorter-wavelength velocity update. In order to retrieve the long wavelengths of the model, low-frequency data are favorable [Sirgue, 2003]. And when low frequencies are missing from the data, which is often the case in real data case, the FWI usually fails to retrieve the long-wavelength part of the model [Claerbout, 1985, Jannane et al., 1989].

Resolution analysis and offset

[Mora, 1989] and [Pratt et al., 1996] show that FWI acts in two modes, a migration mode, related to the short wavelengths of the model which could be obtained by iterative migration, as well as a tomography mode, related to the long wavelengths of the model which could be retrieved with the benefit of long-offset data.

We compute the gradient of FWI using the 1D layer model of Figure 2-8, and using data with different offset ranges. The central frequency of the Ricket wavelet is 12 Hz and the initial model is homogeneous. The offset ranges are respectively [0–0.17] km,

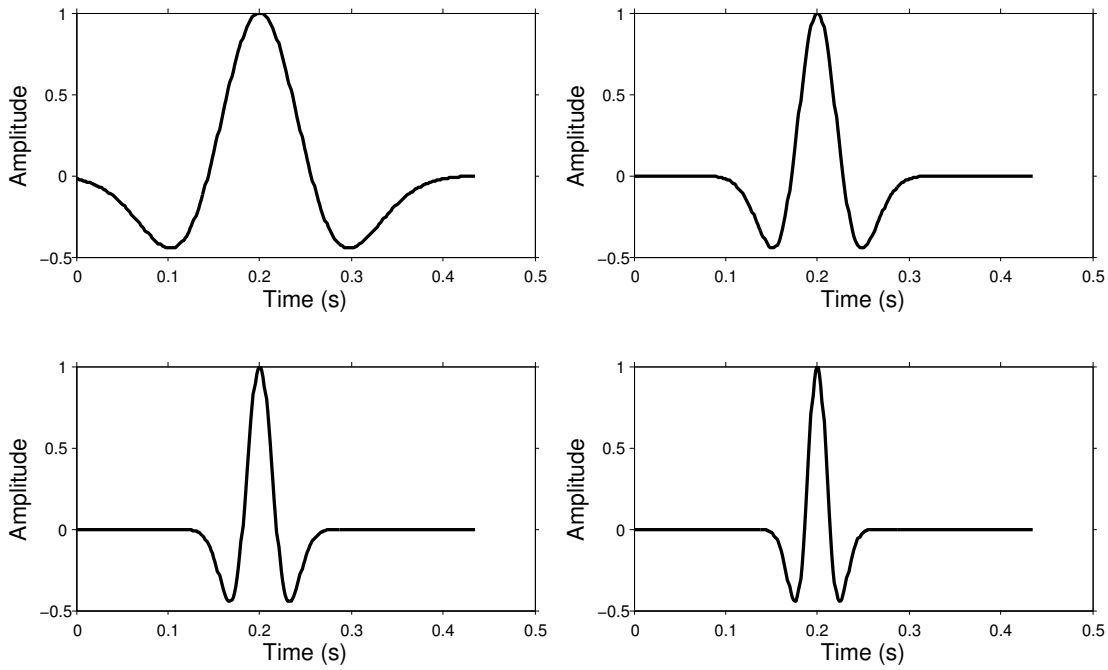


Figure 2-9: Four Ricker wavelets. From top to bottom and from left to right, the central frequency is respectively 4 Hz, 8 Hz, 12 Hz, and 16 Hz.

[0.34–0.51] km, [0.68–0.85] km. The gradients at the first iteration and the wavenumbers of the vertical slice taken in the middle of each gradient are shown in Figure 2-11. As expected, the long-offset data provides longer wavelengths than short-offset data.

From the analysis in sections 2.7.2 and 2.7.3, we conclude that in order to retrieve the long-wavelength components of the model, low frequency and long-offset data are important [Sirgue, 2006]. Based on this idea, offset-windowing [Mora, 1989, Pratt et al., 1996] strategies are proposed for FWI. Moreover, reflection-related transmission [Xu et al., 2012, Tang and Lee, 2013, Wang et al., 2013, Brossier et al., 2015], and multiple-related transmission [Staal et al., 2012] can also be taken into consideration for the reconstruction of the long-wavelength model.

2.8 Velocity model update methods

Supposing that the initial model m_0 is situated in the basin of attraction of the global minimum of the objective function, the local optimization methods should be able to converge to this global minimum in a limited number of iterations. When added to the initial velocity, the velocity perturbations calculated using equation (2.11) lead to an updated velocity model, which is used as the starting model for the next iteration.

Depending on the methods used to calculate the Hessian matrix, the velocity model update methods could be divided into several categories.

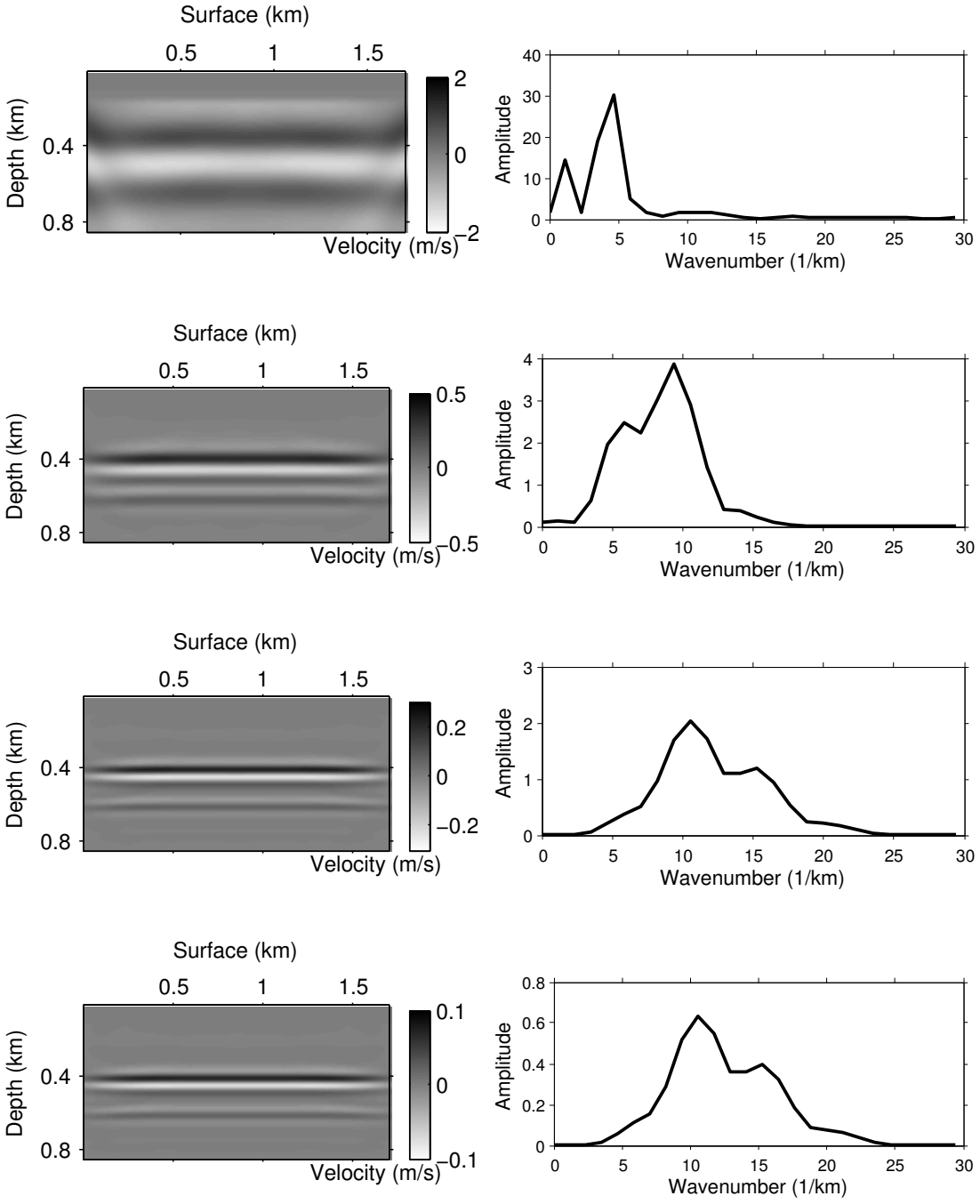


Figure 2-10: (Left) Gradient of the first iteration of FWI with source wavelets displayed in Figure 2-9. The central frequency is respectively 4 Hz, 8 Hz, 12 Hz, and 16 Hz. (Right) The bandwidth of the vertical slice taken at the middle of each model on the left.

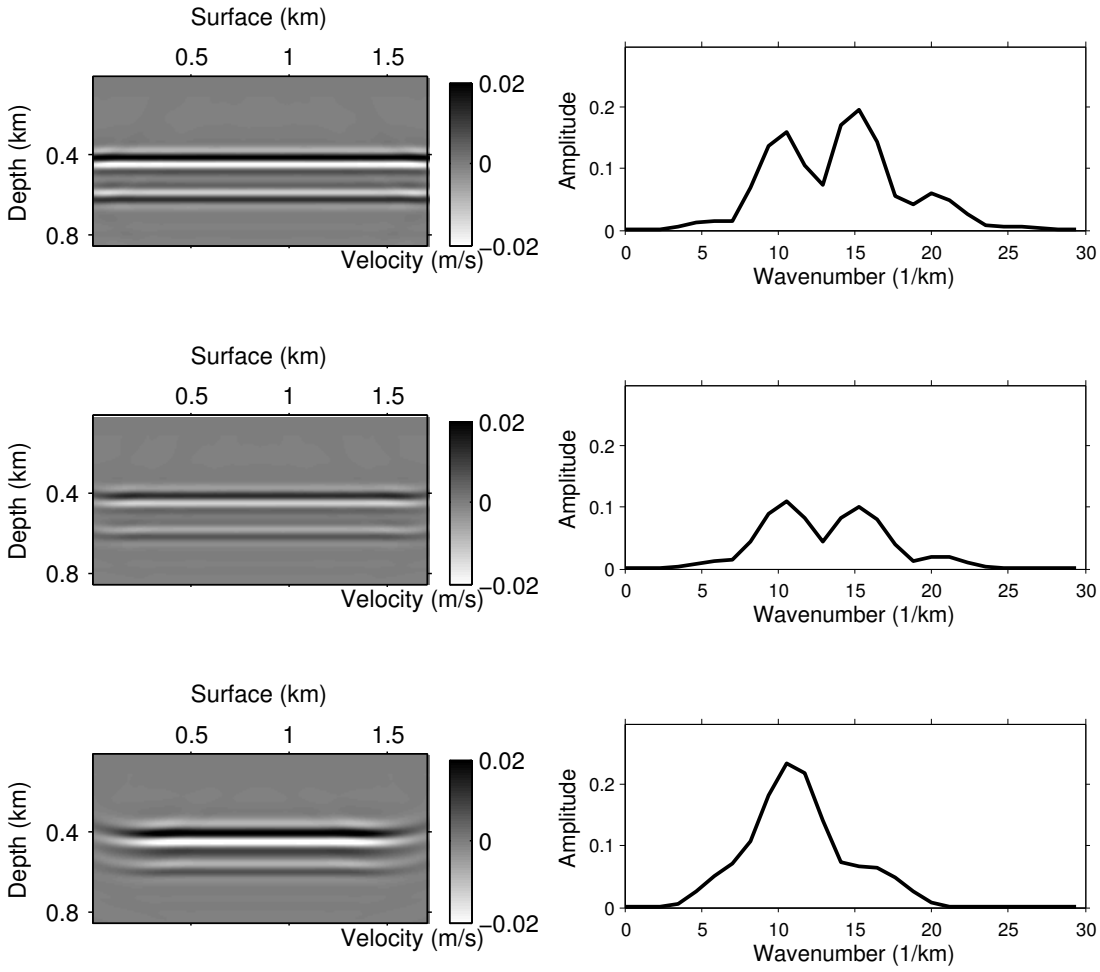


Figure 2-11: (Left) Gradient of the first iteration of FWI for data with offset range [0–0.17] km, [0.34–0.51] km, [0.68–0.85] km respectively. (Right) The bandwidth of the vertical slice taken at the middle of each model on the left.

2.8.1 Steepest descent (or gradient descent) methods

The gradient descent method assumes that the Hessian is a scalar, giving a simple solution for the model perturbation vector:

$$\Delta m = -\alpha G. \tag{2.30}$$

This method does not require the estimation of the inverse of the Hessian and only the descent step length α needs to be determined. The role of the step length can also be considered as converting the unit of the gradient vector G to model unit [Pratt et al., 1998]. Generally, α is determined by line-search methods [Nocedal, 1980, Gauthier et al., 1986, Tarantola, 2005] by minimizing the objective function with respect to α. Alternatively, α can be estimated by quadratic or cubic interpolation [Vigh et al., 2009]. The idea is to determine a pair (α_1,α_2) such that

$$\begin{cases} J(m^k + \alpha_1 \Delta m) < J(m^k) \\ J(m^k + \alpha_2 \Delta m) > J(m^k + \alpha_1 \Delta m) \end{cases}. \tag{2.31}$$

The quadratic interpolation passing through these three points (m^k, $m^k + \alpha_1 \Delta m$, $m^k + \alpha_2 \Delta m$) provides the minimum of the parabola in the interval [0,α_2], giving the optimal descent step α_{opt}.

Knowing that the gradient represents the direction of the steepest descent of the objective function, we can always find an α that reduces the objective function in the opposite direction of the gradient.

2.8.2 Conjugate-gradient method

The convergence rate of the gradient descent method is generally quite slow as the successive directions estimated by the gradient are orthogonal if the step length is optimally chosen. The convergence rate can be improved by the conjugate-gradient method, which combines the current gradient direction with the previous gradient direction to accelerate the convergence rate [Mora, 1987, Crase et al., 1990, Hu et al., 2011].

The current gradient direction P^k uses the present gradient direction G^k and previous conjugate-gradient direction P^{k-1}:

$$P^k = G^k + \beta^k P^{k-1}. \tag{2.32}$$

There are various ways to determine β. One well-known choice of β is expressed as [Polak and Ribiere, 1969]

$$\beta^k = \frac{(G^k - G^{k-1})^t G^k}{(G^{k-1})^t G^{k-1}}. \tag{2.33}$$

2.8.3 Newton and Gauss-Newton methods

The Newton method computes the full Hessian, while the Gauss-Newton only considers the first term of the Hessian in equation 2.20 [Pratt et al., 1998, Virieux and Operto, 2009]. These two methods have a faster convergence rate than the previously described gradient-based methods but they require the calculation of extra direct problems for the

calculation of the inverse of the Hessian and are considered quite time-consuming for real data inverse problems. [Métivier et al., 2013] propose a matrix-free method to compute the full Newton term using 2D adjoint-state method and inner linear optimization.

2.8.4 Quasi-Newton method

The quasi-Newton method relies on the estimation of the Hessian, allowing to take advantage of the information contained in the Hessian without paying the price of a full Newton iteration. The BFGS (named after Broyden, Fletcher, Goldfarb and Shanno) algorithm [Nocedal, 1980] estimates the inverse of the Hessian from gradients and objective functions of previous iterations and increases the precision of the estimation along with iterations. The limited-memory version (L-BFGS) of the algorithm [Nocedal, 1980] is a variant of BFGS. It keeps in memory only a limited number of iterations, typically between 3 to 15 iterations, being less demanding in memory than BFGS. The L-BFGS algorithm will be used in the applications presented in Chapter 4 and Chapter 5 of this thesis.

[Pratt et al., 1998] review different optimization methods and compare their quality and convergence rate. [Pratt et al., 1998] show that the gradient-based methods are less robust than the Newton-based methods, and may fail to converge no matter how many iterations. [Métivier et al., 2012, Castellanos et al., 2015] compare the FWI results using Newton, Gauss-Newton and quasi-Newton methods. Figure 2-12 shows an example of the optimization scheme for different optimization methods.

2.9 Initial model and non-linearities

The success of the local optimization methods depends on the accuracy of the initial model. As the objective functions may present some local minima, the initial model should be located in the neighbourhood of the global minimum to ensure the success of convergence. If the phase shift between the calculated data and the observed data does not exceed half period of the signal, the optimization method will adjust these two data without phase ambiguity. Otherwise, the optimization method will adjust these two data with a phase shift of one or several periods (Figure 2-13), causing the convergence to a local minimum. This is the so-called cycle-skipping phenomenon.

The initial model for FWI can be obtained by traveltime tomography [Bishop et al., 1985, Pratt and Chapman, 1992, Billette and Lambaré, 1998, Woodward et al., 2008] or Migration Velocity Analysis (MVA) [Symes and Carazzone, 1991, Chauris and Noble, 2001, Sava and Biondi, 2004]. Traveltime tomography uses the traveltimes to estimate the wave propagation velocity. Generally, RMO events are picked after prestack depth migration [Guillaume et al., 2008]. Migration Velocity Analysis (MVA) is performed in the image domain and aims at building a relationship between migrated image perturbation and model perturbation. This approach has the advantage of not requiring picking.

In waveform inversion, strategies of using different data at different stages of inversion are proposed to mitigate the non-linearity of the problem. For the same acquisition geometry, lower frequency data provide longer-wavelength velocity update, while higher frequency data provide shorter-wavelength velocity update [Sirgue, 2003]. Based on this idea, [Bunks et al., 1995] perform the inversion in the time domain using scale decomposed

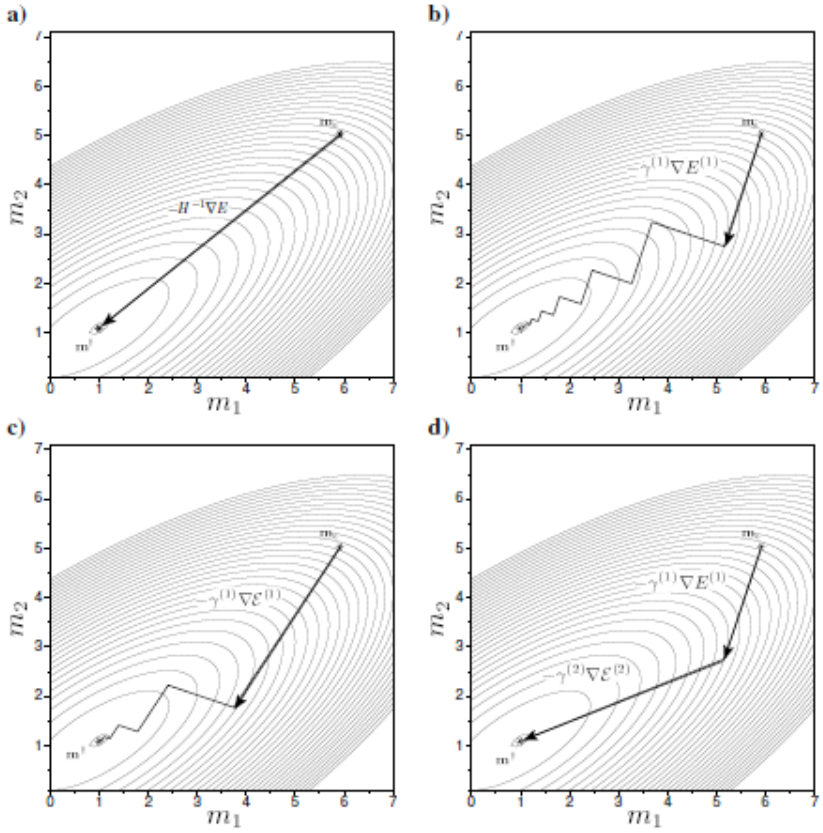


Figure 2-12: Comparison of the convergence rate of local methods applied on a 2D linear inverse problem. The problem is solved using a) Gauss-Newton, b) gradient descent, c) preconditioned gradient and d) conjugate-gradient methods. The gradient in c) was preconditioned with the diagonal term of the inverse Hessian [Sirgue, 2003].

source and data, starting from the lowest available frequencies and gradually increasing the frequency bandwidth. This point is also illustrated in [Mulder and Plessix, 2008] by showing that the basin of attraction of the objective function becomes narrower for increasing frequencies. Similarly to [Bunks et al., 1995], the multiscale inversion can be implemented in the frequency domain using increasing frequencies [Sirgue, 2003, Mulder and Plessix, 2008]. Guidelines to properly choose successive frequencies or frequency ranges are also provided [Sirgue and Pratt, 2004, Brossier et al., 2009]. The elastic inversion using the multiscale scheme is carried out by [Brossier et al., 2009]. Figure 2-14 to Figure 2-19 show an example that compares the result of conventional FWI and multiscale FWI through a 2D synthetic model. The true model (Figure 2-14) is a constant gradient model with a circular anomaly. The initial model (Figure 2-15) is just the constant gradient model. The circular velocity anomaly is missing. The source wavelet used for FWI is a Ricker wavelet with central frequency of 8 Hz. The result of FWI after 100 iterations is displayed in Figure 2-16, and the model perturbation with respect to the initial model is displayed in Figure 2-17. Compared to the exact model, only the edges of the circular anomaly are retrieved. FWI fail to recover the long-wavelength part inside the anomaly. For the multiscale FWI, we use 7 singular frequencies: 1.8 Hz, 2.5 Hz, 3.5 Hz, 5 Hz, 7 Hz, 10 Hz, 14 Hz.

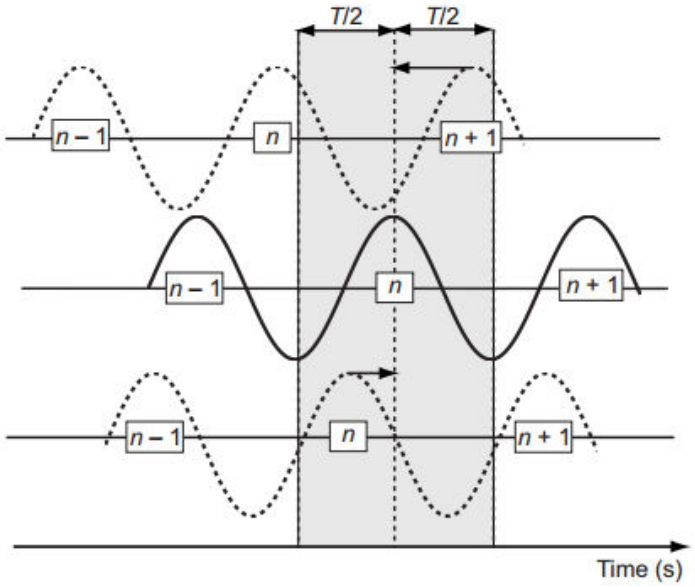


Figure 2-13: Principle of cycle-skipping problem [Virieux and Operto, 2009]. If the phase mismatch is less than one half of the period, the local optimization methods could adjust the correct phase. In the contrary case, the local optimization methods adjust these two signals with one period shift and the inversion falls into a local minimum. [Virieux and Operto, 2009]

For each frequency, we perform 25 iterations of FWI. The wavefield is initially computed in the time domain and then the chosen discrete frequency is extracted. Compared to the conventional FWI, the multiscale FWI (Figure 2-18 and Figure 2-19) better recovers the long wavelength of the anomaly.

Another strategy to avoid local minima is data-windowing [Sirgue, 2003, Brenders and Pratt, 2007]. It consists of inverting the early arrivals first as they provide lower wave numbers in the velocity gradient [Sears et al., 2008]. They mainly update the long-wavelength components of the shallow part of the model [Sirgue, 2003]. Time windowing of early arrivals is achieved by damping the amplitudes of later arrivals in the data residuals.

The strategy of selecting data with different offsets is also proposed. [Mora, 1989] and [Pratt et al., 1996] observe that the short wavelengths of the model could be obtained by iterative migration, and the long wavelength of the model could be retrieved with

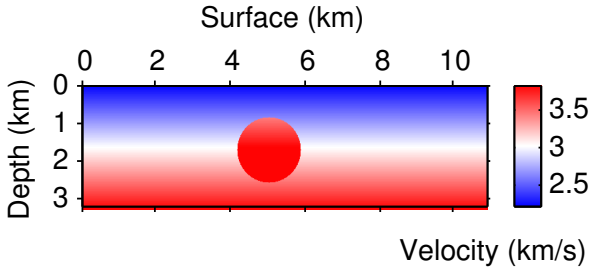


Figure 2-14: Exact velocity model.

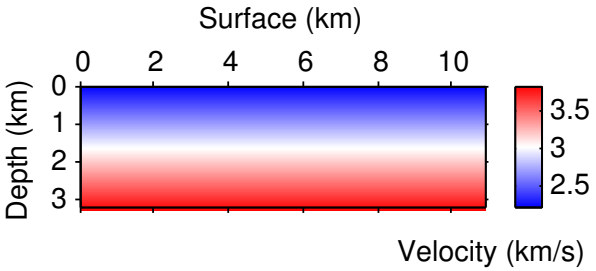


Figure 2-15: Initial velocity model.

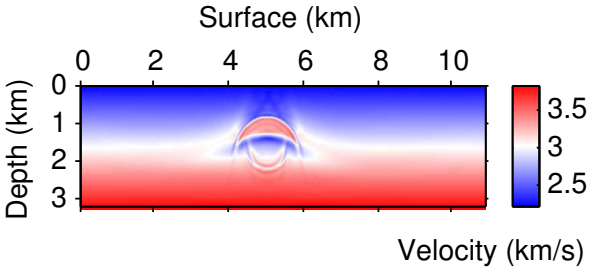


Figure 2-16: Conventional FWI result.

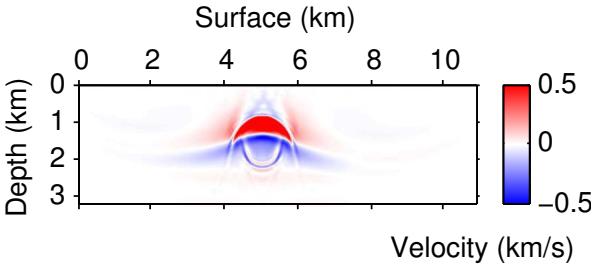


Figure 2-17: Velocity perturbation obtained with conventional FWI.

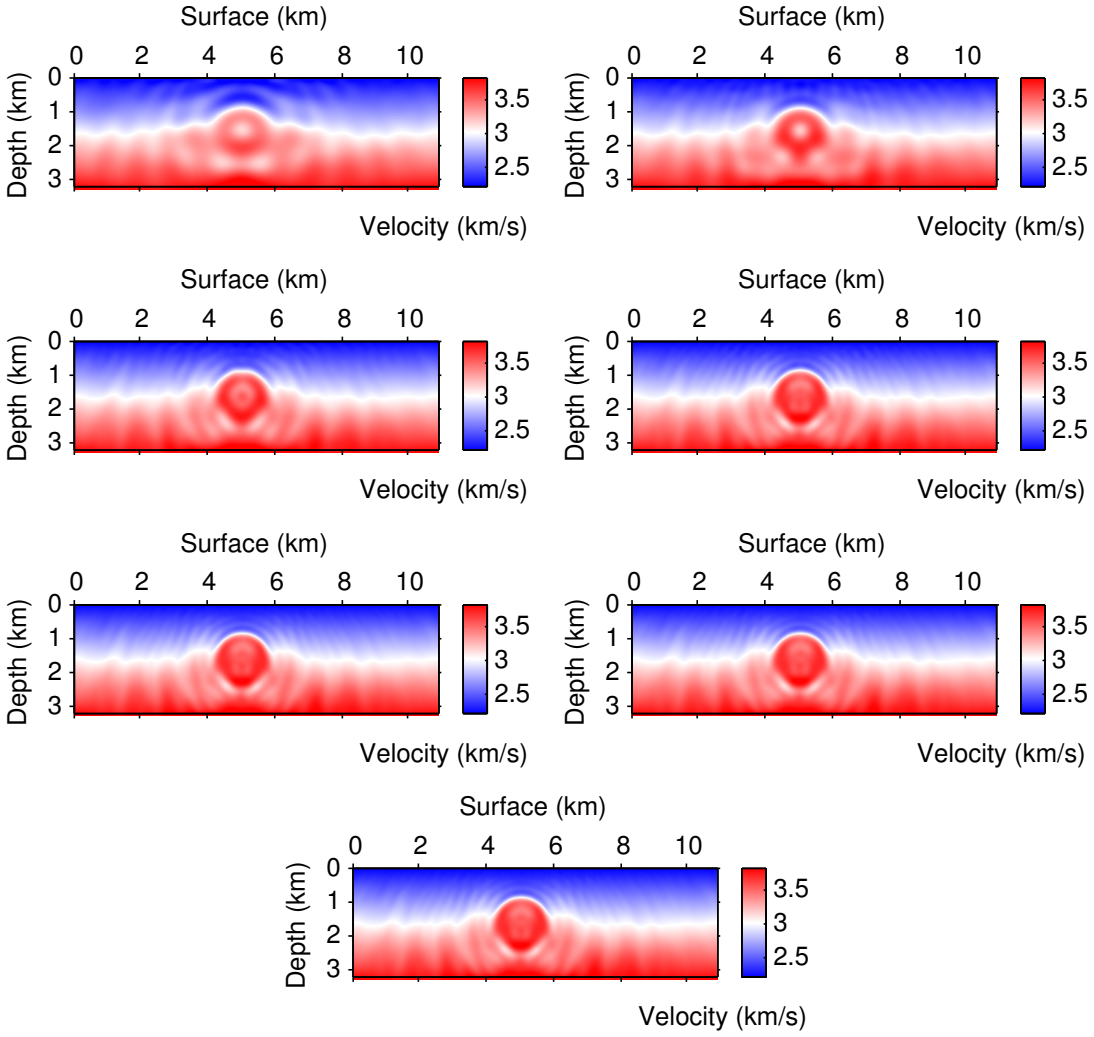


Figure 2-18: Multiscale FWI results. From top to bottom, the frequency used is 1.8 Hz, 2.5 Hz, 3.5 Hz, 5 Hz, 7 Hz, 10 Hz, 14 Hz respectively. For each frequency, 25 iterations of FWI is performed.

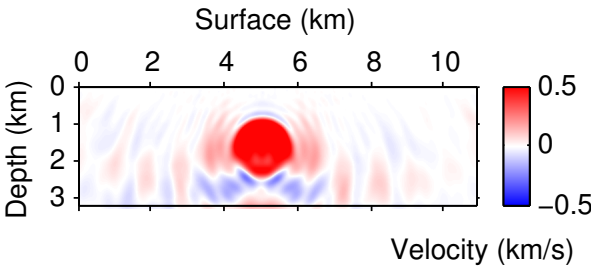


Figure 2-19: Velocity perturbation found with multiscale FWI.

the benefit of long-offset (diving waves, refractions) and transmission data. Moreover, they point out that if a data set mainly contains short-offset reflections, only the short wavelengths of the model are reconstructed by conventional FWI.

Based on this idea, [Shipp and Singh, 2002] propose that the long-offset data should be used in the early stages of inversion, in order to recover the long-wavelength components of the model. At later stages, short-offset data are inverted to provide the short-wavelength components. This strategy of data windowing from long offsets to short offsets is efficient, provided that there is no cycle skipping in the long-offset data. However, as long-offset data correspond to long propagation distance, they introduce increasing non-linearities in an inaccurate model, which makes the application of FWI to long-offset data still challenging [Sirgue, 2006]. The opposite strategy of windowing the data from short-offset to long-offset in order to mitigate the risk of cycle skipping has been also proposed [Sirgue, 2003].

2.10 Conclusion

I have reviewed different aspects of classical FWI, including the objective function, the methods for the numerical modeling of wavefield propagation, the computation of the gradient and the Hessian of the objective function, and the velocity model update methods. I have also presented a resolution analysis of the gradient and show that long-offset data and source with low frequencies help to recover the long-wavelength part of the model. Finally, I have reviewed the difficulties that FWI meet and some methods to mitigate the nonlinearity of FWI.

Chapter 3

Waveform inversion based on wavefield decomposition

Contents

Résumé du chapitre 3

La fonction objective de l'inversion des formes d'ondes peut être fortement non linéaire et présente beaucoup de minima locaux. Si les données contiennent principalement des réflexions, ce problème empêche notamment les méthodes classiques basées sur le gradient de récupérer les grandes longueurs d'ondes du modèle [Pratt et al., 1996]. Avec l'analyse de la résolution du gradient présentée dans le chapitre 2, nous présentons ici une variation de l'approche FWI en profitant de la transmission liée à la réflexion pour construire les grandes longueurs d'ondes du modèle. La méthode est basée sur la séparation du gradient en une partie de courtes longueurs d'ondes et une partie de grandes longueurs d'ondes après la décomposition des champs d'ondes. Nous appelons cette nouvelle méthode en anglais Decomposition-based Waveform Inversion (DWI).

Notons que la formule du gradient de FWI somme le terme des courtes longueurs d'ondes et le terme des grandes longueurs d'ondes avec le même poids. Pour les données de réflexion , le terme des grandes longueurs d'ondes est généralement beaucoup plus faible que le terme des courtes longueurs d'ondes. C'est la raison que la FWI classique souvent n'arrive pas à retrouver les grandes longueurs d'ondes du modèle.

Au lieu de mettre des poids différents, nous proposons d'inverser séparément les deux parties. Supposons que l'on commence par un modèle lisse, d'abord nous inversons les courtes longueurs d'onde en utilisant les données zéro offset. Dans cette étape, nous effectuons l'inversion itérative pour retrouver une réflectivité bien focalisée, qui est essentielle pour l'étape suivante. Notons que cette étape se comporte comme une minimisation au sens des moindres carrés (en anglais, Least-Squares Migration, LSM). L'utilisation des données courts offsets est importante pour ne pas introduire les grandes longueurs d'ondes dans cette étape.

Après l'inversion itérative, le champ d'ondes direct et le champ d'onde rétropropagé maintenant contient les champs descendants et montants. Dans la deuxième étape, nous effectuons la mise à jour des grandes longueurs d'ondes. Dans cette étape, les données avec plus d'offsets sont utilisées pour obtenir les informations moveout. Un lissage sur le gradient est appliqué afin d'éliminer les résidus de courtes longueurs d'ondes, causée par les limitations d'acquisition. Une fois que nous avons mis à jour les grandes longueurs d'ondes du modèle, nous recommençons la mise à jour des courtes longueurs d'ondes du modèle puisque les positions des réflecteurs ne sont pas compatibles avec le modèle de vitesse actuel et devraient être mis à jour en conséquence. Puis nous continuons à alterner entre ces deux étapes. Le point clé de cette méthode d'inversion est ce schéma alternatif.

La limitation d'offset et l'application de l'inversion itérative dans la première étape et le lissage du gradient dans la deuxième étape sont les points importants pour assurer le découplage entre les grandes longueurs d'onde et les courtes longueurs d'onde.

Cette méthode d'inversion basée sur la décomposition est similaire á d'autres méthodes d'inversion basées sur la réflexion. Nous comparons en particulier avec la méthode proposée par [Xu et al., 2012]. La comparaison des gradients calculés dans un modèle 1D montre la similarité entre ces deux méthodes.

Les méthodes de décomposition des champs d'ondes sont également présentées dans ce chapitre. Nous discutons de trois méthodes de décomposition : l'une dans le domaine de Fourier, l'une utilisant le vecteur de Poynting et la troisième dans le domaine des curvelets. La méthode par la transformé de Fourier est la plus populaire, et a été largement util-

isée pour supprimer des artefacts de haute fréquence dans les images migrées, mais avec cette méthode les basses fréquences sont filtrées dans les résultat final. La méthode de Poynting est la plus rapide car elle pourrait être calculée au cours de la propagation des ondes, mais elle souffre du problème d'artefacts numériques fortes (instabilités). La méthode de décomposition par les curvelets donne le meilleur résultat car il peut prendre en compte les différentes directions en même temps, avec le coût de calcul le plus élevé des trois. Dans cette méthode, le champ d'onde est décomposé en champs d'onde localement plans. La corrélation permet de savoir dans quelle direction l'onde se propage, sans avoir à faire d'hypothèse sur la direction de propagation. Ce n'est pas le cas avec la méthode de Fourier pour laquelle il faut savoir si les ondes se propagent essentiellement horizontalement ou verticalement. La méthodologie a été développée pour corréler les champs directement dans le domaine des curvelets. La nouvelle méthode d'inversion et les méthodes de décomposition sont illustrées par un modèle de couche 1D. L'application de la nouvelle méthode d'inversion aux modèles synthétiques et aux données réelles est abordée dans les chapitres 4 et 5.

3.1 Introduction

The objective function of the Full Waveform Inversion (FWI) may be highly nonlinear and has many local minima. If a data set mainly contains reflections, this problem particularly prevents the classical gradient-based methods from recovering the long wavelengths of the model [Pratt et al., 1996]. In this chapter, we present a variation of FWI by separating the gradient of FWI into a migration part and a tomographic part. We call this new method Decomposition-based Waveform Inversion (DWI). The scale separation in the gradient is based on the wavefield decomposition into their one-way components. The wavefield decomposition methods are also presented in this chapter. The new inversion method and the decomposition methods are illustrated through a 1D layer model. The application of the new inversion method on synthetic models and real data set will be addressed in Chapter 4 and Chapter 5.

3.2 Methodology

Based on the resolution analysis of the gradient presented in Chapter 2, we propose a new inversion method taking advantage of the reflection-related transmission to construct the long wavelengths of the model. The method is based on the gradient separation into a short-wavelength part and a long-wavelength part. The scale separation in the gradient is achieved by the wavefield decomposition.

Given a non-smooth model m, both forward wavefield P and back-propagated wavefield Q contain up- and down-going wavefields, as shown in Figure 3-1 (top panel). They can therefore be separated into:

$$\begin{cases} P &= P^+ + P^-, \\ Q &= Q^+ + Q^-, \end{cases} \tag{3.1}$$

where P^+ is the direct forward wavefield, P^- is the reflected forward wavefield, Q^- is the direct backward wavefield and Q^+ is the reflected backward wavefield. By substituting the decomposed wavefields in equation (3.1) into the gradient formula (equation (2.18)), the gradient G can be decomposed into four terms:

$$G = P \otimes Q = P^+ \otimes Q^- + P^- \otimes Q^+ + P^+ \otimes Q^+ + P^- \otimes Q^-, \tag{3.2}$$

where $\otimes$ indicates the correlation between two wavefields. For the first two terms, the wavefields P^+ and Q^-, as well as P^- and Q^+, propagate in the same directions (both downward or upward) and coincide in time at the positions of reflectors (Figures 3-1 second and third panel). Thus, their correlations $P^+ \otimes Q^-$ and $P^- \otimes Q^+$ provide the short-wavelength update of the model (reflectivity) and correspond to the migration imaging principle proposed by [Claerbout, 1971]. The other two terms $P^+ \otimes Q^+$ and $P^- \otimes Q^-$ are illustrated in Figure 3-1 (fourth and fifth panel). As the back-propagated wavefield Q is reversed in time, wavefields P^+ and Q^+, as well as P^- and Q^- propagate in the opposite direction and coincide in time along the whole propagation path that connects the source to the reflector, and the reflector to the receiver. Thus, their correlations provide the long wavelength update of the model (background velocity) [Zhou et al., 1995].

Suppose that the velocity model m consists of a background model m_0 and a perturba-

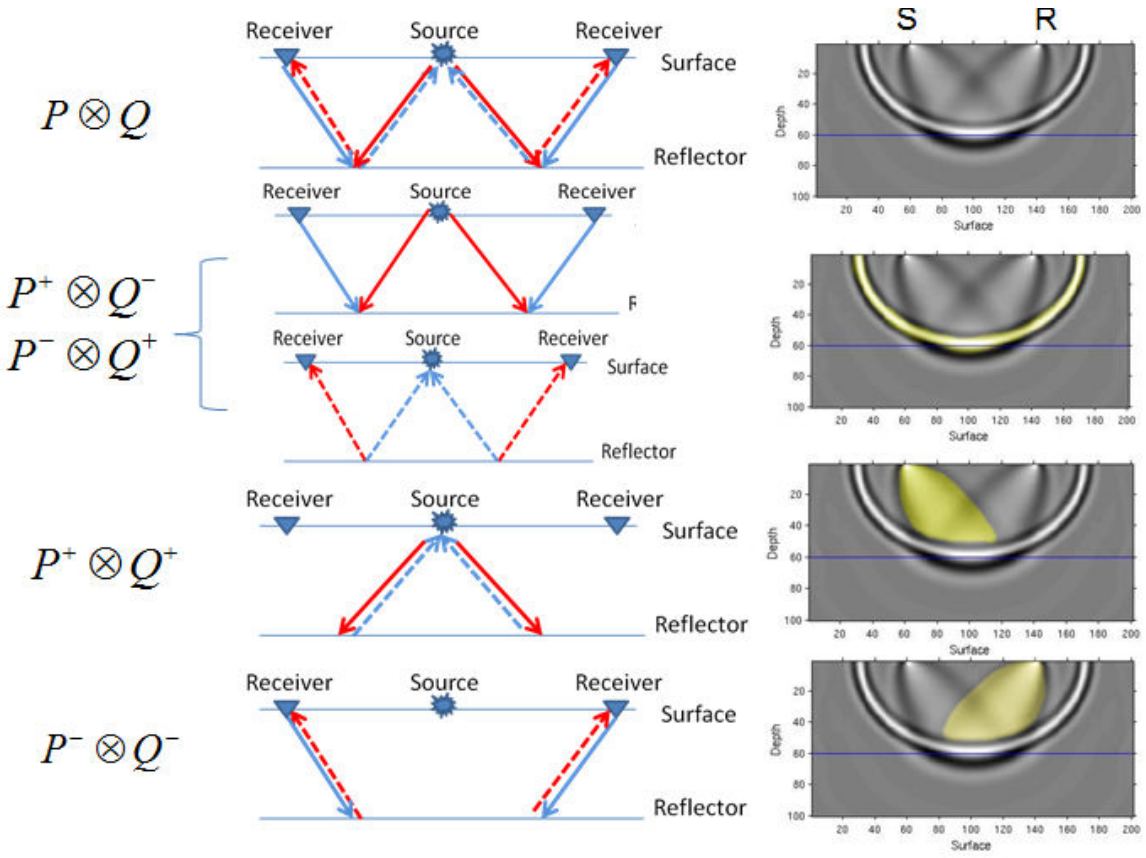


Figure 3-1: Raypaths and illumination in the model space for different components of the gradient. Schematic wave propagation in 1D layer model. P^+: direct forward wavefield (red solid line). P^-: reflected forward wavefield (red dashed line). Q^-: direct backward wavefield (blue solid line) and Q^+: reflected backward wavefield (blue dashed line).

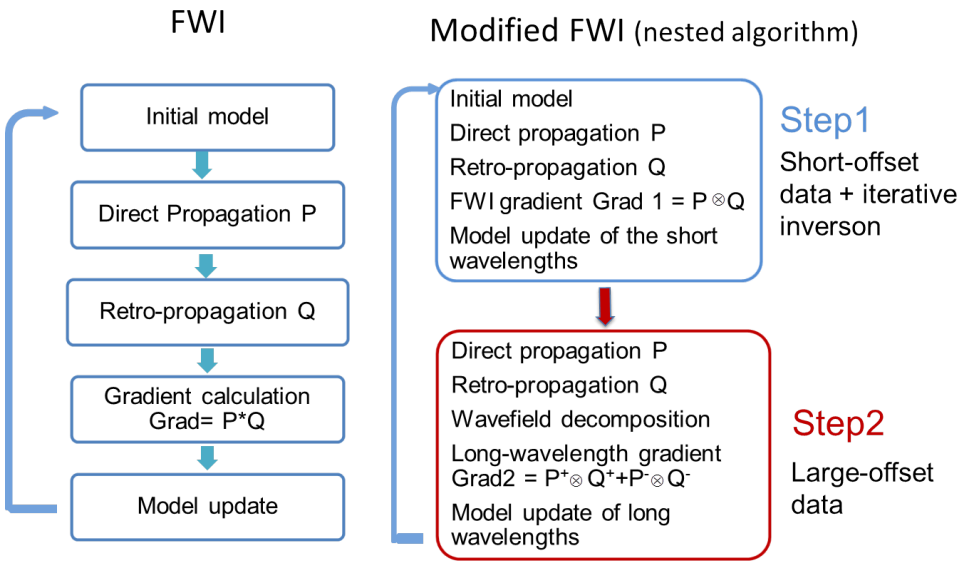


Figure 3-2: Comparison of conventional FWI (left) and the decomposition-based waveform inversion (right).

tion model r: $m = m_0 + r$. If all frequencies are considered at the same time, the gradient of the short-wavelength update is given by

$$G_r = P^+ \otimes Q^- + P^- \otimes Q^+, \tag{3.3}$$

while the long-wavelength update is given by

$$G_{m_0} = P^+ \otimes Q^+ + P^- \otimes Q^-. \tag{3.4}$$

3.3 Inversion strategy

Note that the conventional FWI gradient formula (equation (3.2)) sums the long-wavelength term G_{m_0} and short wavelength term G_r with equal weights. For reflection-dominant data, the long-wavelength term G_{m_0} is usually much weaker than the short wavelength term G_r, of a factor r^2, where r is a typical reflectivity value. [Mora, 1987] and [Tang and Lee, 2013] propose to enhance the long-wavelength term G_{m_0} in the gradient by applying different weights to G_r and G_{m_0}. This is not trivial as the weighting factor is difficult to determine and may vary with models and iterations. [Tang and Lee, 2013] propose to determine the weight at each iteration by solving a linear optimization problem which involves the calculation of the inverse of the Hessian.

Instead of setting different weights, we propose to separately invert the short and long wavelengths of the model. Let start with a smooth initial velocity model m_0. First we retrieve the short wavelengths of the model using equation (3.2) instead of equation (3.3) to reduce the computing cost, as it avoids to decompose the wavefields into different parts. This is feasible because the gradient amplitudes are dominated by the short wavelengths

especially in the first stages of the inversion. It is important to note that in this step, we use zero or short offset data and we perform iterative inversion. The short-wavelength model obtained in this first step contains reflectors which are needed for the successive step.

Note that the short-wavelength update step behaves like a least-squares migration, but it is not exactly the same as the separation of the long wavelength and the short wavelength is not explicit in this step, thus the inversion is not linearized. Instead, we perform the iterative inversion on zero-offset data with all frequencies simultaneously. The use of short-offset data is crucial because the long-wavelength components of the model are not updated during this step, as no moveout information is available. The iterative inversion helps to correct the reflector amplitude as well as to achieve better focusing of the reflectors [Plessix and Mulder, 2004], which allows for a more accurate long wavelength update. However, the reflectors are mispositioned in depth at this stage as the used background velocity model is probably erroneous.

After the iterative inversion, the forward wavefield P and backward wavefield Q found at the end of the first step are now decomposed into their up- and down-going components. We perform the long-wavelength update using equation (3.4). In this step, longer offset data is used to get the moveout information. A smoothing filter is applied to the gradient, in order to remove the residual short-wavelength components, that are caused by acquisition limitations. We carry out one iteration of inversion in this step and the velocity is updated using the smoothed gradient.

Once we have updated the long-wavelength model, we use this smooth model to update once more the short-wavelength model using the iterative inversion, as the positions of reflectors are not consistent with the actual velocity model any more and should be updated accordingly. At each global iteration, the reflectivity is set to zero before being estimated. Then we continue alternating between these two steps. The key point of the proposed inversion method is this alternating scheme. The comparison of the conventional FWI and the proposed method is shown in Figure 3-2.

The offset limitation and the application of iterative inversion in the first step and the smoothing of the gradient in the second step are the key issues to ensure the decoupling between the long wavelength and short wavelength updates of the velocity model, as we will describe in more details in the following subsections. A 1D-layer model will be used to better illustrate the method.

3.3.1 Choice of offset range

In the first step of the inversion, we use zero-offset or short-offset data in order to preserve the vertical traveltime. In this step, we do not want to introduce the long-wavelength update. As the zero-offset traveltime is preserved, there is no moveout information in the data residual.

3.3.2 Iterative inversion

In the first step of inversion, we perform the iterative inversion to derive an optimal reflectivity model. As the second step of inversion is sensitive to the reflectivity model obtained in the first step, building a true-amplitude reflectivity is important. Otherwise,

residuals at short offsets will be introduced in the long-wavelength update and degrade the inversion result. An example describing this issue will be shown in Chapter 4. The reflectivity is usually obtained by migration methods. However, as the migration uses the adjoint of forward-modeling operator instead of the inverse operator, they could not preserve the true amplitude of migration images. For Reflection-based Waveform Inversion (RWI), usually Least-Squares Migration (LSM) is used in the first step of inversion. LSM is an iterative migration method resolving the linearized inversion problem. It has been shown to improve amplitude information and to focus migrated images. The sensitivity of the update of the macro model to the reflectivity model is also analyzed in [Lameloise et al., 2015, Symes, 2015] by showing that the global minimum of the objective function of the Differential Semblance Optimization (DSO) does not necessary coincide with the exact model due to the artifacts present in the reflectivity model, and that the inversion result can be improved by performing the iterative migration or the quantitative migration preceding the update of the long-wavelength part of the model. In our method, as the scale separation is not explicit, instead of performing the LSM, we perform the nonlinear inversion in the first step to maintain the compatibility between the two steps of inversion.

3.3.3 Gradient smoothing

In the second step of the inversion, we try to retrieve the long-wavelength components of the model. In order to remove the residual short-wavelength components in the gradient, we apply a smoothing operator to the gradient. After 2D Fourier transform of the gradient, a 2D low-pass elliptic filter is applied to the transformed gradient. The elliptic filter is defined by the model dimension and the maximum preserved wavenumber. A taper zone is applied to prevent artifacts caused by discontinuity. After filtering the gradient in the wavenumber domain, an inverse Fourier transform is performed to recover the filtered gradient in the space domain.

3.3.4 1D layer model case

Before presenting the wavefield decomposition scheme, let use the 1D layered model in Figure 2-8 to study the characteristics of the proposed method. A Ricker wavelet with central frequency of 12 Hz and temporal sampling of 1.5 ms is used as the source function.

The initial velocity model is displayed in Figure 3-3. This is the reflectivity model and the model contains mainly short-wavelength components. Here, we artificially use the exact reflectivity to analyse the different wavefields. Later, the reflectivity model is obtained through iterative inversion. Starting from this model we calculate the forward wavefield P and backpropagated wavefield Q for the single sources located at 0.85 km at the surface. Usually, all the available source positions are used, but in this case we use a single source to more easily illustrate the behavior of the proposed algorithm. Then, we decompose P and Q into their up- and down-going parts to get four correlation terms: $P^+ \otimes Q^-$, $P^- \otimes Q^+$, $P^+ \otimes Q^+$ and $P^- \otimes Q^-$. The decomposition is achieved using the 2D Fourier-transform based method, which will be described in the next section. For one source and one receiver setup, the four correlation terms are shown in Figure 3-4 and Figure 3-5. In these figures, the source is fixed at 0.85 km and the receivers are at three

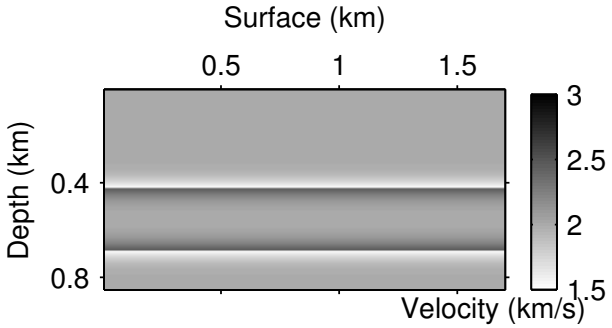


Figure 3-3: The initial velocity model. It is the sum of the homogeneous model at 2 km/s and the true reflectivity model. The true reflectivity is obtained by subtracting the smoothed true velocity model from the exact velocity model of Figure 2-8.

different positions. For one source and all receivers setup, the four correlation terms are shown in Figure 3-6. As expected, these images are in accordance with the wavepath illustration in Figure 3-1.

If we stack for all the sources at the surface, we obtain the results for the correlation of P and Q in Figure 3-7 on the left. The images on the right of Figure 3-7 show the result of taking the vertical velocity profile at 0.85 km and analyze the spectrum using the Fourier transform. The four correlations of decomposed wavefields are shown in Figure 3-8 on the left. Their corresponding spectra of the central vertical slice are displayed in Figure 3-8 on the right. We can observe that the spectrum of $P \otimes Q$ contains both high wavenumbers and low wavenumbers. Instead, for the decomposed case, the spectra of $P^+ \otimes Q^-$ and $P^- \otimes Q^+$ mainly contain the high-wavenumber part of $P \otimes Q$, while $P^+ \otimes Q^+$ and $P^- \otimes Q^-$ mainly contain the low-wavenumber part of $P \otimes Q$.

3.4 Comparison with similar methods

The decomposition-based FWI has many common features with other Reflection-based Waveform Inversion (RWI) methods. Here, we particularly compare our method to the one proposed by [Xu et al., 2012] as an example. The method proposed by [Xu et al., 2012] is based on a linearization (Born approximation) and on the explicit separation of the velocity model m into a smooth background model m_0 and a reflectivity model δm. The forward problem is described by the wave equation:

$$L(m)P = S(\omega)\delta(\mathbf{x} - \mathbf{s}), \tag{3.5}$$

where $L(m) = -\omega^2 m - \Delta$ is the forward propagator. P is the wavefield, composed of P_0 and δP, with $P = P_0 + \delta P$. P_0 is the wavefield propagating in the smooth model m_0 and δP is the wavefield generated by the model perturbation δm. The objective function is defined as:

$$J(\delta m) = \frac{1}{2}\|\delta P - \delta P^{obs}\|^2, \tag{3.6}$$

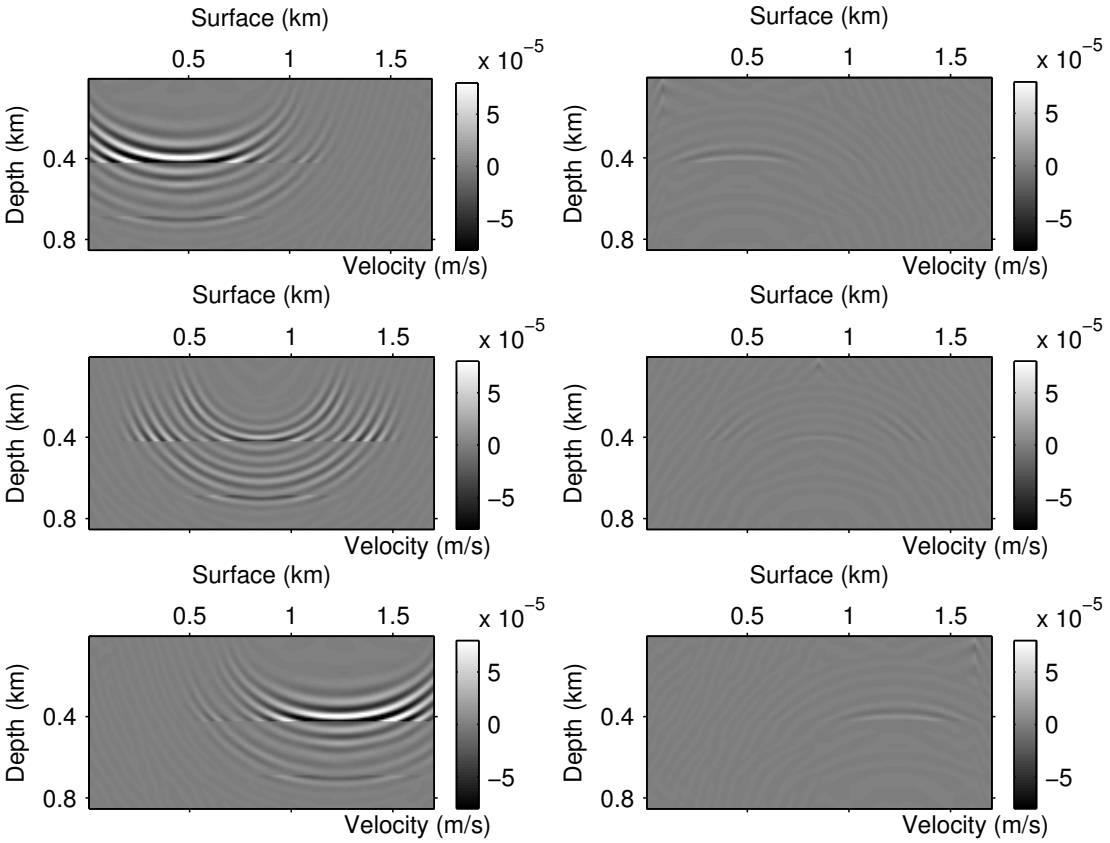


Figure 3-4: Correlation of decomposed forward and backward wavefields for one source-receiver pair with the model in Figure 3-3. The source is located at 0.85 km. From top to bottom, the receivers are at 0.085 km, 0.85 km and 1.615 km respectively. Correlation of P^+ and Q^- (left); P^- and Q^+ (right).

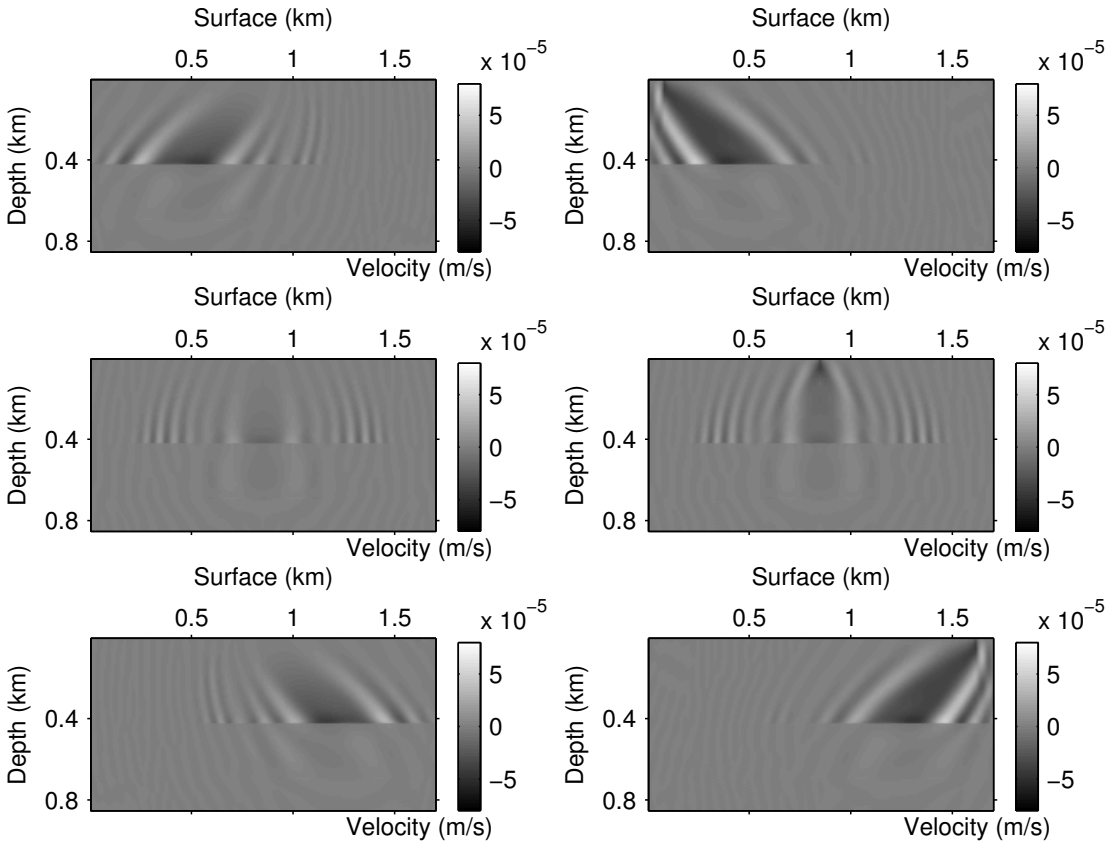


Figure 3-5: Correlation of decomposed forward and backward wavefields for one source-receiver pair with the model in Figure 3-3. The source is located at 0.85 km. From top to bottom, the receivers are at 0.085 km, 0.85 km and 1.615 km respectively. Correlation of P^+ and Q^+ (left); P^- and Q^- (right).

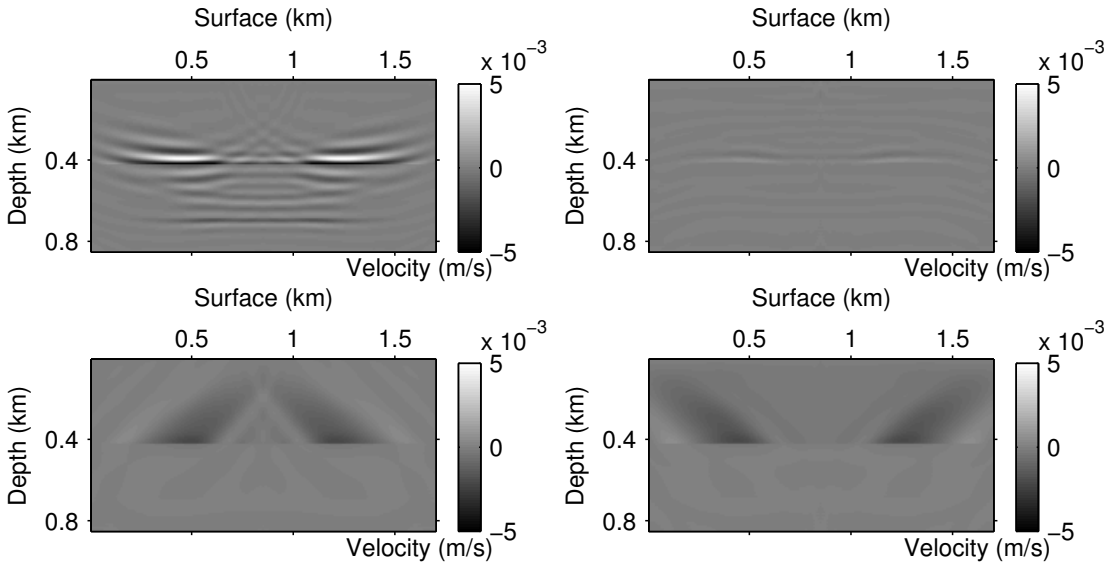


Figure 3-6: Correlation of decomposed forward and backward wavefield for the source located at 0.85 km with the model in Figure 3-3. Correlation of P^+ and Q^- (top left); P^- and Q^+ (top right); P^+ and Q^+ (bottom left); P^- and Q^- (bottom right).

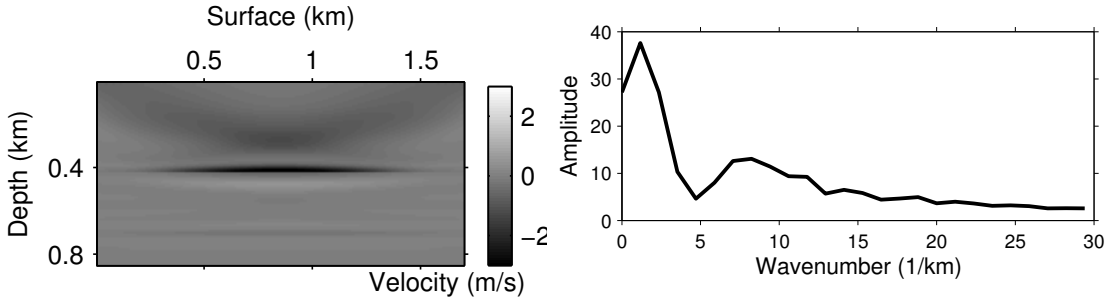


Figure 3-7: (Left) Correlation of forward wavefield P and backward wavefield Q using all the available sources at the surface, starting from the model in Figure 3-3. (Right) Amplitude of spectrum of the vertical velocity profile taken at 0.85 km in the figure on the left.

where δP^{obs} is the observed wavefield generated by the true model perturbation. At first, we deduce δP by using the Born approximation for the wavefields P and P_0:

$$\begin{cases} L(m)P = S(\omega)\delta(\mathbf{x}-\mathbf{s}), \\ L(m_0)P_0 = S(\omega)\delta(\mathbf{x}-\mathbf{s}). \end{cases} \tag{3.7}$$

By subtracting these two equations, we get the formula to compute δP:

$$L(m_0)\delta P = \omega^2 \delta m P_0. \tag{3.8}$$

The adjoint-state method [Plessix, 2006] is used to compute the gradient of the objective function in equation (3.6) with respect to m_0 and δm. Two adjoint variables Q_0 and δQ

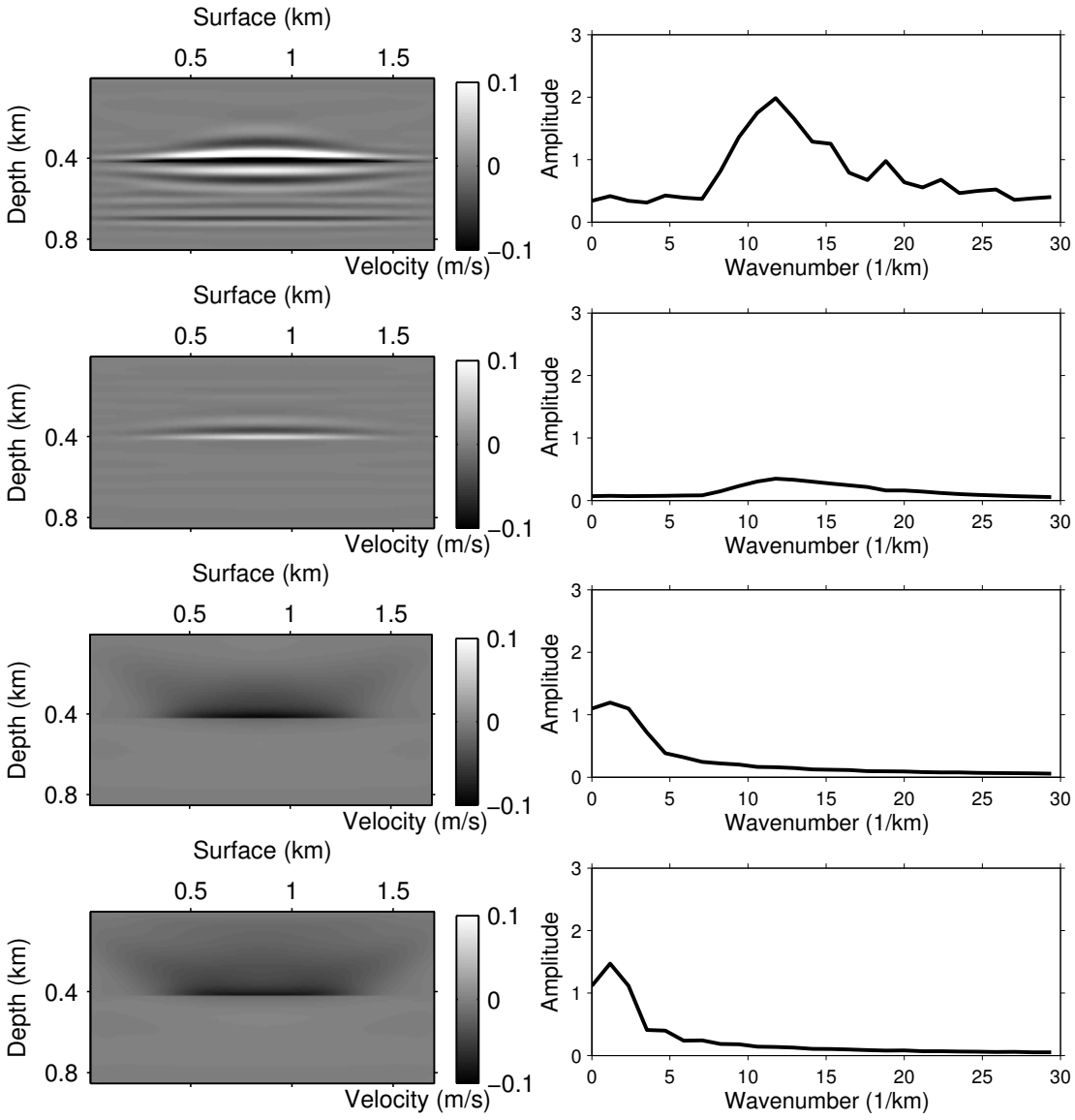


Figure 3-8: Correlation of decomposed forward wavefield and backward wavefield using all the available sources at the surface, starting from the model in Figure 3-3 (left). From top to bottom, correlation of P^+ and Q^-; P^- and Q^+; P^+ and Q^+; P^- and Q^- respectively. (right) Spectra of the vertical velocity profiles taken at 0.85 km for figures on the left respectively.

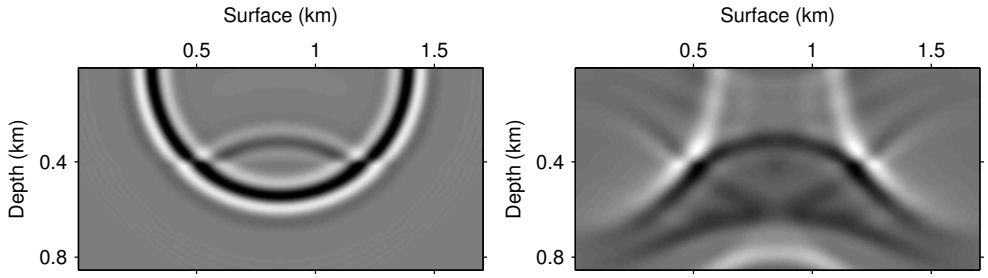


Figure 3-9: Direct and backpropagated wavefield in the model in Figure 2-8.

are introduced and the augmented function $\widetilde{J}$ is:

$$\begin{aligned}\widetilde{J} &= \frac{1}{2}\|\delta P - \delta P^{obs}\|^2 \\ &+ \iiint \delta Q[L(m_0)P_0 - S(\omega)\delta(\mathbf{x}-\mathbf{s})]\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{x}\mathrm{d}\omega \\ &+ \iiint Q_0[L(m_0)\delta P - \omega^2\delta m P_0]\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{x}\mathrm{d}\omega.\end{aligned} \tag{3.9}$$

We compute the gradient of $\widetilde{J}$ with respect to m_0 and δm respectively:

$$\frac{\partial \widetilde{J}}{\partial \delta m} = \iiint (-\omega^2 P_0(\mathbf{s},\mathbf{x},\omega)Q_0(\mathbf{s},\mathbf{x},\omega))\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{x}\mathrm{d}\omega, \tag{3.10}$$

$$\frac{\partial \widetilde{J}}{\partial m_0} = \iiint [-\omega^2 P_0(\mathbf{s},\mathbf{x},\omega)\delta Q(\mathbf{s},\mathbf{x},\omega) - \omega^2\delta P(\mathbf{s},\mathbf{x},\omega)Q_0(\mathbf{s},\mathbf{x},\omega)]\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{x}\mathrm{d}\omega. \tag{3.11}$$

Q_0 and δQ are computed by the two back-propagation wave equations:

$$\begin{cases} \frac{\partial \widetilde{J}}{\partial \delta P} = L^* Q_0 + (\delta P - \delta P^{obs}) = 0, \\ \frac{\partial \widetilde{J}}{\partial P_0} = L^* \delta Q - \omega^2 Q_0 \delta m = 0. \end{cases} \tag{3.12}$$

where $*$ denotes the adjoint operator.

From equation (3.11) we observe that four wavefields are involved in the gradient calculation for the background model m_0. They are respectively the direct source wavefield P_0 and the direct receiver wavefield Q_0, both propagating from the surface to the subsurface reflector, as well as the perturbed source wavefield δP and perturbed receiver wavefield δQ, both propagating from the reflector back to the surface. The reflectivity δm serves as a secondary source in depth that emits up-going waves towards the surface.

The decomposition-based gradient in equation (3.4), related to the long-wavelength update, has a similar structure compared to the gradient in equation (3.11). We first show the comparison using the model in Figure 3-3. We compute the original forward wavefield P and the back-propagated wavefield Q (Figure 3-9), then the decomposed wavefields as described in equation (3.4) (Figure 3-10). The four components in the long-wavelength gradient (equation (3.11)) are shown in Figure 3-11. The gradient calculated with all the

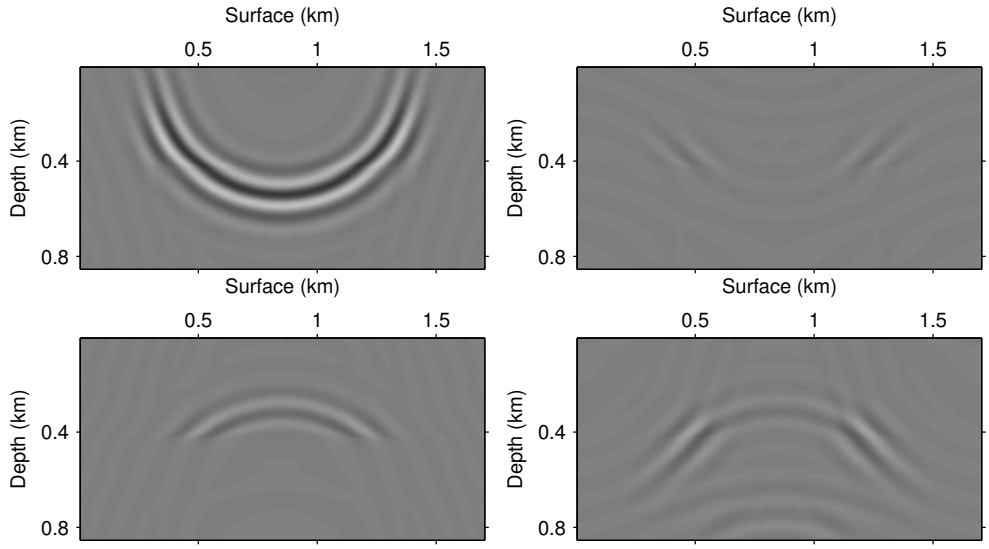


Figure 3-10: Decomposed wavefields for the initial wavefields in Figure 3-9. P^+ (top left), Q^+ (top right), P^- (bottom left) and Q^- (bottom right).

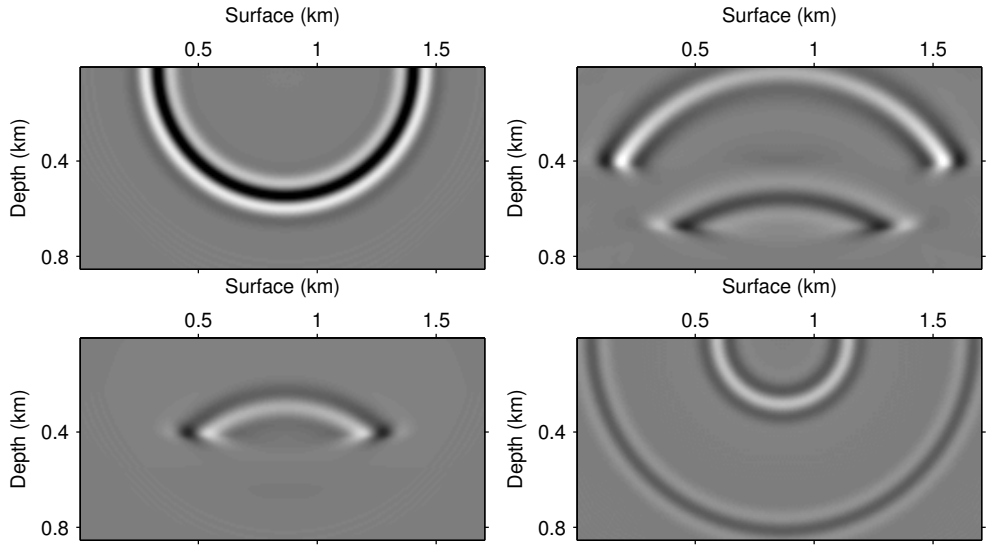


Figure 3-11: Four wavefields in equation (3.11). P^0 (top left), δQ (top right), δP (bottom left) and Q^0 (bottom right).

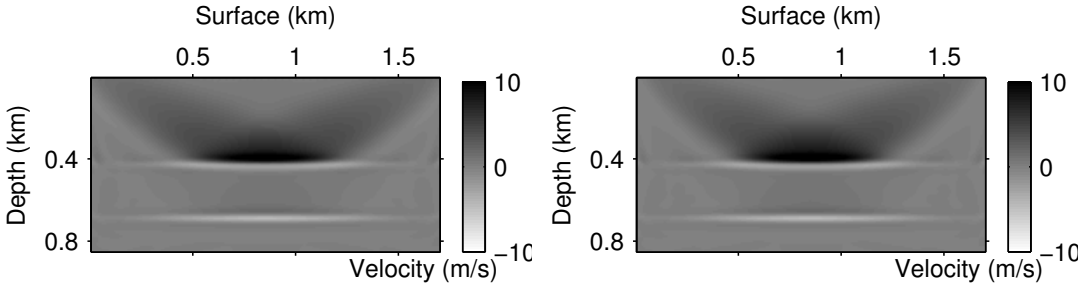


Figure 3-12: Gradients in equation (3.11). Correlation of (left) P^0 and δQ, (right) δP and Q^0. To be compared with Figure 3-8 (last two panels).

sources and receivers, using equation (3.11) is displayed in Figure 3-12.

If we compare the correlation of P^+ and Q^+, with the correlation of P^0 and δQ, we can observe that with the time, they both move along the wavepath from the surface to the reflector. The same conclusion could be drawn when we compare the correlation of P^- and Q^-, with the correlation of δP and Q^0, while they both move along the wavepath from the reflector to the surface. The gradients in Figure 3-12 and in Figure 3-8 (last two panels) are also similar. In this sense, we could consider decomposition-based FWI as an analogue of the method proposed by [Xu et al., 2012]. The mayor difference resides in how to calculate the four components in the gradient. For the method proposed by [Xu et al., 2012] they are obtained by resolving the linearized wave equation, and for the method proposed in this thesis, they are obtained by decomposing the full wavefields after resolving the nonlinear wave equation.

However, as the method of [Xu et al., 2012] is based on the linearized modeling, it assumes that the velocity perturbations are small and the data contain only primaries. While similarly to classical FWI, the proposed decomposition-based FWI method resolves the nonlinear wave equation; therefore refractions, diving waves and multiples could potentially be included in the dataset. [Symes, 2008] suggests that MVA is a solution method for the linearized waveform inversion problem.

Besides the method of [Xu et al., 2012], there are in the literature other similar methods, also based on separating different components in the gradient. [Tang and Lee, 2013] also propose the gradient separation by the wavefiled decomposition. The difference is that they recombines the two parts in the gradient with different weights. [Brossier et al., 2015, Zhou et al., 2015] propose to get the tomographic part by subtracting the migration part that is obtained in the smooth model from the full gradient. [Zhou et al., 2015] propose to combine the reflections with the diving waves to enhance the long-wavelength components of the model. [Wang et al., 2015b] propose to combine the refractions with reflections. For reflection data, the long-wavelength part is updated using the tomographic part in the gradient by including the migration part within the density model. [Alkhalifah, 2015] propose to do the separation in the angle domain by filtering the gradient in the wavenumber domain according to the scattering angles and then to use the gradient corresponding to big scattering angles for long-wavelength update and the gradient corresponding to small scattering angles for short-wavelength update. Migration Based Traveltime Tomography (MBTT) [Chavent et al., 1994] combines the least-squares migration and the multiscale FWI to mitigate the nonlinearity.

3.5 Wavefield decomposition methods

The gradient decomposition into four terms is possible with the benefit of the wavefield decomposition. The objective of the wavefied decomposition is to obtain the one-way component of the wavefield with respect to the vertical direction (donwgoing/upgoing) and the horizontal direction (leftgoing/rightgoing). The wavefield decomposition is mainly used in the literature for:

1. Providing a new imaging condition. The conventional zero-lag cross-correlation imaging condition of RTM is subject to strong migration artifacts [Liu et al., 2011].

The decomposition of the full wavefields to their one-way components and the application of the imaging condition to the appropriate combinations of the wavefield components could remove the undesired noise.

2. Retrieving the long wavelength components of the velocity model. The gradient of FWI could be separated into a migration part and a tomography part, after the wavefield decomposition. Thus, the wavefield decomposition is used for recovering the long-wavelength part of the model [Tang and Lee, 2013, Wang et al., 2013, Wang et al., 2015a].

Here, we give a review of different methods to produce partial wavefields or to decompose the full wavefields. In general, the wavefield decomposition methods could be divided into one-way based methods and two-way based methods.

3.5.1 Methods based on one-way wave equation

To produce one-way wavefields, the straightest way is to use the one-way approximation of the full wave equation. The one-way wave equation is easy to implement, but has dip limitations. In [Zhang et al., 2006], the authors propose a modified version of one-way wave equation to decompose wavefields. With this method, turning waves can be properly imaged and the imaging capability of one-way wave equation is greatly improved.

3.5.2 Methods based on two-way wave equation

Methods based on two-way wave equation has no dip limitation, and are able to deal with multiples and turning waves. They produce a better resolution and are more robust with respect to strong velocity variations. There are two major categories to obtain the one-way component of wavefields using the two-way wave equation. The first one is by suppressing internal reflections, and the second one is by decomposing wavefields.

Suppression of internal reflections

These methods aim at modifying the impedance of the media to suppress internal reflections:

- **Impedance matching**. [Whitmore et al., 1983] and [Baysal et al., 1984] proposed a method to suppress undesired reflections by forcing the acoustic impedance, $I = \rho(x, y, z)v(x, y, z)$, to be a constant, and deriving an appropriate impedance-matched counterpart to the full wave equation.

- **Smoothing of slowness field [Loewenthal et al., 1987]**. This method is based on the smoothing of the slowness field throughout the media by an operator which is longer than the wavelengths associated with the acoustic wave. The reason of using slowness instead of velocity is that the integral of the original slowness curve and the integral of the smoothed slowness curve are the same, while the integral of the smoothed velocity curve differs from the original one and causes a shift in the arrival time.

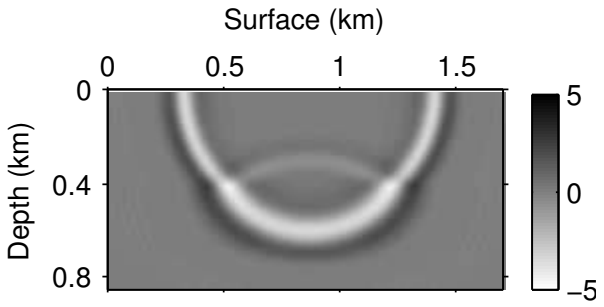


Figure 3-13: One snapshot of the wavefield propagated in the model in Figure 2-8 at $t = 0.525$ s.

- **Add of a damping term [Fletcher et al., 2005]**. This method applies a directional damping term to the wave equation in areas of the velocity model where unwanted reflections occur. A linear derivative operator is used. This linear derivative operator radiates in one direction and damps in the opposite direction.

Decomposition of wavefields

These methods compute the full wavefield and then decompose it according to the propagation direction. We analyze here three methods for achieving the decomposition of wavefields, namely the 2D Fourier transform, the Poynting-vector method and the curvelet transform.

- **2D Fourier Transform (FT) decomposition**.
  [Liu et al., 2011] applied wavefield decomposition in the F-K domain to suppress the low-frequency artifacts in RTM images. This method was first applied to vertical seismic profiles [Hu and McMechan, 1987]. The wavefield is decomposed into its upgoing and downgoing components by using the 2D Fourier transform:

$$\begin{aligned} \widetilde{P}_{z+}(f,k_z) &= \begin{cases} \widetilde{P}(f,k_z) & \text{for } fk_z \geq 0 \\ 0 & \text{for } fk_z < 0 \end{cases}, \\ \widetilde{P}_{z-}(f,k_z) &= \begin{cases} 0 & \text{for } fk_z \geq 0 \\ \widetilde{P}(f,k_z) & \text{for } fk_z < 0 \end{cases}, \end{aligned} \tag{3.13}$$

  where $\widetilde{P}(f,k_z)$ is the 2D Fourier transform of $P(t,z)$. $P_{z+}(t,z)$ and $P_{z-}(t,z)$ are the decomposed down- and up-going wavefields. They are the inverse Fourier transform of $\widetilde{P}_{z+}(f,k_z)$ and $\widetilde{P}_{z-}(f,k_z)$ respectively.

  The horizontal decomposition is obtained similarly, by replacing z by x.

$$\begin{aligned} \widetilde{P}_{x+}(f,k_x) &= \begin{cases} \widetilde{P}(f,k_x) & \text{for } fk_x \geq 0 \\ 0 & \text{for } fk_x < 0 \end{cases}, \\ \widetilde{P}_{x-}(f,k_x) &= \begin{cases} 0 & \text{for } fk_x \geq 0 \\ \widetilde{P}(f,k_x) & \text{for } fk_x < 0 \end{cases}. \end{aligned} \tag{3.14}$$

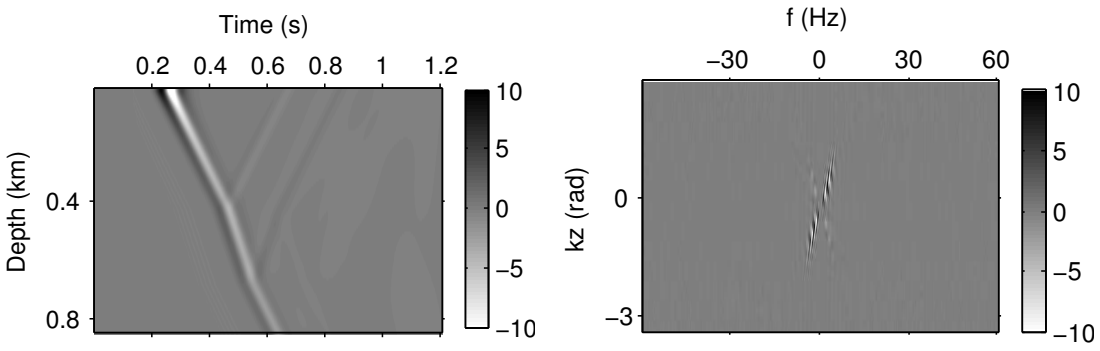


Figure 3-14: (Left) The section of the wavefield in the $t-z$ domain for $x = 0.85$ km in the middle of the model. (Right) FT of the figure on the left in the $f-k_z$ domain.

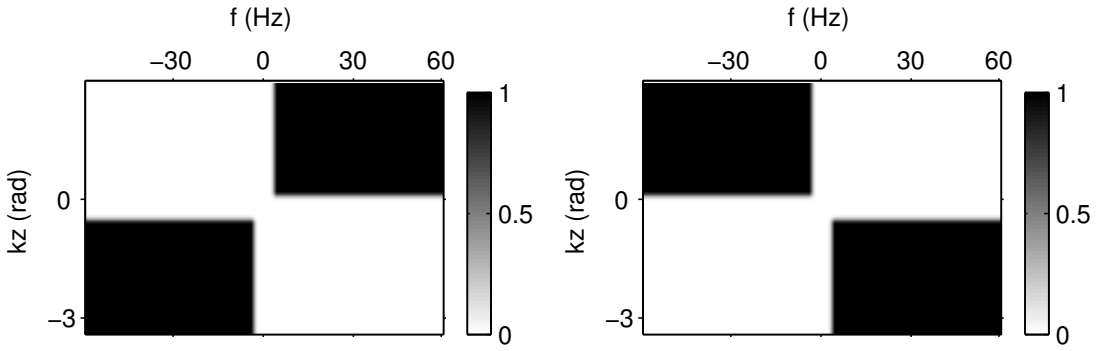


Figure 3-15: The down-going and up-going filter in the $f-k_z$ domain.

We apply this method on a 2D wavefield $P(t,z,x)$. The wavefield is computed in the model of Figure 2-8 with the source at $x = 0.85$ km at the surface. A snapshot at $t = 0.525$ s is shown in Figure 3-13. The algorithm for the vertical decomposition is described as follows.
For the central trace ($x = 0.85$ km) of the image:

1. The image in $t-z$ domain $P(t,z)$ is shown in Figure 3-14 (left).
2. The 2D FT of $P(t,z)$ gives $\widetilde{P}(f,k_z)$ in the $f-k_z$ domain, as shown in Figure 3-14 (right).
3. We select appropriate quadrants using the filters in Figure 3-15. $\widetilde{P}(f,k_z)$ is separated into a down-going part $\widetilde{P}_{z+}(f,k_z)$ and an up-going part $\widetilde{P}_{z-}(f,k_z)$ (Figure 3-16).
4. Then we apply the 2D inverse FT on $\widetilde{P}_{z+}(f,k_z)$ and $\widetilde{P}_{z-}(f,k_z)$ to get the decomposed images in the $t-z$ domain (Figure 3-17).
5. We repeat steps (1)-(4) for each horizontal coordinate x, and we sum all the contributions to get $P^+(t,z,x)$ and $P^-(t,z,x)$, shown in Figure 3-18.

The drawbacks of the FT-based decomposition are:

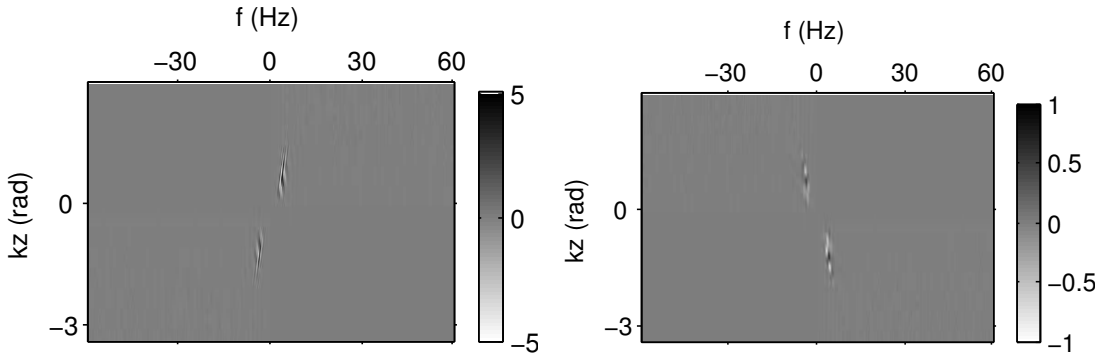


Figure 3-16: Filtered parts of the initial image in Figure 3-14.

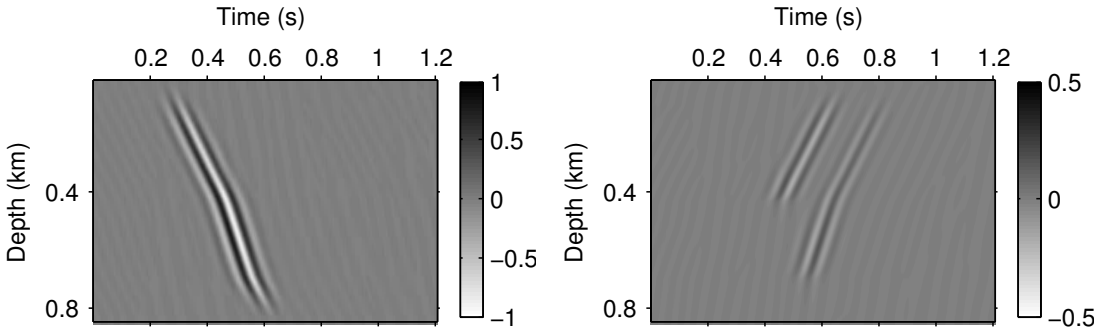


Figure 3-17: Decomposed wavefields in the $t - z$ domain. Notice the different color-scale between the two images.

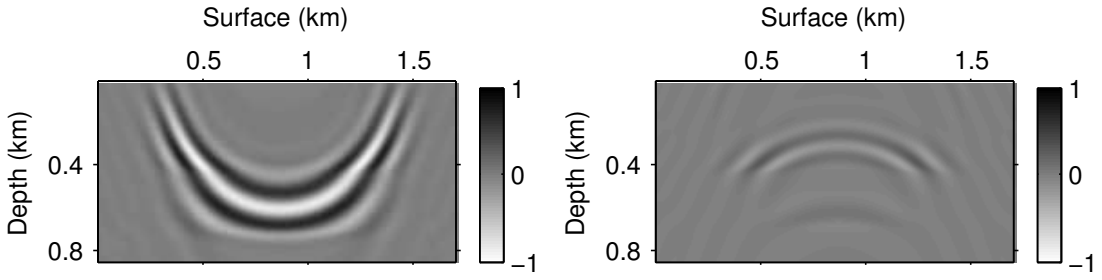


Figure 3-18: Decomposed wavefields in the space domain using the 2D FT-based decomposition. P^+ (left) and P^- (right).

1. It can not handle vertical directions and horizontal directions at the same time. Usually, vertical decomposition is performed for models with sources and receivers at the surface. Therefore, in this case, artifacts will be generated by wavefield propagating laterally. When a point source is used, direct waves propagate horizontally along the surface. That is why artefacts tend to appear near the surface.

2. As we separate data in the $f-k$ domain, we need to apply a smooth cut in the $f-k$ domain (Figure 3-15) to reduce the artefacts due to discontinuities. Thus information near the zero axes in the F-K domain is partially filtered and this information correspond to the low frequency component of the wavefield. Compared to Figure 3-13, the two images in Figure 3-17 look like a low-cut filtered version. However, applying the smooth cut has also the advantage of reducing the artefacts near the surface as presented in the previous paragraph (1).

- **Poynting-vector decomposition**.
  Another way to decompose a wavefield is by analyzing the instantaneous wave propagation directions of the wavefields using the Poynting vector [Yoon and Marfurt, 2006]. The Poynting vector measures the energy flow and computes the wave propagation direction. It is defined as:

$$Poynting \quad vector = -\nabla P \frac{\partial P}{\partial t}. \tag{3.15}$$

  The principle of this method is to evaluate the sign of the product $\nabla P \frac{\partial P}{\partial t}$. This idea could be interpreted using the expression under the high frequency approximation of the wavefield:

$$P \simeq A(\mathbf{s},\mathbf{x}) S^{(k)}(t-\tau(\mathbf{s},\mathbf{x})), \tag{3.16}$$

  where τ is the traveltime between the source and the image point. Is satisfies the Eikonal equation. $A(\mathbf{s},\mathbf{x})$ is the amplitude term, solution of the transport equation. $S(t)$ is the signature of the source. The term k acting at the source depends on the dimension (1D, 2D, 3D). For the derivation of the spatial gradient, both τ and A depends on $\mathbf{x}$. But the smooth variation of A can be neglected in front of the oscillatory kernel, yielding

$$\nabla P \simeq -A(\mathbf{s},\mathbf{x}) S^{(k+1)}(t-\tau(\mathbf{s},\mathbf{x})) \nabla\tau(\mathbf{s},\mathbf{x}), \tag{3.17}$$

  and

$$\frac{\partial P}{\partial t} = A(\mathbf{s},\mathbf{x}) S^{(k+1)}(t-\tau(\mathbf{s},\mathbf{x})). \tag{3.18}$$

  Then the product of ∇P and $\frac{\partial P}{\partial t}$ yields

$$-\nabla P \frac{\partial P}{\partial t} = A^2(\mathbf{s},\mathbf{x}) [S^{(k+1)}(t-\tau(\mathbf{s},\mathbf{x}))]^2 \nabla\tau. \tag{3.19}$$

  The sign of this product only depends on $\nabla\tau$, which indicates the direction of wavefield propagation with time.

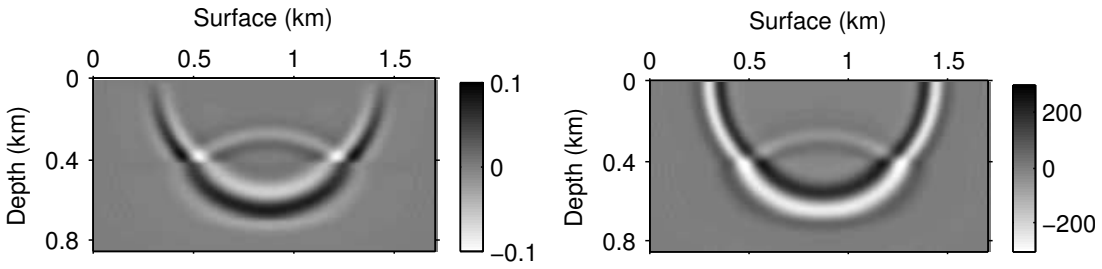


Figure 3-19: (Left) Partial derivative of the wavefield with respect to z. (Right) Partial derivative of the wavefield with the respect to t.

Therefore, the wavefield is decomposed in the vertical direction as follows:

$$\begin{aligned} P_{z+}(t,z) &= \begin{cases} P(t,z) & \text{for } \frac{\partial P}{\partial z}\frac{\partial P}{\partial t} \leq 0 \\ 0 & \text{for } \frac{\partial P}{\partial z}\frac{\partial P}{\partial t} > 0 \end{cases}, \\ P_{z-}(t,z) &= \begin{cases} 0 & \text{for } \frac{\partial P}{\partial z}\frac{\partial P}{\partial t} \leq 0 \\ P(t,z) & \text{for } \frac{\partial P}{\partial z}\frac{\partial p}{\partial t} > 0 \end{cases}. \end{aligned} \tag{3.20}$$

The decomposition of left- and right-going wavefields can be deduced in a similar way:

$$\begin{aligned} P_{x+}(t,x) &= \begin{cases} P(t,x) & \text{for } \frac{\partial P}{\partial z}\frac{\partial P}{\partial t} \leq 0 \\ 0 & \text{for } \frac{\partial P}{\partial z}\frac{\partial P}{\partial t} > 0 \end{cases}, \\ P_{x-}(t,x) &= \begin{cases} 0 & \text{for } \frac{\partial P}{\partial z}\frac{\partial P}{\partial t} \leq 0 \\ P(t,x) & \text{for } \frac{\partial P}{\partial x}\frac{\partial P}{\partial t} > 0 \end{cases}. \end{aligned} \tag{3.21}$$

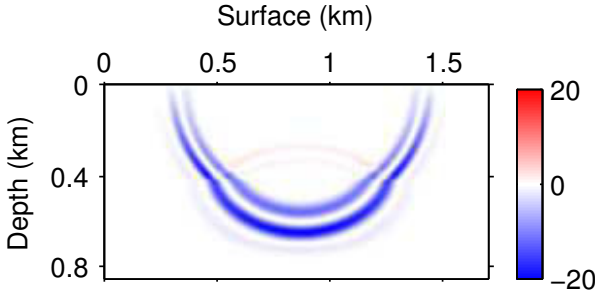


Figure 3-20: The product of the two images in Figure 3-19.

We show an example of the implementation of this method using the same wavefield as in the FT-based method.

1. We calculate the partial derivative of the wavefield P with respect to t and z, and we get $\frac{\partial P}{\partial z}$ and $\frac{\partial P}{\partial t}$. This can be done during the wavefield propagation, thus requires no extra computation. Two snapshots of $\frac{\partial P}{\partial z}$ and $\frac{\partial P}{\partial t}$ are shown in Figure 3-19.

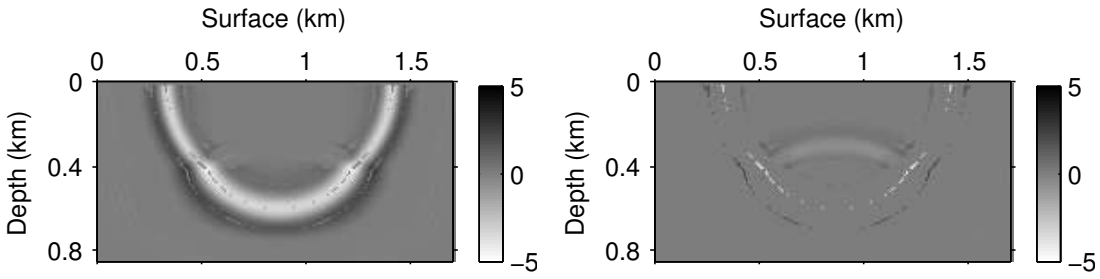


Figure 3-21: The decomposed wavefields using Poynting-vector method. P^+ (left) and P^- (right).

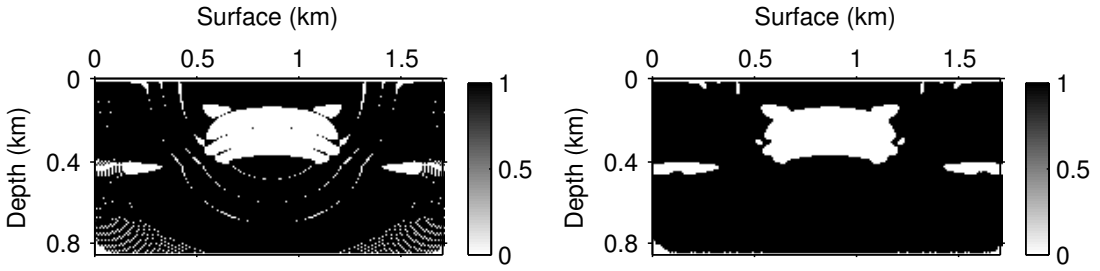


Figure 3-22: Original mask (left) and median-filtered mask (right) for the decomposition by the Poynting-vector method.

2. We then correlate these two wavefields $\frac{\partial P}{\partial z}$ and $\frac{\partial P}{\partial t}$. A snapshot of the product $\frac{\partial P}{\partial z}\,\frac{\partial P}{\partial t}$ is shown in Figure 3-20. The original wavefield is separated into down-going and up-going parts according to the sign of the product at each location. According to equation (3.20), at each location, if the product is positive, then the original wavefield at this location corresponds to the down-going part, otherwise, it corresponds to the up-going part. The decomposition result is shown in Figure 3-21. We notice that the down-going and up-going wavefields are not completely separated. Parasite residuals are present, especially in the middle of the wavefield. If we set the positive values in Figure 3-20 to 1 and negative values to 0, we get the image in Figure 3-22 on the left. In this image, points in black correspond to the locations of down-going wavefield, while points in white correspond to the up-going wavefield. This image betters shows the locations of artefacts. These artefacts are like "salt and pepper" artefacts, and they are due to the numerical approximation of the partial derivative. For the spatial positions whose partial derivatives are close to zero, the sign may be incorrect, thus yielding a wrong estimation of the mask. The points corresponding to near-zero partial derivatives are a) points at the intersections of downgoing and upgoing wavefield; b) points propagating perpendicularly to the decomposition direction, for example, horizontally propagating wavefield when the vertical decomposition is performed, which are usually near the surface; and c) points

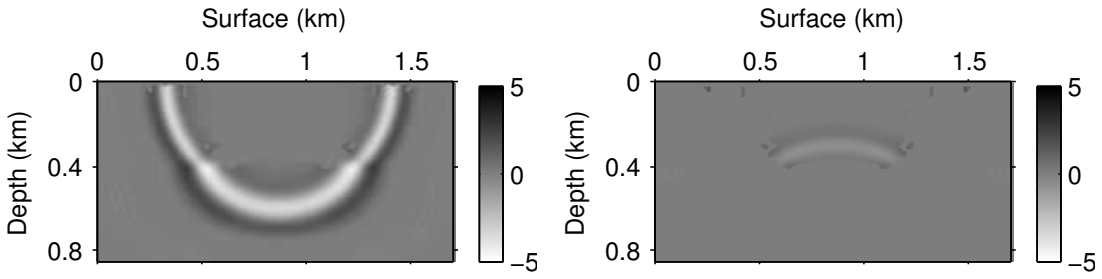


Figure 3-23: Decomposed wavefields using the mask in Figure 3-22 on the right. P^+ (left) and P^- (right).

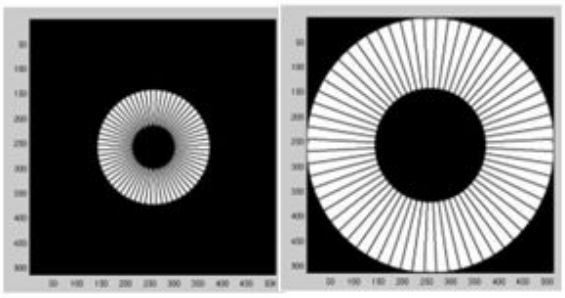


Figure 3-24: A scale example of curvelet [Chauris et al., 2006].

where P is maximal, which are along the central part of the wavelet of the wavefield. The artifacts decrease with the increasing of the spatial discretization, because the derivative value is better estimated.

3. We apply a median filter to suppress the salt and pepper artefact of the mask. The result is shown in Figure 3-22 on the right. The decomposed wavefields using the median-filtered mask are shown in Figure 3-23. These two images present less artifacts than the initial results in Figure 3-21.

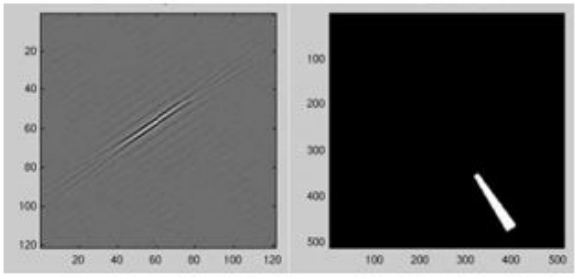


Figure 3-25: A direction example of curvelet [Chauris et al., 2006].

- **Curvelet decomposition**.
  The main limitation of the FT- and Poynting-based methods is that they can not deal with the vertical and the horizontal directions at the same time. To circumvent this

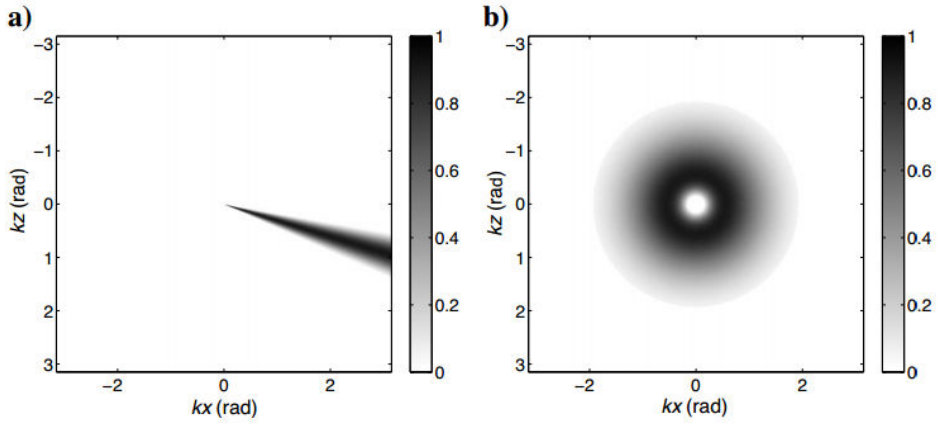


Figure 3-26: The direction (left) and scale (right) filter in the curvelet domain [Chauris and Nguyen, 2008].

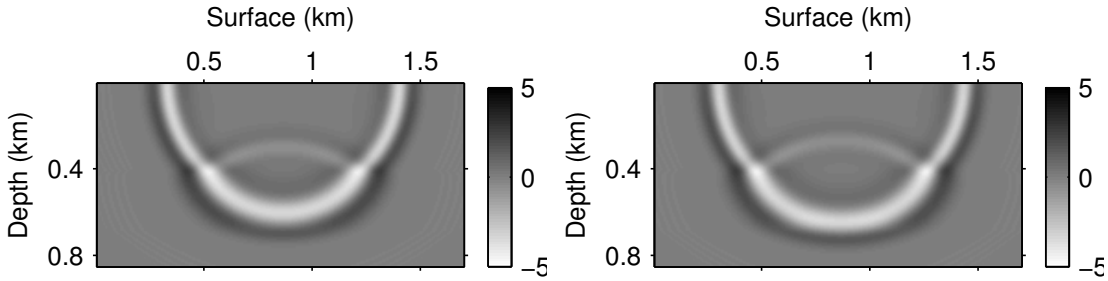


Figure 3-27: Two snapshots at different recording time for $t = 0.525$ s (left) and $t = 0.54$ s (right).

problem, we have designed a new method to decompose wavefields using curvelets in order to attenuate the artifacts that FT- or Poynting-based decomposition methods cause.

Curvelets form a basis well-adapted for representing smooth images. Curvelets are multi-directional [Candes et al., 2006, Chauris and Nguyen, 2008] and can decompose wavefields in several directions at the same time. A curvelet is characterized by three parameters: scale, direction and position [Candes et al., 2006]. A direction and scale example of curvelets are shown in Figure 3-24 and 3-25. Curvelets form a tight frame, such that a function can be decomposed as a sum of curvelets. Besides, [Candes et al., 1999] show that the curvelet representation of wave propagators is optimally sparse. For a 2D image with smooth discontinuities $f(x, z)$, the reconstructed image $f_n(x, z)$, obtained with the n first most important coefficients, approximates the original image by:

$$\|f - f_n\|_{L^2}^2 \approx \frac{(\log n)^3}{n^2}. \tag{3.22}$$

The optimal convergence rate for all the decomposition methods is $\frac{1}{n^2}$. Therefore, compared to the convergence rate for Fourier transform ($\frac{1}{\sqrt{n}}$) and for the wavelet transform ($\frac{1}{n}$), the curvelet representation is more efficient. The curvelet is widely used in geophysics for data denoising [Hennenfent and Herrmann, 2006] and data

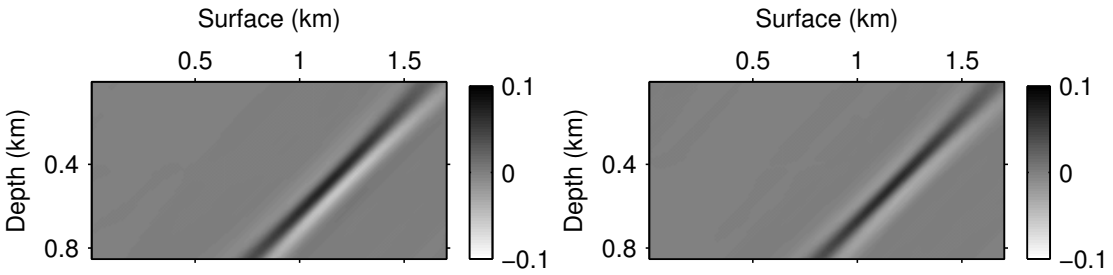


Figure 3-28: One direction after the curvelet transform for the two snapshots in Figure 3-27.

compression [Ma, 2011]. [Chauris and Nguyen, 2008] apply the demigration operator in the curvelet domain to reduce computation cost.

From a practical point of view, we implement the curvelet transform as [Chauris and Nguyen, 2008] did. First, we compute the 2D Fourier transform of the original image. Then, for all scales and all directions, we filter the transformed image by the predefined curvelet functions. Finally, we apply an inverse 2D Fourier transform. This provides the curvelet coefficients. The inverse curvelet transform consists of applying a forward 2D Fourier transform for each scale and direction, filtering once more, adding the contributions for all scales and all directions, and performing an inverse Fourier transform to obtain the original image. In this thesis, we consider all scales at the same time, and the numbers of directions are 64 for all the scales. The number of directions could also be reduced by two at every two scales to satisfy the parabolic scaling [Chauris and Nguyen, 2008]. To define curvelet filters, we use polar coordinates instead of Cartesian coordinates. This naturally leads to consider rotations instead of shears [Candes and Demanet, 2004] and avoids special considerations for handling edge effects. The design of the filters is based on the approach proposed by [Simoncelli et al., 1992].

Each filter $F_{d,l}(\theta, r)$ is defined in the polar coordinate system (θ, r) and constructed as the product of two filters, the direction filter $G_d(\theta)$ and the scale filter $H_l(r)$ (Figure 3-26). Suppose we have 2^{N+1} different directions. The directional part $G_d(\theta)$ is defined as follows:

$$G_d(\theta) = \begin{cases} \frac{2}{\sqrt{5}} cos^3[N\theta - \frac{Pi}{4}(d-1)] & \text{for } \theta \in [-\frac{Pi}{2N} + \frac{Pi}{4N}(d-1), \frac{Pi}{2N} + \frac{Pi}{4N}(d-1)] \\ 0 & \text{otherwise} \end{cases} . \tag{3.23}$$

The factor $\frac{2}{\sqrt{5}}$ is a normalization factor, and one could check that the sum of all squared filters equal to 1. This ensures a perfect reconstruction scheme. The radial filters $H_l(r)$ are defined in a similar way, with a change of variable from r to $\log r$, as described in [Simoncelli et al., 1992].

Let $P_1(\mathbf{x})$ and $P_2(\mathbf{x})$ be two successive snapshots, for different 2D spatial positions $\mathbf{x} = (x, z)$, as shown in Figure 3-27. The objective is to decompose the wavefields in two different parts: a part for which the wavefield expands from the source, and the remaining contracting part. More precisely,

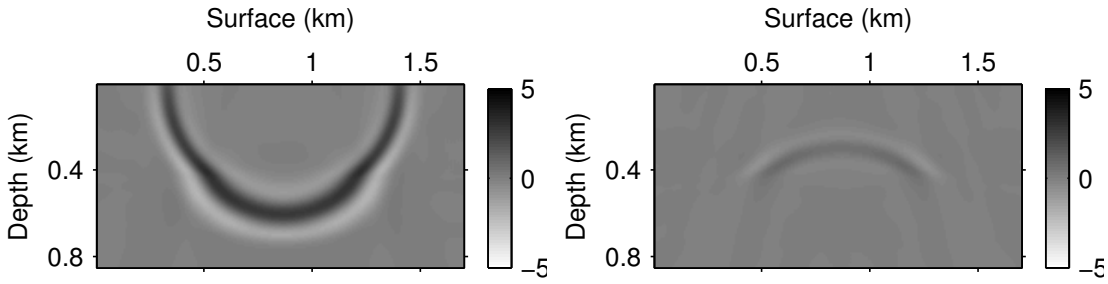


Figure 3-29: Decomposed wavefields using the curvelet method. P^+ (left) and P^- (right).

1. We first decompose the two images $P_1(\mathbf{x})$ and $P_2(\mathbf{x})$ in the curvelet domain for different directions. Let now consider two curvelets $c_d^1(\mathbf{x})$ and $c_d^2(\mathbf{x})$, for the selected direction d.

2. We decompose the curvelet domain in sub-domains. For each sub-domain Ω_i, the reconstructed images in the space domain for direction d are $I_1(\mathbf{x})$ and $I_2(\mathbf{x})$ (Figure 3-28), obtained by

$$I_1(\mathbf{x}) = \int_{\Omega_i} \mathrm{d}\mathbf{y}\, c_d^1(\mathbf{y}) F_d^*(\mathbf{x}-\mathbf{y}), \tag{3.24}$$

$$I_2(\mathbf{x}) = \int_{\Omega_i} \mathrm{d}\mathbf{y}\, c_d^2(\mathbf{y}) F_d^*(\mathbf{x}-\mathbf{y}), \tag{3.25}$$

   where F_d is the filter defining the inverse curvelet transform for direction d. This is simply a shifted version of the filter, with a weight equal to the curvelet coefficient. The star $*$ indicates that we consider the complex conjugate because the c_d coefficients and possibly the filters are indeed complex.

3. We compute a correlation function $\xi_d(\lambda)$ that computes the correlation between I_1 and I_2. The local correlation is evaluated along the d direction, and characterized by the normal n_d as follows

$$\xi_d(\lambda) = \int \mathrm{d}\mathbf{x}\, I_1(\mathbf{x}) I_2^*(\mathbf{x}+\lambda\mathbf{n}_d), \tag{3.26}$$

   where λ is a scalar value, denoting the shift along the d direction. The meaning of ξ_d is the following: if the maximum of ξ_d is obtained for a positive λ value, then the wavefield is considered as in expansion, otherwise in contraction.

4. We repeat steps (1) to (3) for each direction and each sub-domain, and we sum all the contributions. The final result is shown in Figure 3-29.

In the following, we will show that instead of reconstructing images from curvelet coefficients and then correlating the two images in the space domain, as done in equation (3.26), we could also achieve this goal in the curvelet domain directly. The correlation function only depends on the curvelet coefficients.

After step (1), in the curvelet domain, let select the coefficients belonging to a particular spatial zone Ω_i. This is the zone where we want to evaluate the behavior

of the wavefields. For the evaluation of the local correlation, we consider coefficients in Ω_i. For the correlation, the summation is performed over the full space. We replace I_1 and I_2 in equation (3.26) by their expression in equation (3.24) and 3.25 and interchange the order of integration, leading to

$$\xi_d(\lambda) = \int_{\Omega_i} \mathrm{d}\mathbf{y}\, c_d^1(\mathbf{y}) \int_{\Omega_i} \mathrm{d}\mathbf{z}\, c_d^{2*}(\mathbf{z}) \int \mathrm{d}\mathbf{x}\, F_d^*(\mathbf{x}-\mathbf{y}) F_d(\mathbf{x}+\lambda\mathbf{n}_d-\mathbf{z}). \tag{3.27}$$

In practice, the F_d filters are very oscillatory in the spatial domain. The main contribution is thus obtained for $\mathbf{z} = \mathbf{y} + \lambda\mathbf{n}_d$. We thus have

$$\begin{aligned} \xi_d(\lambda) &\propto \int_{\Omega_i} \mathrm{d}\mathbf{y}\, c_d^1(\mathbf{y}) \int_{\Omega_i} \mathrm{d}\mathbf{z}\, c_d^{2*}(\mathbf{z}) \delta(\mathbf{z}-\mathbf{y}+\lambda\mathbf{n}_d) && (3.28) \\ &= \int_{\Omega_i} \mathrm{d}\mathbf{y}\, c_d^1(\mathbf{y}) c_d^{2*}(\mathbf{y}+\lambda\mathbf{n_d}). && (3.29) \end{aligned}$$

We observe that the original correlation in the space domain (equation (3.26)) is equivalent to the correlation in the curvelet domain (equation (3.29)). Compared to the direct implementation in the space domain, we do not have to reconstruct the images for all directions nor for all sub-domains. All the computations are performed with curvelet coefficients. Note that by selecting coefficients in the curvelet domain, the reconstructed images would have natural tapering on the edges.

In summary, the first step consists of decomposing the two wavefields in the curvelet domain, then of performing the local correlation. As before, depending on if the maximum value for ξ_d is obtained for a positive or negative λ, the selected coefficients are stored in two different arrays. After having investigated all spatial positions and all directions, the inverse curvelet transform is applied to the two intermediate arrays, providing two reconstructed images.

For the curvelet-based decomposition, we can observe some artifacts of residual directional element as the curvelet is pseudo localized in the space domain. The artifacts get reduced when the number of directions considered in the curvelet decomposition increase.

We investigate a more complex wavefield propagated in the Marmousi model. The decomposed results for the snapshot in Figure 3-30 using the Fourier transform method, the Poynting vector method and the curvelet method are shown in Figure 3-31. For this wavefield, the drawbacks of the Fourier transform method and the Poynting vector method are obvious. We note that the Fourier transform method filters out low frequency data and wavefields propagating horizontally near the surface, and the Poynting vector method suffers from strong artefacts at intersections of different directions. The curvelet method yields the best result of the three.

From the point of view of computational cost, the Poynting vector method is the most efficient one, as the partial derivative wavefields could be obtained during the wave propagation and the computational complexity is $O(nz * nx)$, where nz and nx are the model dimensions.

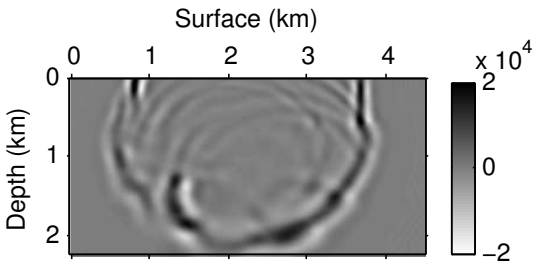


Figure 3-30: One snapshot of the wavefield propagated in the Marmousi model.

The Fourier transform method is more expensive than the Poynting vector method as we have to compute the Fourier transform of the wavefield in the $t-z$ domain for each horizontal coordinate x. If we use the FFT algorithm, the computational complexity is $O(nx{*}nz{*}nt{*}log(nz{*}nt))$, where nz and nx are the model dimensions, and nt is the number of time samples.

The curvelet method is also expensive as we have to compute the curvelet transform for each snapshot of the wavefield. Recall that the curvelet transform consists of one Fourier transform, one filtering for each direction and one inverse Fourier transform for each direction. The computational complexity is then $O(nt{*}nz{*}nx{*}log(nz{*}nx))$, where nz and nx are the model dimensions, and nt is the number of time samples.

Note that the Poynting vector method is the most efficient one. The difference of computational complexity between the Fourier transform method and the curvelet method is not obvious and should be evaluated case by case. In general, the decomposition based on the curvelet transform is more expensive than the decomposition based on the Fourier transform as the curvelet transform is about 10 times more expensive than the Fourier transform.

From the point of view of memory, for the Fourier transform method, we have to store the whole wavefield on the hard disk. For the Poynting method, as the partial derivative wavefields can be obtained during the wave propagation, we do not have to keep in memory the whole wavefield. For the curvelet transform method, at each time, we compare two neighbouring snapshots, thus the decomposition can be done during the wave propagation and we do not need to store the whole wavefield. Furthermore, since the decomposition is carried out for each snapshot separately, the curvelet decomposition method is more easily parallelizable. Further investigations on the curvelet implementation are needed before being able to apply it to large datasets.

For the tests shown in Chapters 4 and 5, we use the 2D Fourier transform decomposition method for computational cost and accuracy considerations. At the time these tests were performed, the correlation was done in the space domain but not in the curvelet domain, and the method was even more expensive. For this reason, the 2D Fourier domain was preferred.

For 3D applications, the Poynting vector method may be the best choice, as the need of the Fourier transform method for storing the whole 3D wavefield is difficult to meet. For the curvelet method, the significant computational cost of the 3D

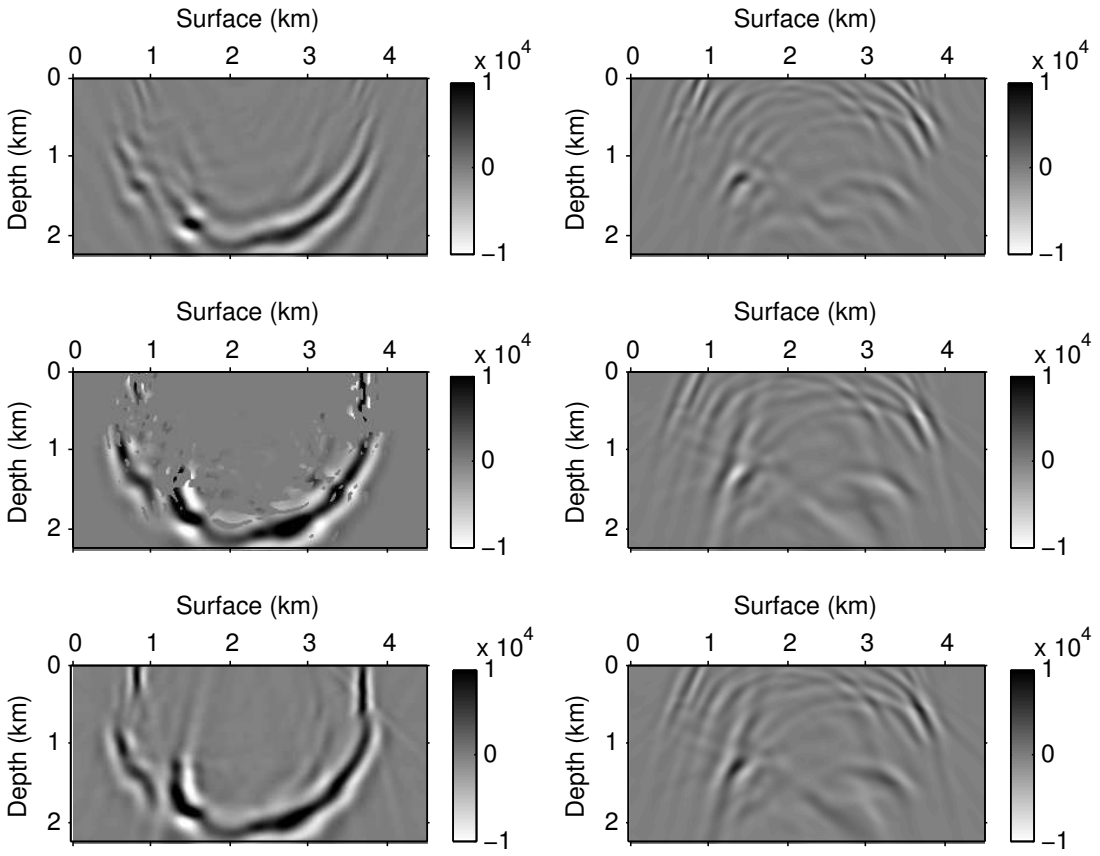


Figure 3-31: Decomposed wavefields using the 2D FT method (first panel), Poynting method (second panel) and curvelet method (third panel). P^+ (left) and P^- (right).

curvelet transform makes this method difficult to be widely used for large dataset applications.

3.6 Conclusion

In this chapter, we propose a reflection-based waveform inversion method. The principle of the proposed method is to decompose the gradient of FWI into a long-wavelength part and a short-wavelength part. The inversion is performed in an alternating fashion between these two parts. This method allows to automatically separate the long and short wavelengths of the model without having a-priori knowledge of the model. The gradient decomposition is achieved by the wavefield decomposition into their one-way components. Different wavefield decomposition methods, especially those based on two-way wave equation, are presented. The 2D FT-based wavefield decomposition method will be used for the 2D synthetic model in Chapter 4 and for the real data set in Chapter 5.

Chapter 4

Inversion of synthetic data

Contents

Résumé du chapitre 4

Dans ce chapitre, j'applique l'inversion des formes d'ondes (Full Waveform Inversion, FWI) et l'inversion basée sur la décomposition des champs d'ondes (Decomposition-based waveform inversion, DWI) sur des modèles synthétiques 1D et 2D. Rappelons que DWI consiste à décomposer le gradient en une partie de grandes longueurs d'ondes et une partie de courtes longueurs d'ondes après la décomposition des champs d'ondes. L'inversion est effectuée d'une manière alternée entre ces deux parties. Pour la modélisation numérique, nous utilisons un schéma aux différences finies d'ordre 8 en espace et d'ordre 4 en temps. Des conditions aux limites absorbantes sont imposées aux bords du modèle pour simuler un milieu infini. Nous utilisons la méthode d'optimisation de quasi-Newton, plus précisément la L-BFGS avec 5 itérations conservées en mémoire.

D'abord, nous testons FWI et DWI sur un modèle de couche 1D. Pour ce test, la source contient les basses fréquences, mais comme il y un grand écart entre le modèle exact et le modèle initial, la FWI classique n'arrive pas à retrouver les basses fréquences du modèle.

Ensuite, nous étudions un modèle synthèse 2D et comparons les résultats de l'inversion avec FWI, FWI multi-échelle et DWI. Nous faisons quatre tests avec deux choix de la source et deux choix du modèle de vitesse initial. Pour les deux modèles de vitesse initiale, l'un est proche, et l'autre est loin du modèle de vitesse exacte pour évaluer les résultats de l'inversion. Celui proche du modèle exact est un modèle de gradient constant, bien qu'il ne soit pas si proche du modèle exact, car il n'y a pas d'information a priori dans ce modèle. Celui loin du modèle exact est un modèle homogène. Pour la source, nous utilisons une source sans les basses fréquences pour être plus réaliste. Dans ce cas, la fonction objectif de FWI est moins linéaire, et l'inversion a plus de mal à converger. Cependant, nous avons également testé FWI en utilisant une source avec les basses fréquences comme une référence. La source qui contient les basses fréquences est une ondelette de Ricker avec une fréquence centrale de 8 Hz. Celle sans les basses fréquences est la même ondelette Ricker, mais les fréquences au dessous de 5 Hz sont enlevées. Notons que la densité reste constante pendant l'inversion. A travers des quatre tests, nous observons que la FWI multi-échelle est plus robuste que la FWI classique, et la DWI est plus robuste que la FWI classique et que la FWI multi-échelle face à l'absence de basses fréquences dans la source et aux erreurs importantes dans le modèle initial.

Nous avons également analysé l'influence de plusieurs paramètres clés pendant l'inversion, comme l'application de l'inversion itérative non-linéaire dans l'étape de mise à jour des courtes longueurs d'ondes, et en utilisant les données de hautes fréquences et zéro offset dans cette étape. Grâce à cette analyse, nous pouvons conclure que lorsque ces paramètres clés sont tous proprement appliqués à l'inversion, la mise à jour des grandes longueurs d'ondes est plus proche de la perturbation exacte du modèle que dans les cas où les paramètres sont partiellement appliqués. Du point de vue du coût de calcul, DWI est très cher, car l'inversion itérative est effectuée à chaque étape de la mise à jour des courtes longueurs d'ondes pour ajuster les données au lieu de juste imager les réflecteurs. Toutefois, cette étape itérative est importante, car elle facilite la mise à jour des grandes longueurs d'ondes du modèle. Cette conclusion est compatible avec l'importance d'introduire la migration itérative dans l'analyse de vitesse par migration (Migration Velocity Analysis, MVA) [Symes, 2015, Lameloise et al., 2015].

4.1 Introduction

In this chapter, I apply the classical FWI and the Decomposition-based Waveform Inversion (DWI) on 1D and 2D synthetic models. Recall that the decomposition-based waveform inversion consists of decomposing the gradient of FWI into a long-wavelength part and a short-wavelength part. The inversion is performed in an alternating fashion between the short-wavelength update and the long-wavelength update. For the numerical modeling, we use a finite-difference scheme with 8^{th} order in space and 4^{th} order in time. Absorbing boundary conditions are imposed at the borders of the model to simulate an infinite medium. We use the quasi-Newton optimization method, more precisely the L-BFGS with 5 iterations kept in memory.

We first test FWI and DWI on a 1D layer model. Then we investigate a 2D synthetic model and compare the inversion results of FWI, multiscale FWI and DWI. We also analyse the influence of key inversion parameters on the inversion result.

4.2 1D layer model with low frequency

We use the simple 1D layer model that has been analyzed in Chapter 3. Figure 4-1 shows the true velocity model. The model dimensions are 1.7 km large by 0.85 km deep. The model consists of a strong velocity contrast, the velocities for the top layer, middle layer and bottom layer are 2 km/s, 3 km/s, 2 km/s, respectively. The initial model is a homogeneous model at 2 km/s, thus there is no a priori information in the initial model except the velocity at the surface. A Ricker wavelet with a central frequency of 12 Hz and temporal sampling of 1.5 ms is used as the source function. The sources and receivers are evenly distributed at the surface, with the source spacing being 17 m and receivers spacing being 8.5 m.

Figure 4-2 on the top displays the result after running 80 iterations of conventional FWI with L-BFGS scheme. All offsets available at the surface (up to 1.7 km) are used. It fails into a local minimum when starting from this homogeneous initial model. We note that it does not recover the long wavelengths of the model when all frequencies are inverted at the same time. Figure 4-2 in the second panel displays the final result of DWI for 20 global iterations, and for each global iteration associated to the long-wavelength update, we run 20 iterations for the short-wavelength update using zero-offset data. For the long-wavelength update, we use a mean filter of size 9x9 grid points to remove the potential residual short-wavelength components. In this step, we use all offsets available at the surface. Figure 4-2 in the third panel displays the result after 20 iterations of FWI with the result in the second panel as the initial model. We observe that DWI can better invert the long-wavelength components of the middle layer than the conventional FWI. Figure 4-2 on the bottom is the comparison of the vertical profile taken at x = 0.85 km. Compared to the conventional FWI, the decomposition-based waveform inversion better recovers the long-wavelength part of the middle layer, and as a result the bottom of the middle layer is better positioned.

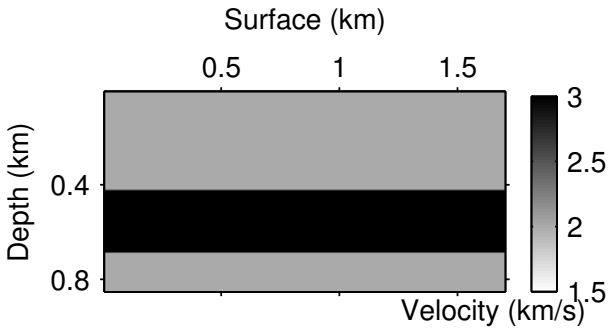


Figure 4-1: 1D layer exact velocity model.

4.3 2D synthetic model

After the test on the simple 1D layer model, we investigate a more complex 2D synthetic model. The exact velocity model is shown in Figure 4-3. This model is inspired from [Perrone, 2013]. The dimensions of the model are 3.12 km by 1.212 km. The model consists of several layers and a velocity anomaly in the central part. Sources and receivers are evenly distributed at the surface, with source spacing being 48 m and receiver spacing being 12 m. The maximum offset we use is 1.2 km in order to mainly simulate reflected events. This maximum offset is about the same dimension as the model depth. The density is considered constant with $\rho = 1000$ kg/m^3.

For the initial velocity model, we consider two models, one that is close, and the other that is far from the exact velocity model to evaluate convergence results in FWI and DWI. The one close to the exact model is a gradient-constant model (Figure 4-4 on the top), however it is not so close to the exact model, as there is no a prior information in the initial model. The one far from the exact model is a homogeneous model at 1.5 km/s (Figure 4-4 on the bottom). For the source wavelet, we use a source without low frequencies to be more realistic. In this case, the objective function of FWI is more nonlinear for data without low frequencies, and the inversion is more difficult to converge. However, we also test FWI using source wavelet with low frequencies as a reference. The source wavelet containing low frequencies is a Ricker wavelet with central frequency of 8 Hz (black line in Figure 4-5). The one without low frequencies is the same Ricker wavelet, but low frequencies lesser than 5 Hz are filtered (red line in Figure 4-5). The source wavelet is filtered from the initial Ricker wavelet with a high-pass Butterworth filter of order 9 and with a cut frequency of 8 Hz . With these two initial models and two source wavelets we test the four combinations (Table 4.1). We start from the optimal case: good initial model and source wavelet with low frequencies, and gradually degrade the inversion conditions.

N°	Frequency content	Initial model
Test1	with low frequency	gradient-constant model
Test2	with low frequency	homogeneous model
Test3	without low frequency	gradient-constant model
Test4	without low frequency	homogeneous model

Table 4.1: List of tests for the 2D model.

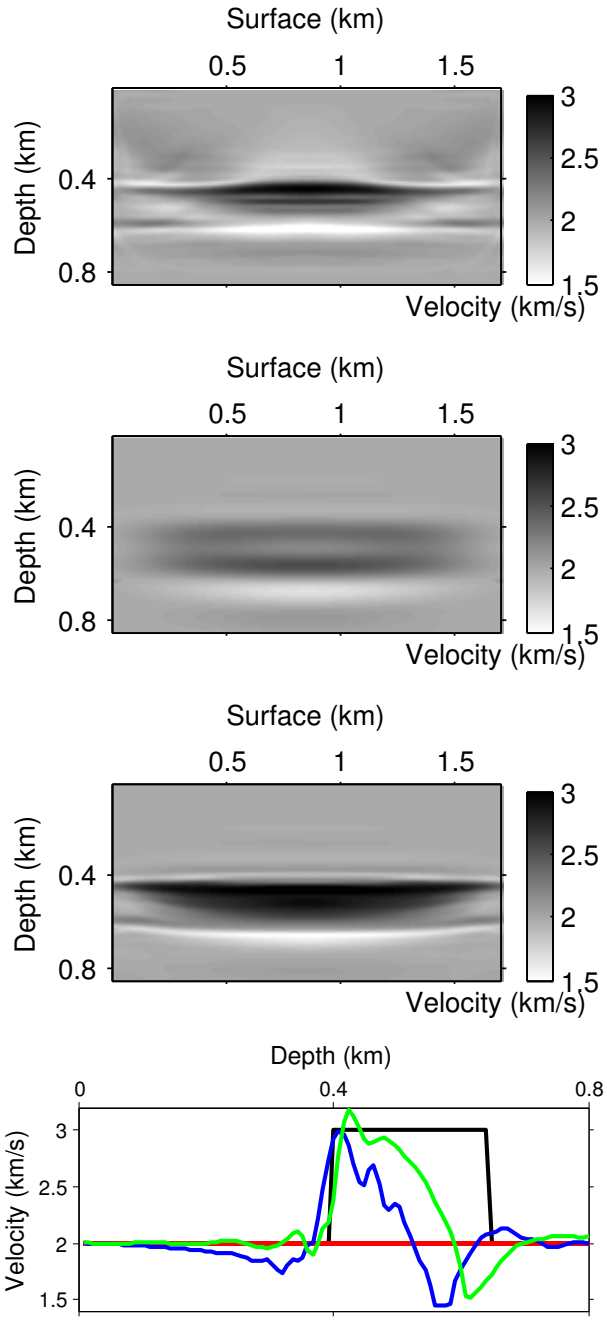


Figure 4-2: Final inversion results using a homogeneous initial model for FWI (first panel), long-wavelength model obtained with DWI (second panel), and FWI result starting from model in the second panel (third panel). Velocity profiles at 0.85 km (fourth panel) for the exact model (black line), initial model (red line), FWI (blue line) and DWI + FWI (green line).

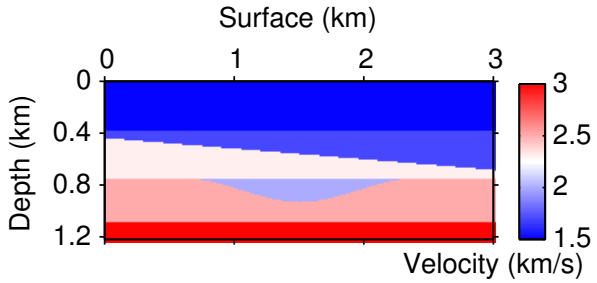


Figure 4-3: 2D exact velocity model [Perrone, 2013].

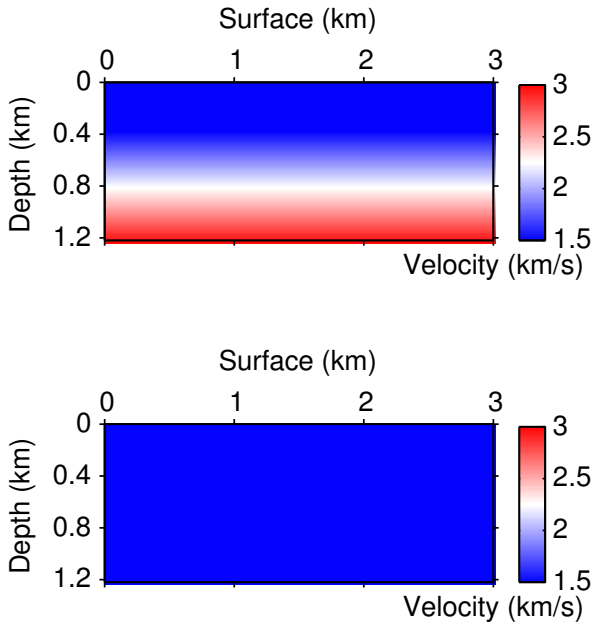


Figure 4-4: Initial velocity models considered for 2D model in Figure 4-3. Gradient-constant model (top) and homogeneous model (bottom).

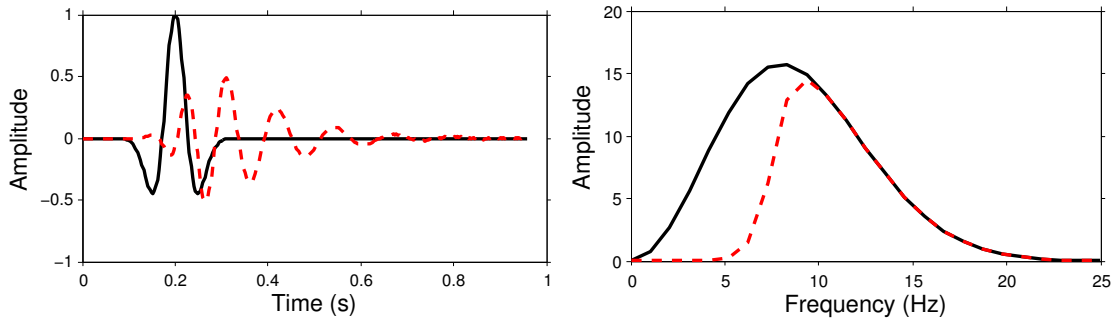


Figure 4-5: Source wavelet in the time domain (left) and in the frequency domain (right). The black lines represent the initial Ricker wavelet with central frequency of 8 Hz and the red lines represent the wavelet after the high-pass Butterworth filtering with cut frequency of 8 Hz.

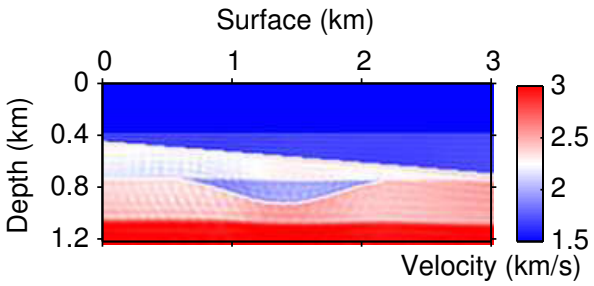


Figure 4-6: Test 1. FWI result after 100 iterations.

4.3.1 Test 1: inversion with low frequency and good initial model

In the first test, we apply FWI with the gradient-constant initial model shown in Figure 4-4 on the top. A Ricker wavelet with central frequency of 8 Hz is used as the source function. In this case, the source wavelet and the initial model are favorable to the success of FWI. The result of FWI after 100 iterations is shown in Figure 4-6 and 4-7. The comparison of the exact data, the initial data and final data is displayed in Figure 4-8. We observe that the conventional FWI succeeds in this case. The result is considered as a reference for the inversion without low frequencies.

4.3.2 Test 2: inversion with low frequency and poor initial model

In the second test, we apply FWI starting from the homogeneous model shown in Figure 4-4 on the bottom. The source wavelet remains the same as in the first test. In this test, as the initial model is very far from the true model, the conventional FWI fails to converge to the exact model. The final FWI result is shown in Figure 4-9 and Figure 4-10, and is considered as a reference for the fourth test with a poor initial model and without low frequencies. The comparison of the exact data, the initial data and final data is displayed in Figure 4-11. We note that for this initial model FWI falls into a local minimum.

4.3.3 Test 3: inversion without low frequency and good initial model

In the third test, the initial model is the same as in the first test (Figure 4-4 on the top). But low frequencies are filtered in the original Ricker wavelet to be more realistic (Figure 4-5).

FWI

Firstly, we run FWI on this model. The FWI result after 100 iterations is shown in Figure 4-12. Compared to the first test, as low frequencies are missing from the source wavelet, the problem becomes more nonlinear, and the FWI yields a poorer result, especially for the long-wavelength part in the central anomaly zone. In order to retrieve the long-wavelength components of the model, multi-scale FWI is widely used [Bunks et al., 1995]. We perform the multi-scale inversion to see if it could improve the inversion result.

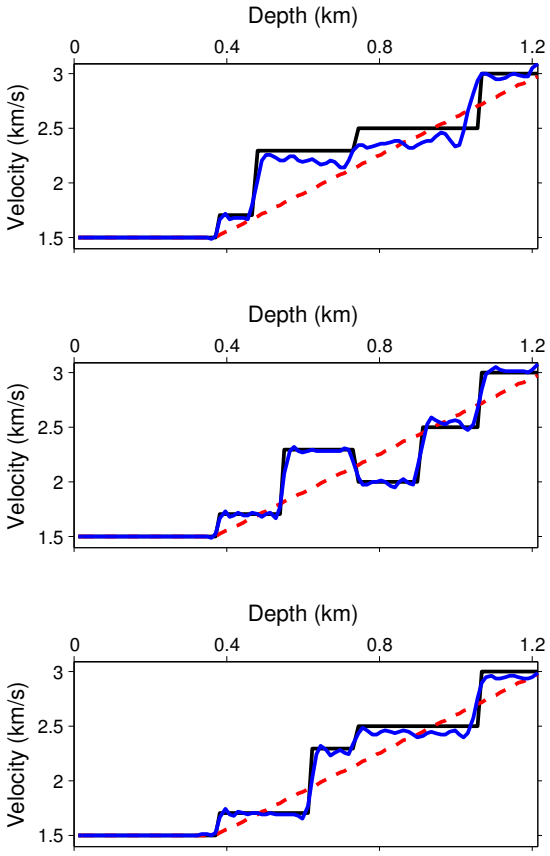


Figure 4-7: Test 1. 1D vertical profiles of exact (black), initial (red) and FWI model (blue). From top to bottom: $x = 0.4$, 1, 1.6 km.

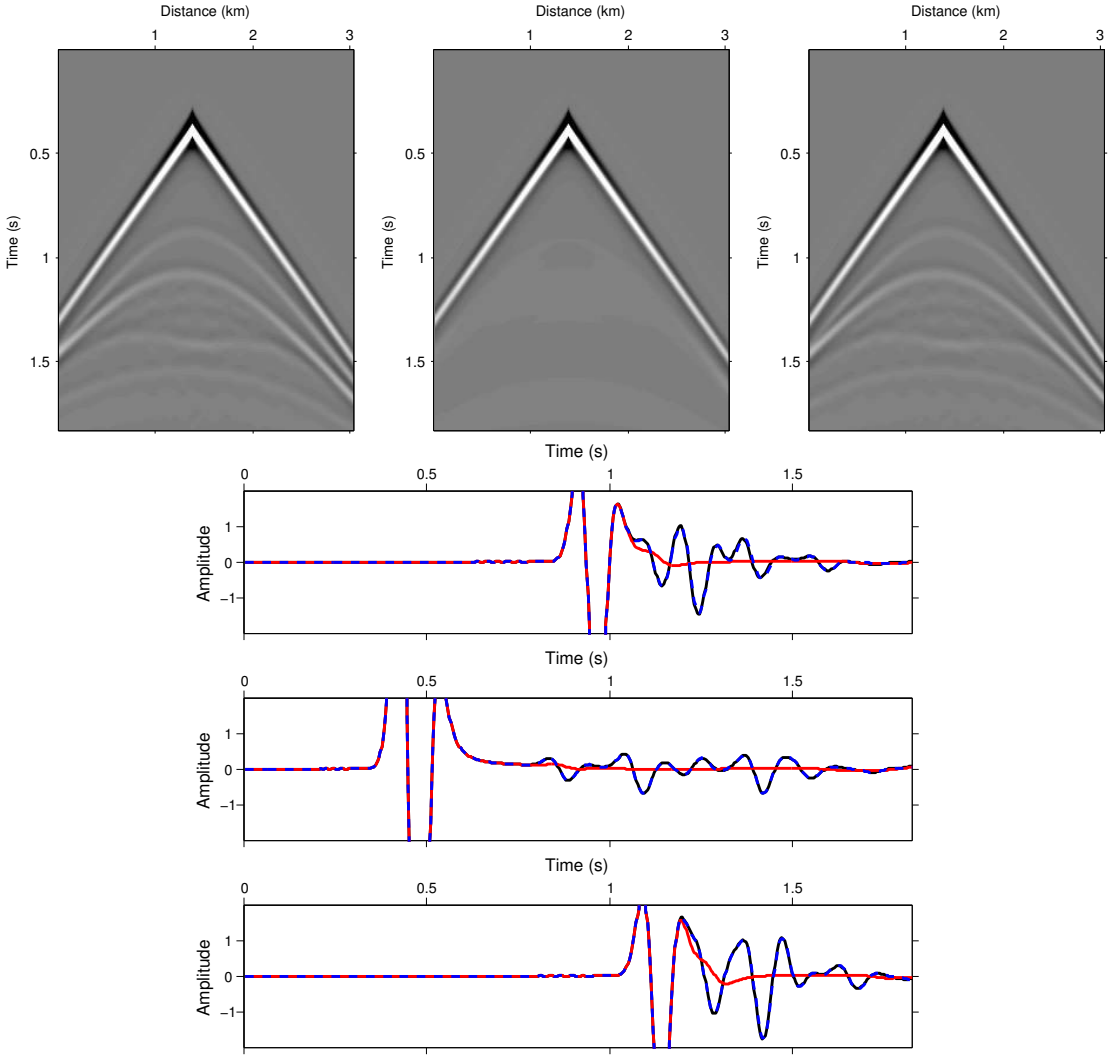


Figure 4-8: Test 1. Top panels: shot gathers for observed data (left), initial data (middle) and final data obtained with FWI (right). Bottom panels: observed (black), initial (red) and modeled (blue) seismic traces taken at three offset positions: 0.5 km (top), 1.5 km (middle) and 2.5 km (bottom). Note that the observed data and the modeled data are overlapped.

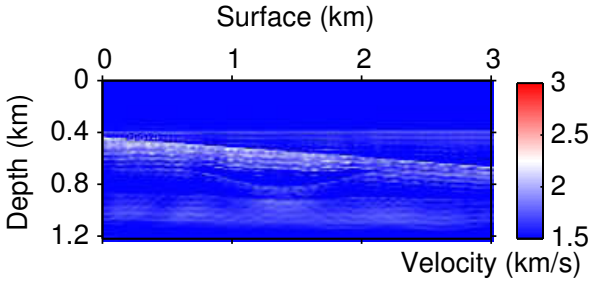


Figure 4-9: Test 2. FWI result after 100 iterations.

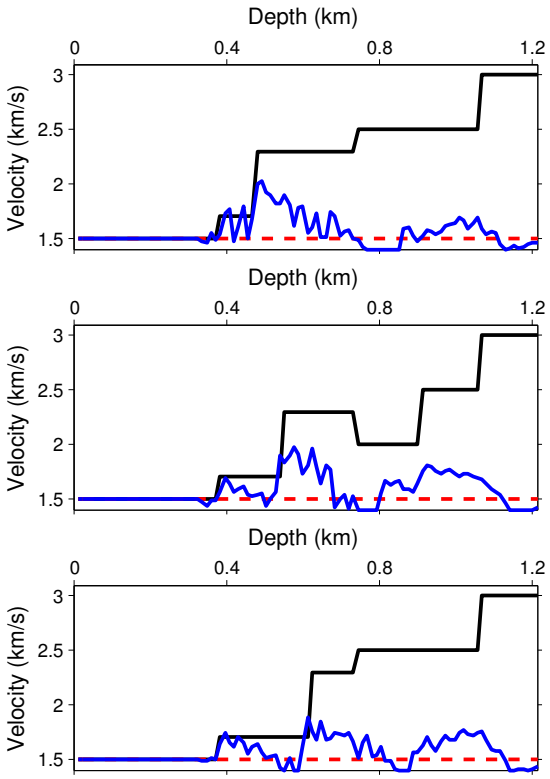


Figure 4-10: Test 2. 1D vertical profiles of exact (black), initial (red) and FWI model (blue). From top to bottom: $x = 0.4$, 1, 1.6 km.

Multi-scale FWI

The multi-scale FWI is performed here in the time domain [Bunks et al., 1995], starting from the lowest available frequency band (5 Hz) and then gradually adding higher frequency bands. Here, we perform 4 FWI steps. The source wavelets are filtered from the initial Ricker wavelet with a Butterworth filter of order 9 and in the frequency band of [8–9] Hz, [8–12] Hz, [8–15] Hz, and [8–Inf] Hz respectively (Figure 4-13). We perform 100 iterations of FWI at each step. The FWI result at each step is considered as the initial model for the next step. The results at each step are shown in Figure 4-14. Compared to the FWI result in Figure 4-12, the multi-scale FWI provides a better result (Figures 4-14 and 4-15) and allows to better retrieve the main features of the model. However, the central velocity anomaly between 0.72 km and 0.9 km in depth is not well retrieved. As a result, two reflectors at 0.72 km and 1.06 km are not positioned correctly due to the background velocity error.

DWI

We test the same model using DWI. As presented in Chapter 3, we use a nested algorithm. Within each global iteration, we perform 20 iterations of FWI using zero-offset data and quasi-Newton (L-BFGS) algorithm for the model update [Nocedal, 1980]. The long-wavelength update step is then performed using the gradient described in equation (3.4). A single iteration of inversion is performed using the gradient descent algorithm [Sirgue,

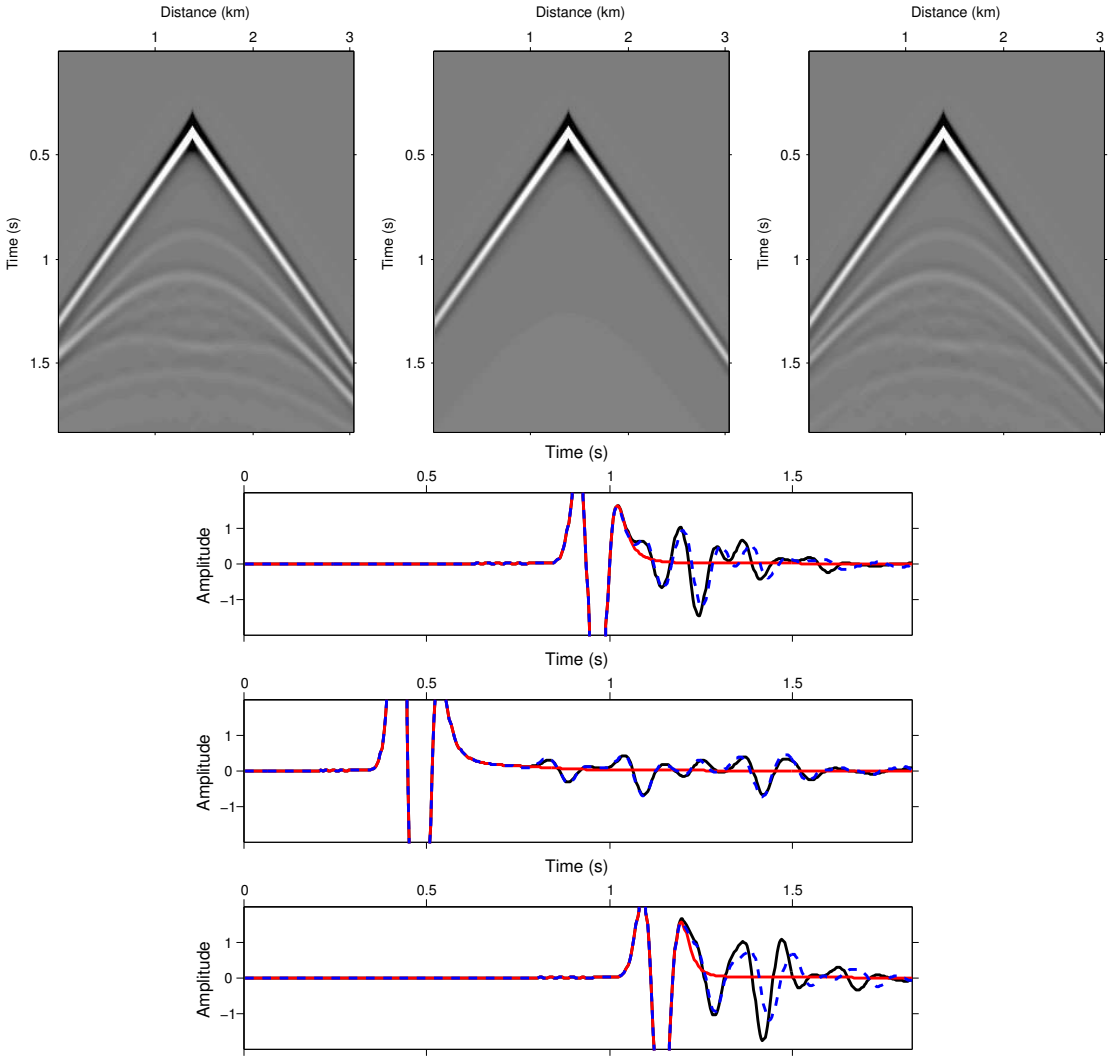


Figure 4-11: Test 2. Top panels: shot gathers for observed data (left), initial data (middle) and final data obtained with FWI (right). Bottom panels: observed (black), initial (red) and modeled (blue) seismic traces taken at three offset positions: 0.5 km (top), 1.5 km (middle) and 2.5 km (bottom).

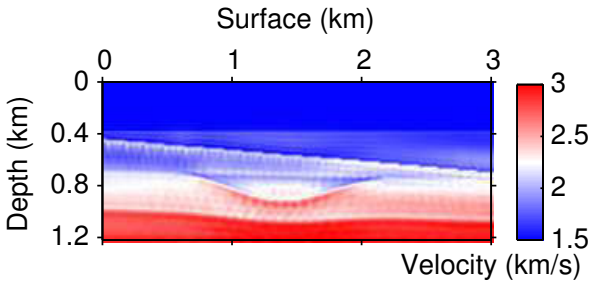


Figure 4-12: Test 3. FWI result after 100 iterations.

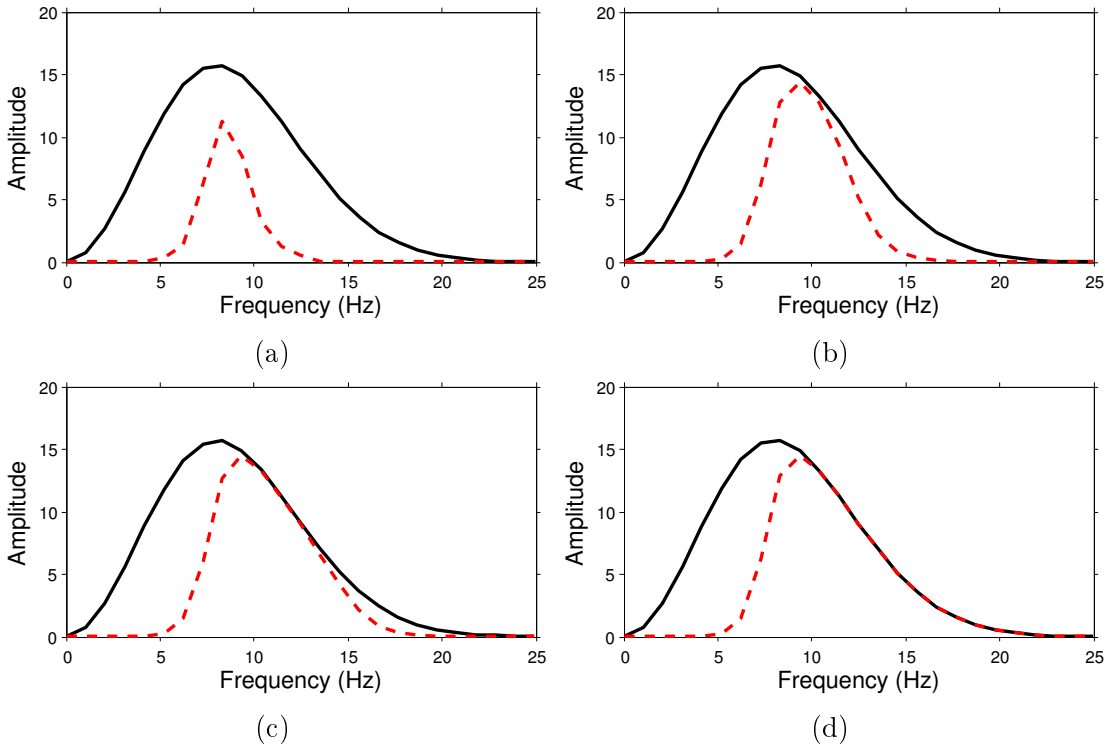


Figure 4-13: Test 3. Source bandwidths used for the four steps in multiscale FWI. The black lines represent the initial Ricker bandwidth with central frequency of 8 Hz and the red lines represent the source bandwidth after the Butterworth filter. The cut frequencies for the band-pass filter are respectively (a) 8–9 Hz, (b) 8–12 Hz and (c) 8–15 Hz. A high-pass filter with cut frequency of 8 Hz is used in (d).

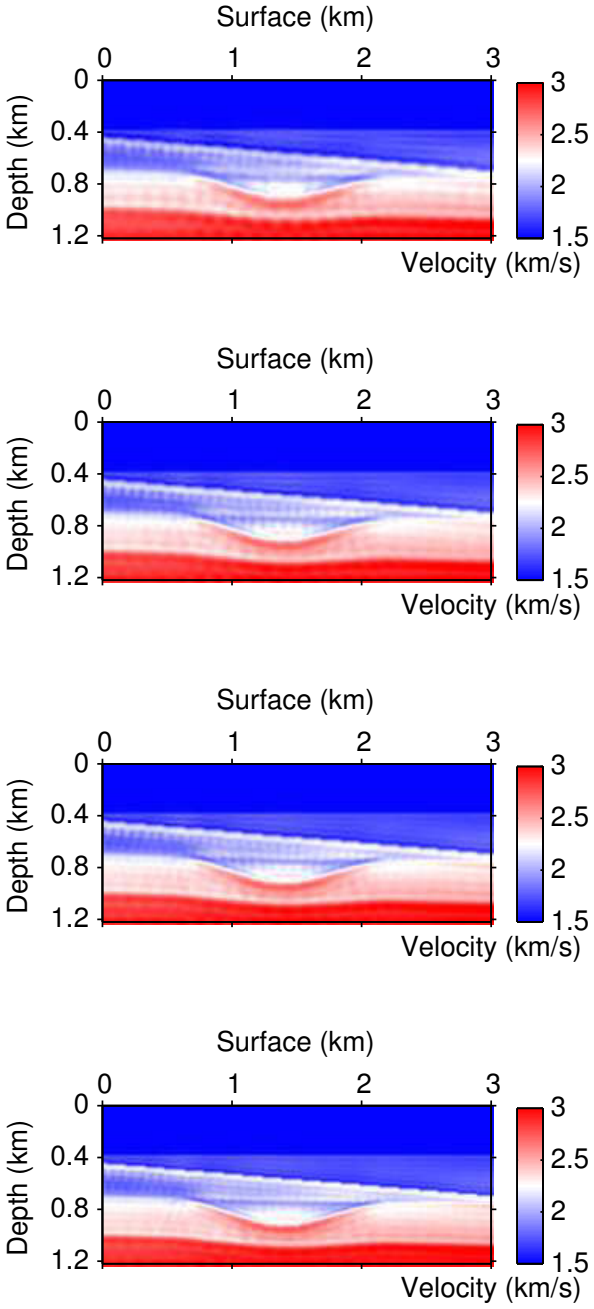


Figure 4-14: Test 3. Multiscale FWI results at each steps. From top to bottom, the source bandwidths used are respectively the one in Figure 4-13 a-d.

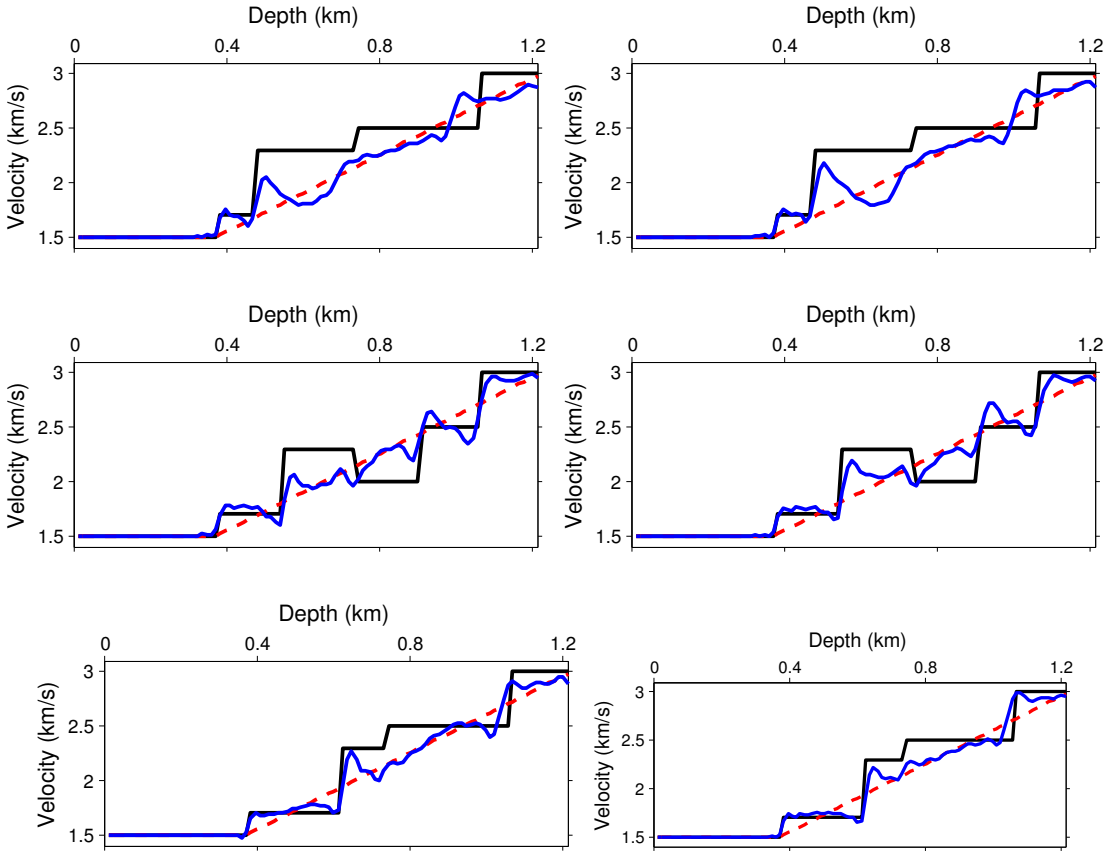


Figure 4-15: Test 3. 1D vertical profiles of exact (black line), initial (red line) and FWI model (blue line) (left), or multiscale FWI model (blue line) (right). From top to bottom: $x = 0.4$, 1, 1.6 km.

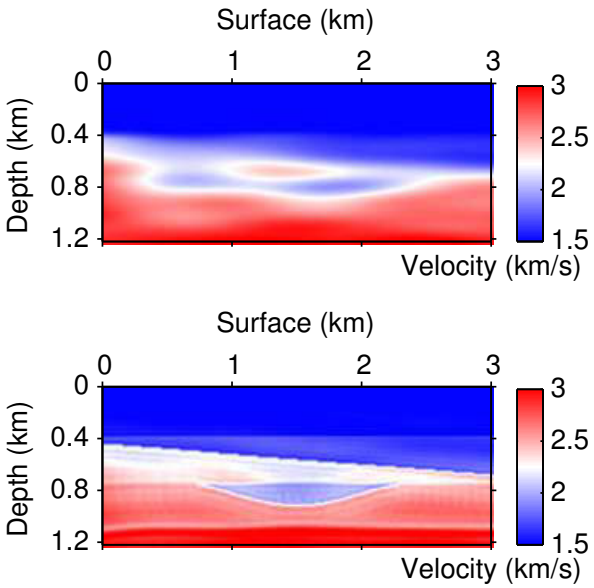


Figure 4-16: Test 3. Long-wavelength model obtained for DWI (top), and FWI result starting from model on the top (bottom).

2003], choosing α in equation (2.11) such that the product of α and the maximum value of the gradient remains a constant fixed at the first iteration. We smooth the gradient in this step to remove the residual short-wavelength components that are caused by acquisition limitations and decomposition artefacts. The smoothing is performed in the frequency domain as shown in Chapter 3. Once the background velocity is updated, a new global iteration is started.

In the long-wavelength update step, as we perform only one iteration of inversion, to compensate the geometrical spreading, we need to weight the gradient by the inverse of the Hessian to speed up convergence. The inverse of the Hessian helps to deconvolve the gradient, and to compensate for the geometrical spreading and for illumination. But the computation of the inverse of the Hessian is extremely expensive [Clément et al., 2001]. As the test model is relative simple, we use a linear preconditioning to compensate the geometrical spreading assuming that the velocity increases linearly with depth. We also tested the preconditioning presented in [Plessix and Mulder, 2004] which calculates an approximation of the inverse of the Hessian by the inverse of the norm of the wavefield. However, the linear preconditioning is sufficient in this case and provides a similar result.

The long-wavelength velocity model we obtain after 30 global iterations is shown in Figure 4-16 on the top. Starting from this model, we run 100 iterations of the conventional FWI (inverting for all frequencies at once) and the final result is shown in Figure 4-16 on the bottom. The whole procedure could be regarded as a two-step FWI workflow, consisting of the DWI followed by the conventional FWI. Compared to the multi-scale FWI (Figure 4-17 and 4-18), we can observe that the reflectors are better positioned and the central velocity anomaly, which is absent in the initial model, is successfully recovered.

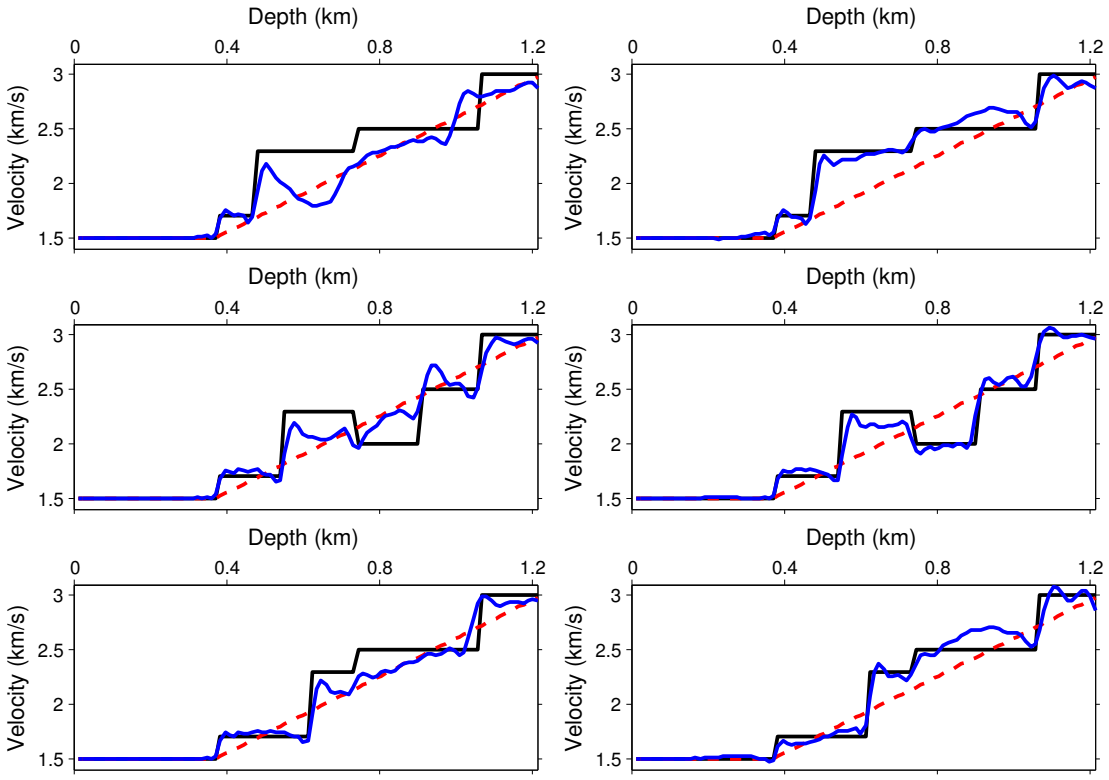


Figure 4-17: Test 3. 1D vertical profiles of exact (black line), initial (red line) and multi-scale FWI model (blue line) (left), or DWI + FWI model (blue line) (right). From top to bottom: $x = 0.4$, 1, 1.6 km.

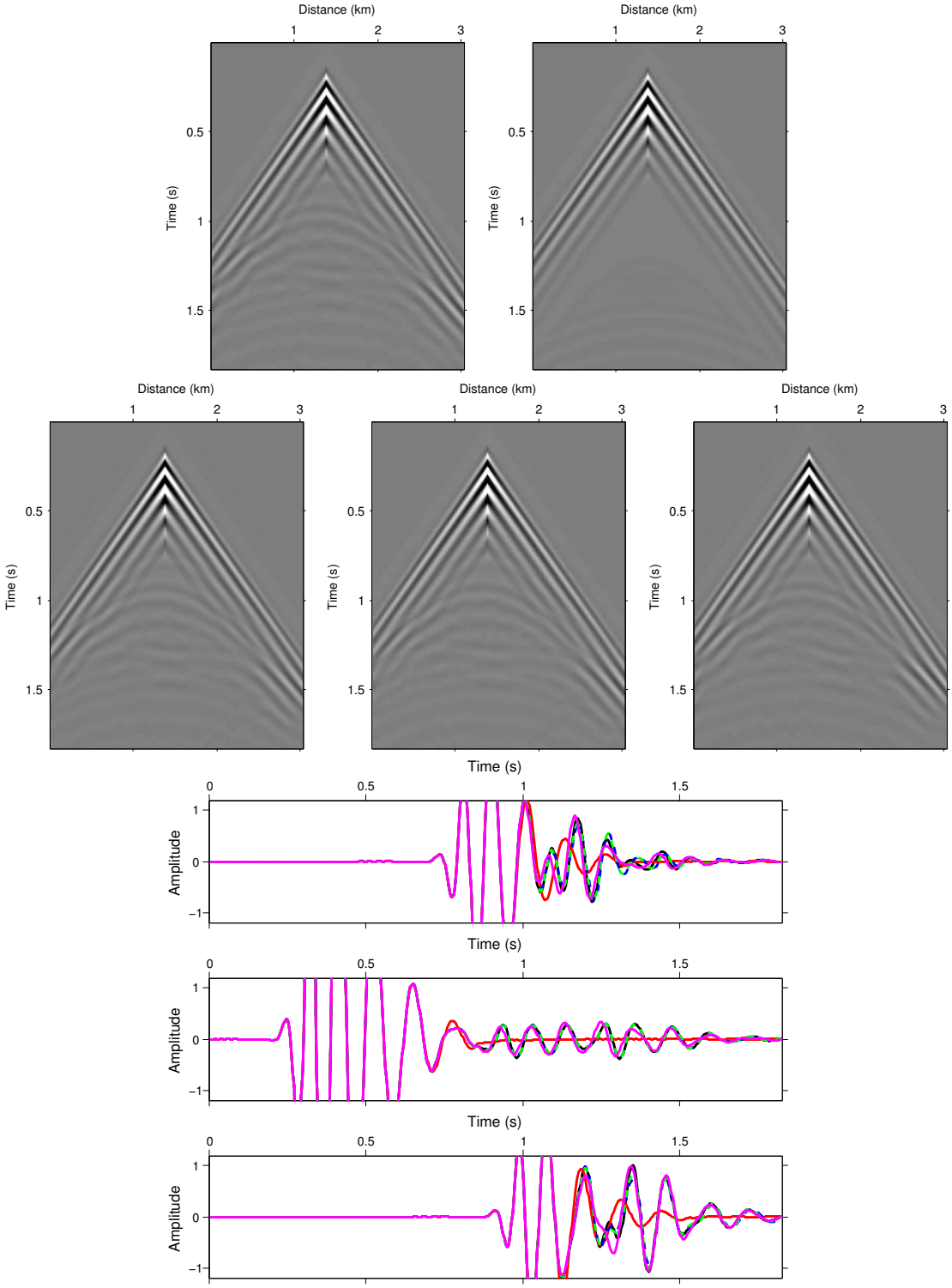


Figure 4-18: Test 3. Top panels: shot gathers for observed data (left), initial data (right). Middle panels: Final data obtained with FWI (left), multi-scale FWI (middle) and DWI (right). Bottom panels: observed (black), initial (red), FWI modeled (blue), multi-scale FWI modeled (green) and DWI modeled (magenta) seismic traces taken at three offset positions: 0.5 km (top), 1.5 km (middle) and 2.5 km (bottom).

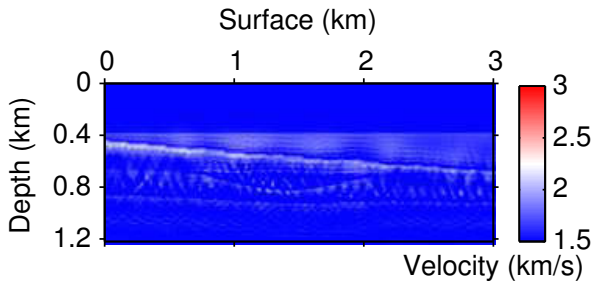


Figure 4-19: Test 4. FWI result

4.3.4 Test 4: inversion without low frequencies and poor initial model

In this test, the inversion conditions are the poorest of the four cases, as the initial model is the homogeneous model at 1500 m/s (Figure 4-4 on the bottom) and the low frequencies are filtered in the source wavelet (Figure 4-5).

FWI

Firstly, we apply the FWI on the initial model. The result after 100 iterations is shown in Figure 4-19. As low frequencies are absent in the source wavelet, and the initial model is too far from the exact model, the problem becomes quite nonlinear in this case, and the FWI falls into a local minimum (Figure 4-19). We perform the multi-scale inversion to see if it could improve the inversion result.

Multi scale FWI

The parameters for multi-scale FWI remain the same as in the third test, and the inversion results at each step are shown in Figure 4-20. Compared to the conventional FWI, multi-scale FWI improves the long-wavelength part of the model (Figure 4-21). But with this initial model and this source wavelet, the multi-scale FWI is still unable to retrieve the long wavelengths of the velocity model and falls into a local minimum.

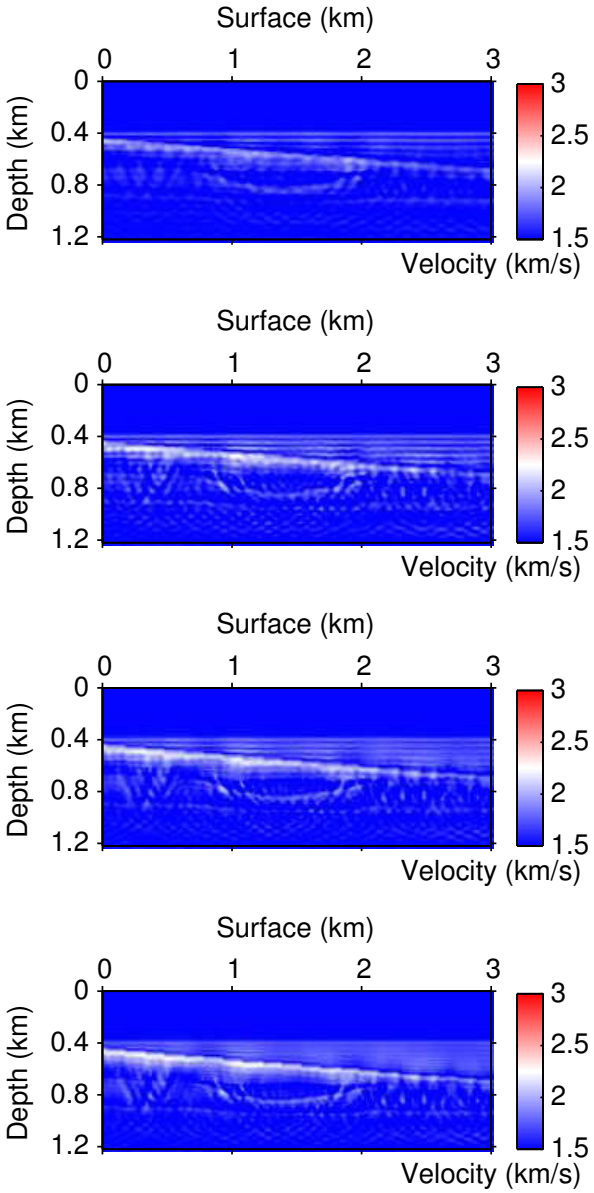


Figure 4-20: Test 4. Multiscale FWI results at each steps. From top to bottom, the source bandwidths used are respectively from (a) to (d) those shown in Figure 4-13.

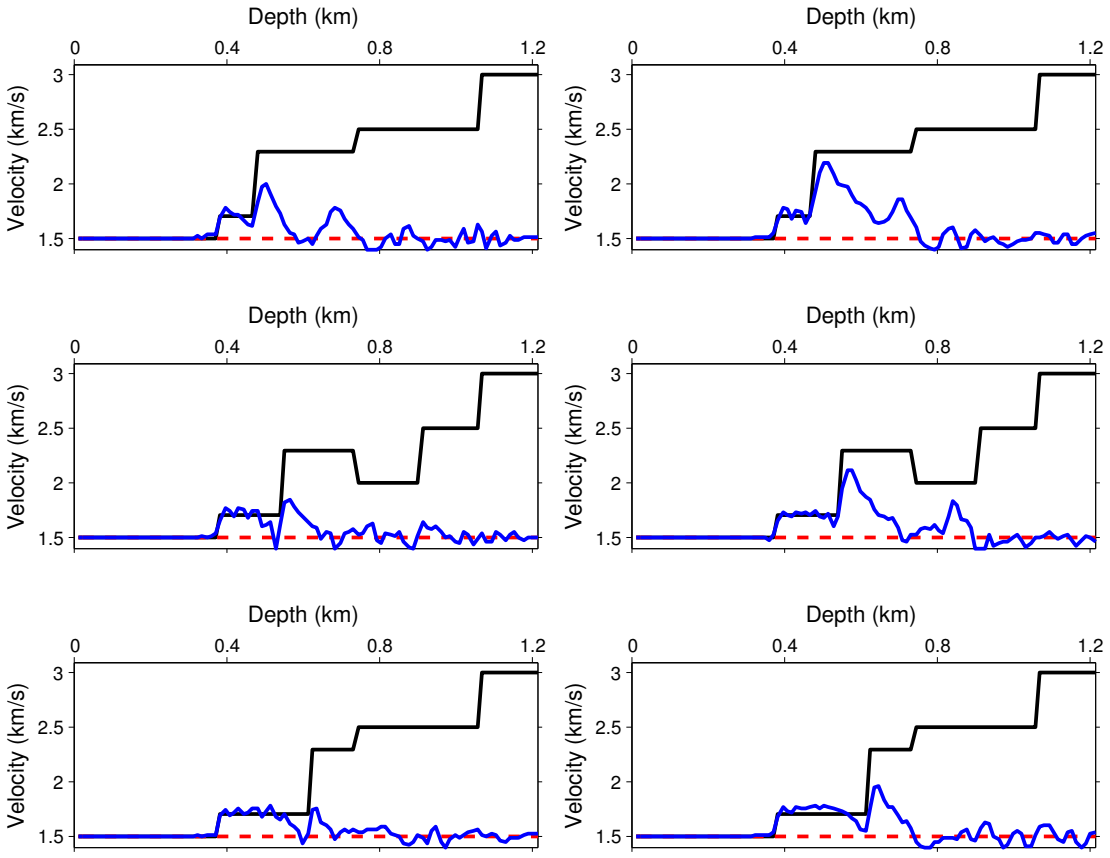


Figure 4-21: Test 4. 1D vertical profiles of exact (black line), initial (red line) and FWI model (blue line) (left), or multiscale FWI model (blue line) (right). From top to bottom: $x = 0.4$, 1, 1.6 km.

DWI

Finally, we apply DWI for this test. Except for the initial model, the inversion parameters are the same as in the third test. The DWI allows retrieving the long-wavelength component of the model (Figure 4-22 on the top) such that FWI using this model as starting model allows getting a good result (Figure 4-22 on the bottom). Note that the deepest layer of the model cannot be retrieved as there are no waves reflected back from the bottom of the model. The comparison of the vertical velocity profiles with multi-scale FWI are shown in Figure 4-23. The comparison of the exact data, the initial data and final data obtained with FWI, multiscale FWI and DWI is displayed in Figure 4-24.

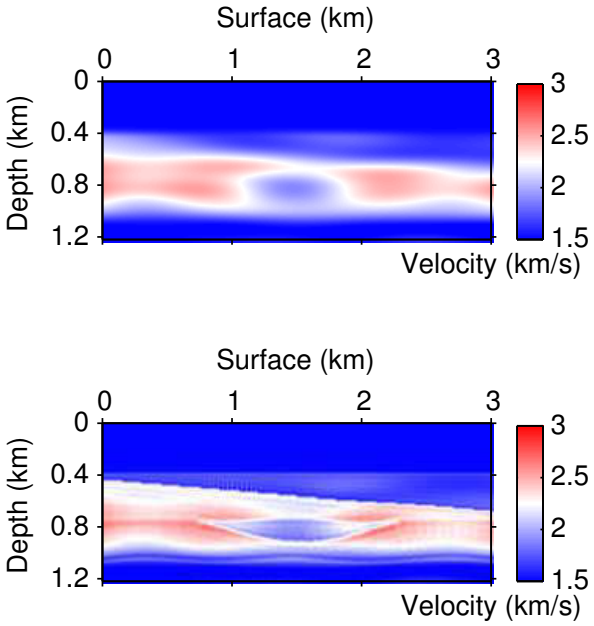


Figure 4-22: Test 4. Long-wavelength model obtained for DWI (top), and FWI result starting from model on the top (bottom).

4.3.5 Analysis of the model perturbation

As presented in Chapter 3, the construction of the short-wavelength reflectivity model is a critical step of the proposed algorithm, since a more correct reflectivity allows for a better long-wavelength update. The proposed inversion scheme combines several elements in the short-wavelength update step to build a true-amplitude, focused and sharp reflectivity model: (1) zero-offset data; (2) nonlinear iterative inversion (20 iterations in our example) and (3) a source with high-frequency content. In this section, we show the importance of these elements by analyzing and comparing the model perturbations that we obtain after the first iteration of the long-wavelength update. The analysis of the model perturbation is similar to the analysis of the sensitivity kernel in the gradient.

In the following tests, the initial velocity model is the gradient-constant velocity model (Figure 4-4 on the top). The true model perturbation (the difference between the true velocity model and the initial velocity model) is displayed in Figure 4-25 on the top. The model perturbation retrieved after the first global iteration (short-wavelength update followed by long-wavelength update) by using the three above-mentioned elements is shown in Figure 4-25 on the bottom. This is the reference for comparison. The result is close to the smoothed version of the true model perturbation shown in Figure 4-25 in the middle as if the new method reduces the nonlinearity of the problem and speeds up the convergence.

In the following, we analyze the effect of partly applying all of the above-mentioned elements and compare the velocity perturbation to the one displayed in Figure 4-25 on the bottom.

Figure 4-26 shows the model perturbation after the first iteration of long-wavelength update when full-offset data (1.2 km offset) instead of zero-offset data is used in the short-

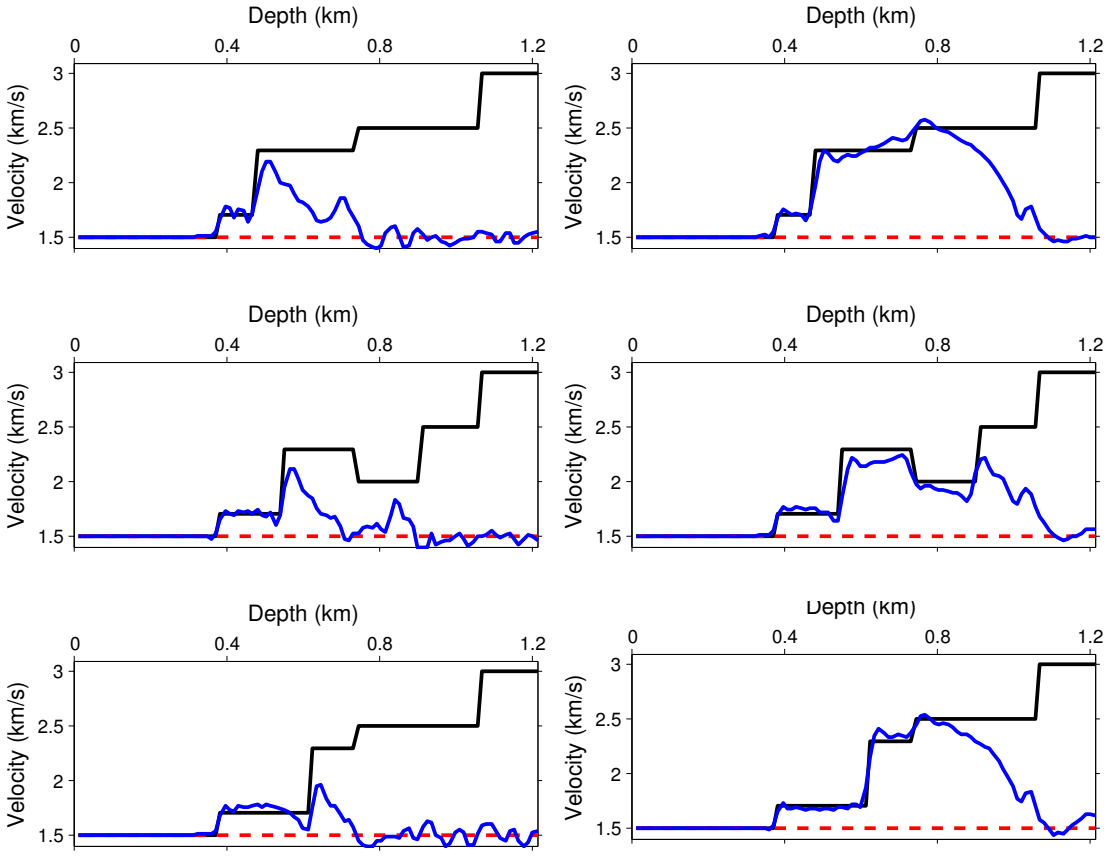


Figure 4-23: Test 4. 1D vertical profiles of exact (black line), initial (red line) and multiscale FWI model (blue line) (left), or decomposition-based waveform inversion + FWI model (blue line) (right). From top to bottom: $x = 0.4$, 1, 1.6 km.

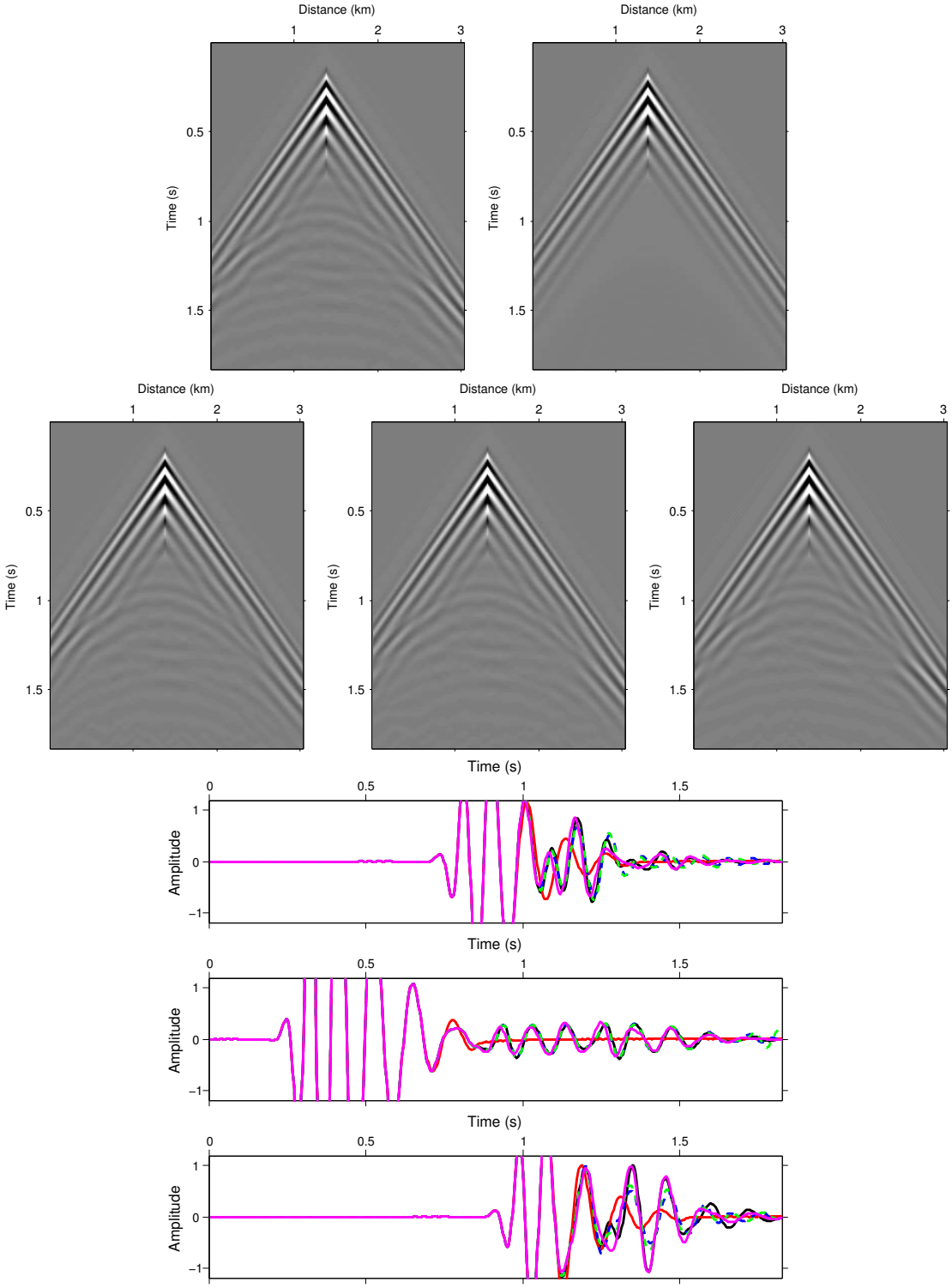


Figure 4-24: Test 4. Top panels: Shot gathers for observed data (left), initial data (right). Middle panels: Final data obtained with FWI (left), multi-scale FWI (middle) and DWI (right). Bottom panels: Observed (black), initial (red), FWI modeled (blue), multi-scale FWI modeled (green) and DWI modeled (magenta) seismic traces taken at three offset positions: 0.5 km (top), 1.5 km (middle) and 2.5 km (bottom).

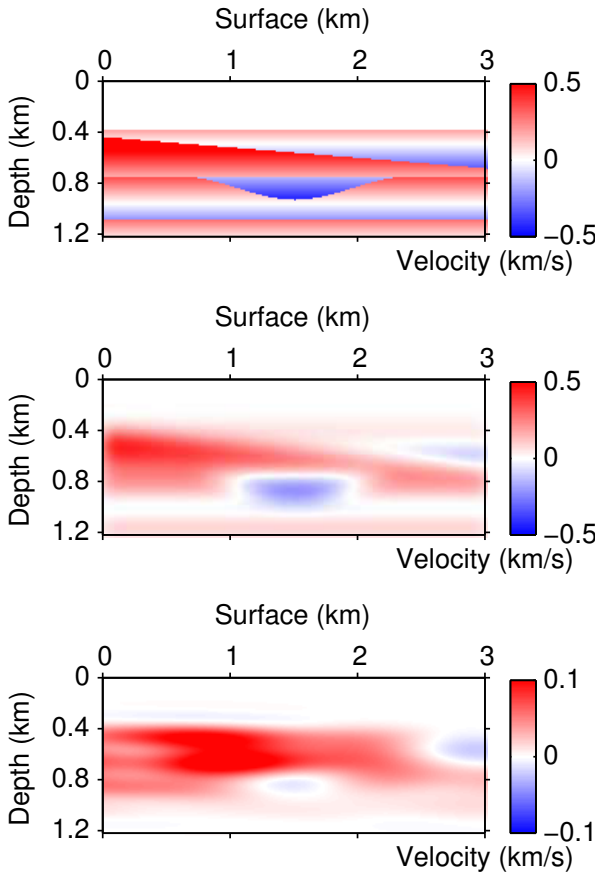


Figure 4-25: (Top) The difference between the exact velocity model (Figure 4-3) and the gradient-constant initial velocity model (Figure 4-4 on the top). (Middle) The smoothed version of the image on the top. (Bottom) The model perturbation of the first global iteration for the proposed decomposition-based inversion using zero-offset data and 20 iterations of FWI in the first step of inversion. Source wavelet without low frequencies less than 5 Hz is used for the two steps of inversion.

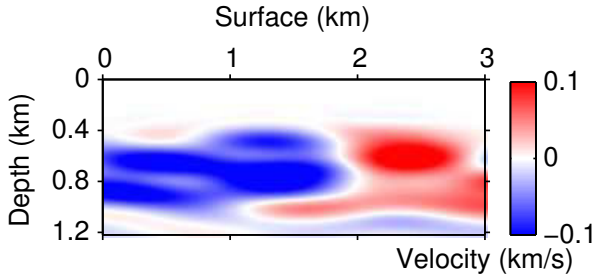


Figure 4-26: Model perturbation of the first iteration of long-wavelength update for the proposed decomposition-based waveform inversion when using large offset data (1.2 km) in the short-wavelength update step of the inversion.

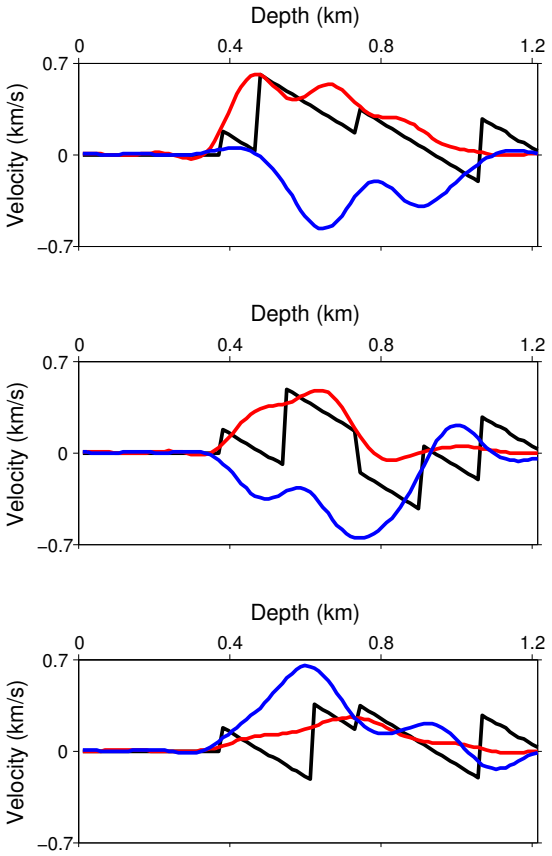


Figure 4-27: 1D vertical profiles of exact model perturbation (black line), model perturbation for DWI applying all the key parameters (red line) and model perturbation for DWI when large offset data (1.2 km) is used in the short-wavelength update step of the inversion (blue line). From top to bottom: $x = 0.4$, 1, 1.6 km.

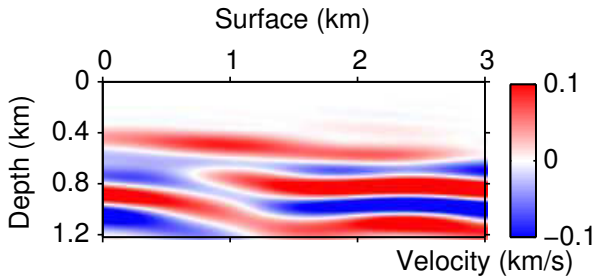


Figure 4-28: Model perturbation of the first iteration of the long-wavelength update for DWI when using 1 iteration instead of 15 iterations in the short-wavelength update step of the inversion.

wavelength update. Compared with the one in Figure 4-25, the result is far from the true model perturbation (refer to Figure 4-27). The sign is opposite. As by using large-offset data, move-out information is introduced and long-wavelength components are also updated in the first step, which is contrary to the idea of decoupling long wavelengths and short wavelengths, the inversion falls into a local minimum after the first global iteration. Figure 4-28 shows the model perturbation after a single iteration of the long-wavelength update step of inversion when running 1 iteration (6 iterations for L-BFGS) instead of 15 iterations (20 iterations for L-BFGS) of zero-offset FWI in the short-wavelength update step of inversion. Similarly, as shown in Figure 4-29, the result is far from the true model perturbation as iterations help to correct the reflectivity. If just one iteration is performed, the amplitude of the reflectivity is not correctly recovered and the amplitude error in the data residual will be interpreted as background velocity error in the second step.

Figure 4-30 and 4-31 shows the model perturbation when full-spectrum source is used in the first step. The image is close to the true model perturbation except on the edge. This is caused by the fact that low frequencies in the source will bring long wavelength update in the first step.

Figure 4-32 shows the model perturbation when using a linearized inversion based on the Born approximation in the short-wavelength update step as [Symes and Carazzone, 1991] and [Xu et al., 2012] proposed. Note that the nonlinear inversion is performed in the second step as discussed previously. The poorer result (Figure 4-33) highlights the importance of the compatibility of waveform between these two steps even though the linearized inversion requires less iterations to converge and provides mainly short wavelength update. Figure 4-34 shows the comparison of a vertical slice taken on the reflectivity models obtained by FWI and linearized inversion respectively, both using zero-offset data. We observe that the results are quite similar. But as the velocity contrast at interfaces is quite large, about 15% in this case, the Born approximation is no longer strictly valid. As a result, small differences in reflectivity could result in large differences in the long-wavelength model update. If we use the Born approximation in the first step, more iterations may be required to get a similar long-wavelength model.

Through these four comparisons we conclude that the three key ingredients: high frequency source, zero-offset data and iterative inversion in the first step, help to build a more correct short-wavelength reflectivity which is favorable to the update of the long-wavelength model in the second step. Besides, we note from the comparison of vertical

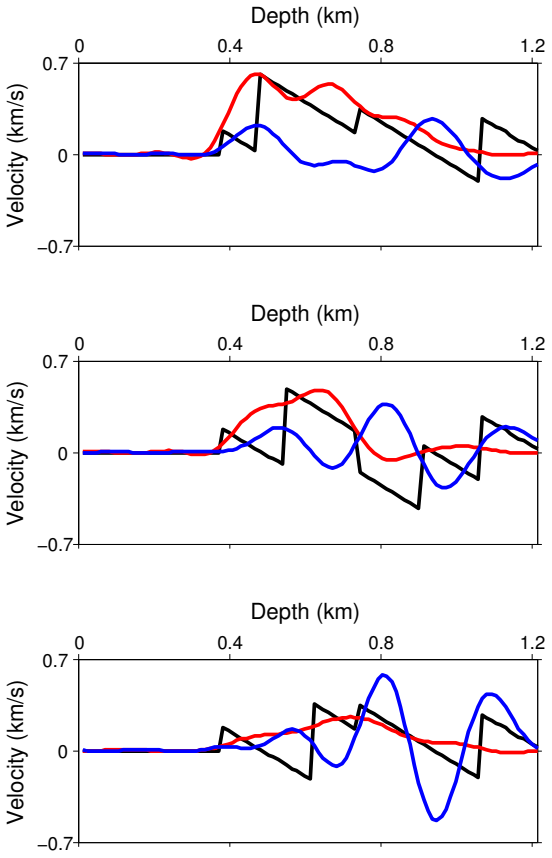


Figure 4-29: 1D vertical profiles of exact model perturbation (black line), model perturbation for DWI (red line) and model perturbation for DWI when 1 iteration instead of 15 iterations is used in the short-wavelength update step (blue line). From top to bottom: $x = 0.4$, 1, 1.6 km.

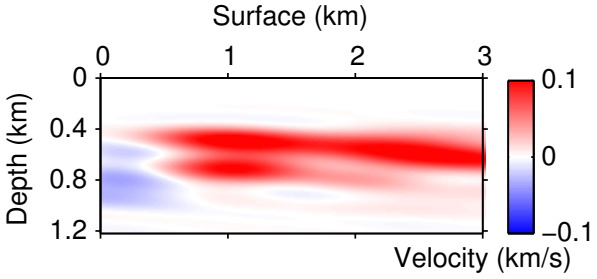


Figure 4-30: Model perturbation of the first global iteration for DWI when using source with low frequency in the short-wavelength update step of the inversion.

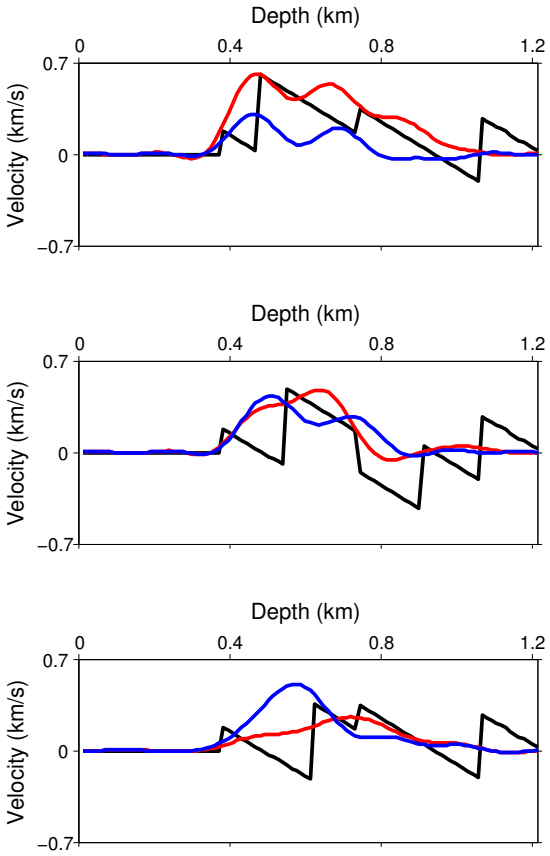


Figure 4-31: 1D vertical profiles of exact model perturbation (black line), model perturbation for DWI (red line) and model perturbation for DWI when using source with low frequency in the short-wavelength update step of the inversion (blue line). From top to bottom: $x = 0.4$, 1, 1.6 km.

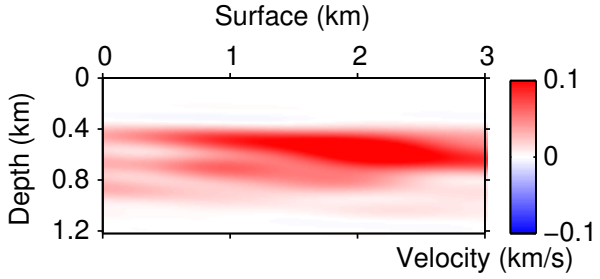


Figure 4-32: Model perturbation of the first global iteration for DWI when using Born approximation in the short-wavelength update step of the inversion.

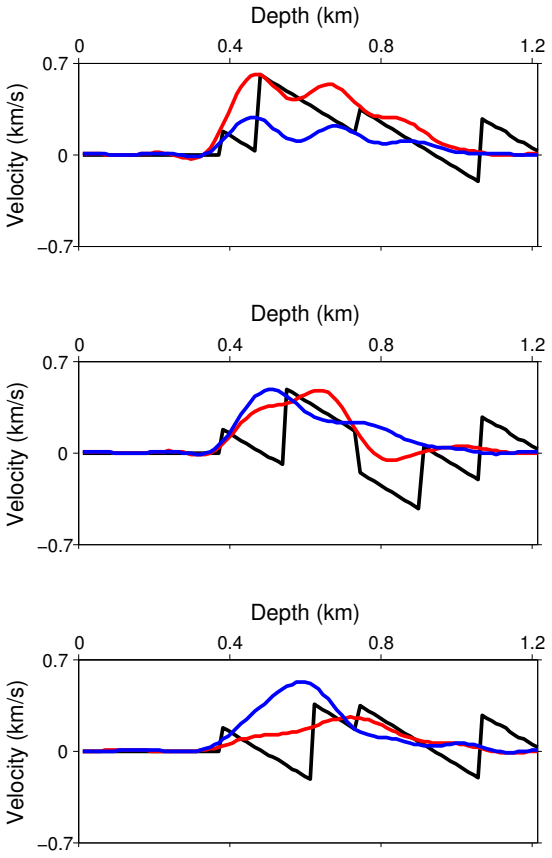


Figure 4-33: 1D vertical profiles of exact model perturbation (black line), model perturbation for DWI (red line) and model perturbation for DWI when using Born approximation in the short-wavelength update step of the inversion (blue line). From top to bottom: $x =$ 0.4, 1, 1.6 km.

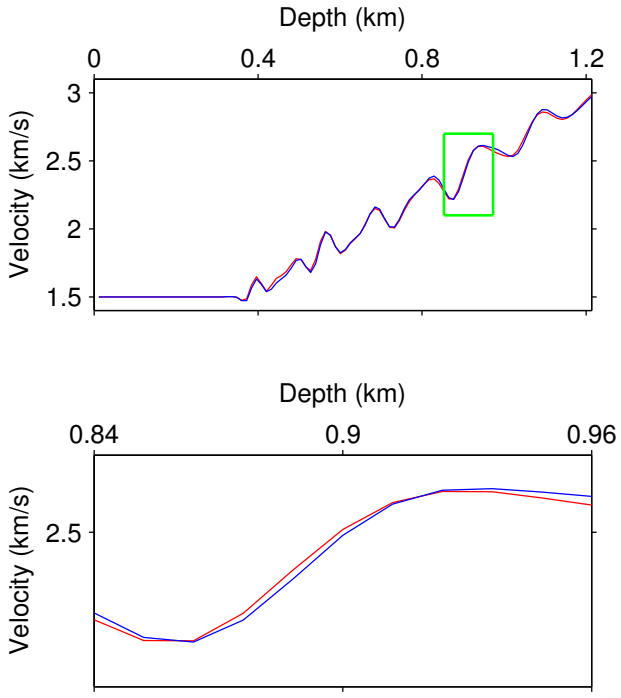


Figure 4-34: (Top) Velocity profiles at 1.5 km for the model obtained after the first step of inversion by using zero-offset FWI (red line) and zero-offset iterative migration based on the Born approximation (blue line). (Bottom) Zoom of the zone in green box in (a).

profiles (Figures 4-27, 4-29, 4-31, 4-33), that these key issues also help to better balance the amplitude for different parts of the model.

Figure 4-35 shows the model perturbation when performing the first step (short-wavelength update) with the high-frequency source and then using full-spectrum source in the second step (long-wavelength update). The similarity of the results shown in Figure 4-36, when compared to the one displayed in Figure 4-25 on the bottom, confirms that the proposed method is capable of retrieving long-wavelength components without low frequency in the source. This property is similar to that of DSO. As shown by [van Leeuwen and Mulder, 2010], the DSO functional does not depend on the frequency content of the data.

There are two possible interpretations to explain the success of the decomposed gradi-

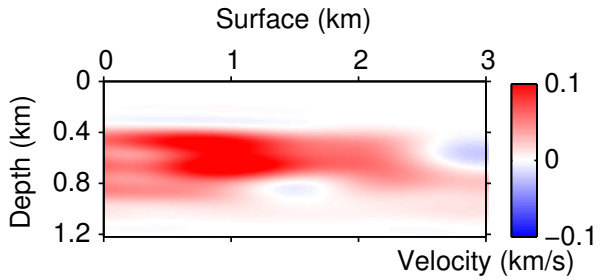


Figure 4-35: Model perturbation of the first global iteration for DWI when using source with low frequency in the short-wavelength update step of the inversion.

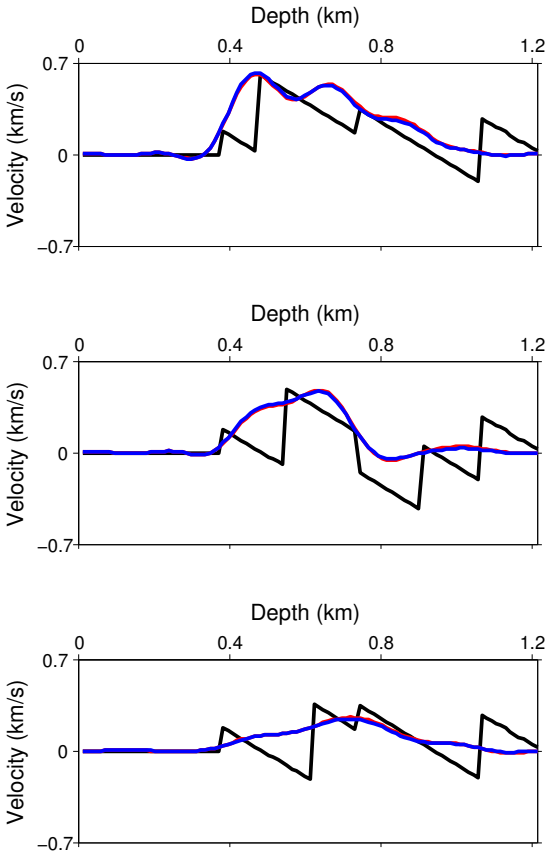


Figure 4-36: 1D vertical profiles of exact model perturbation (black line), model perturbation for DWI (red line) and model perturbation for DWI when using source with low frequency in the short-wavelength update step of the inversion (blue line). From top to bottom: $x = 0.4$, 1, 1.6 km.

ent to reduce the nonlinearity of the inversion problem. The first one is that the decomposed gradient is a filtered version of the original gradient of FWI. The filter serves as a projector of the original gradient direction into the direction of the true model perturbation. The second interpretation is that since we use only one part of the FWI gradient, the objective function in equation (2.1) is no longer valid. The new objective function corresponding to the new gradient is less nonlinear and has a larger basin of attraction.

4.4 Conclusion and discussion

In this chapter, we apply the FWI and DWI on two synthetic models and show that compared to the conventional FWI and the multi-scale FWI, DWI is more robust with respect to the absence of low frequencies in the source wavelet and to the large errors in the initial model. We have also analyzed the influence of three key parameters during the inversion and show that when these three key parameters are all applied in the inversion, the update of the long-wavelength part is closer to the true velocity perturbation.

From the computation cost point of view, the new method is computational expensive, as iterative inversion is performed at the short-wavelength update step to fit the data. However this iterative step is important, since it facilitates the long-wavelength update of the model. For migration velocity analysis, [Symes, 2015, Lameloise et al., 2015] also underline the importance of iterative migration to fit the data in the short-wavelength update, rather than merely image the reflectivity. A possible remedy to reduce the computation time is to perform the iterative inversion in the pseudo-time domain to avoid the re-computation of the reflectivity at each step [Plessix, 2012, Brossier et al., 2015].

We have also tested the idea that in the second step of inversion, we do not decompose wavefields, but directly smooth the FWI gradient to update the long-wavelength part of the model. However, this method fails to give a satisfactory result. Our analysis is that the limit between the short wavelengths and the long wavelengths is not fixed at each iteration, their spectrum may overlap or vary during the iterations, thus it is not trivial to determine the limit at each iteration.

Finally, all the examples shown in this chapter use models with constant and known density. We have also tested the more realistic case with the exact model having variable density, and the model used for the inversion having constant density. In this case, similarly as for conventional FWI, DWI fails to retrieve the correct model. For future work, we would like to extend the method to be suitable for elastic models with variable density. This aspect is described in the perspectives (Chapter 6).

Chapter 5

Real data application

Contents

Résumé du chapitre 5

Dans ce chapitre, nous appliquons la FWI classique et la DWI au jeu de données réelles 2D Brunei fourni par Total. L'acquisition des données Brunei a été réalisée en 2011, et la géométrie d'acquisition est montrée dans la Figure 5-1. Nous présentons d'abord les paramètres d'acquisition et les traitements de données que nous effectuons, y compris la vérification de la cohérence entre traces, le pointé des premières réflexions, la suppression des ondes directes, le filtrage passe-bas des données, le ré-échantillonnage et le débruitage des données. En particulier, nous testons et comparons trois méthodes de pointé et observons que celle proposée par [Arbelaez et al., 2011] est la plus efficace.

Nous estimons la source avec la méthode proposée par [Pratt et al., 1996]. La bande de fréquence estimée pour la source correspond à la bande de fréquence des données. En théorie, l'ondelette de la source doit être estimé à chaque itération de l'inversion, en même temps que la mise à jour du modèle de vitesse. Toutefois, pour simplifier, l'ondelette de la source que nous utilisons reste inchangée au cours de l'inversion. Néanmoins, cette hypothèse pourrait pénaliser le résultat d'inversion.

Nous soulignons aussi les défis que nous avons rencontrés. Tout d'abord, nous étudions le cas de la densité constante. Nous observons que avec l'hypothèse de densité constante la variation de l'amplitude avec offset diffère considérablement entre les données observées et calculées. En outre, pour les grands offsets, nous pouvons aussi constater un déphasage entre ces deux données. Comme nous utilisons l'approximation acoustique de la propagation des ondes avec densité constante, l'influence de la densité et d'autres paramètres physiques du sous-sol sur l'amplitude des données de réflexion ne pouvait pas être incorporé correctement. Notons que pour des ondes plongeantes, l'influence de la variation de densité n'a pas beaucoup d'effet sur l'amplitude des données, mais que pour les réflexions, cet effet est important. Nous essayons d'utiliser un modèle de densité variable pour compenser la différence d'amplitude. Nous testons la formule de Gardner et observons que l'amplitude est mieux expliquée que dans le cas de densité constante. Mais le problème de déphasage aux grands offsets existe toujours. Il semble donc difficile de trouver un modèle de densité qui pourrait mieux expliquer la variation d'amplitude avec offset. Nous pensons que l'atténuation est un autre facteur important, surtout dans les zones superficielles en présence de gaz. De ce fait, nous avons décidé de prendre en compte les offsets limités dans les données.

Le deuxième test préparatoire nous effectuons est d'utiliser le modèle de vitesse fourni par Total pour faire la migration. Nous observons que l'image migrée dans ce modèle de vitesse n'est pas focalisée. Si l'on migre avec les très courts offsets, les images sont bien focalisées mais diffèrent les unes des autres et la somme des contributions n'est plus cohérente, surtout dans les zones de gaz.

Avec les difficultés que nous avons rencontrées, nous montrons quand même les résultats d'inversion en utilisons FWI classique et DWI. D'après le résultat de FWI, nous observons que la partie des grandes longueurs d'ondes du modèle n'est pas mis à jour. En outre, en raison de l'hypothèse de densité constante, la mise à jour de la densité se retrouve dans la mise à jour de vitesse, visible à la fond de l'eau. Pour DWI, nous observons que la poche de gaz est mieux récupérée que dans FWI. Mais le problème de l'amplitude existe toujours, qui est aussi visible sur le fond de l'eau. D'autres études sur ce sujet vont certainement aborder l'utilisation du modèle visco-élastique. En outre, plus de travail serait nécessaire

pour analyser l'inversion multi-paramétrique avec la méthode DWI.

5.1 Introduction

In this chapter, we present an application on the 2D Brunei real dataset as provided by Total. The acquisition of the Brunei data set was carried out in 2011, and the acquisition geometry is displayed in Figure 5-1. We first present the seismic acquisition parameters and data processings we performed. We point out the challenges we met when dealing with this dataset. Then we show the first inversion results using the conventional Full Waveform Inversion (FWI) and Decomposition-based Waveform Inversion (DWI). More work is still needed to correctly image the subsurface.

5.2 Seismic dataset

The seismic dataset consists of a sub-line (that mimics a 2D acquisition) with 608 shot gathers with sources spaced 50 m apart and receivers spaced 12.5 m apart. Within each shot gather, there are 516 channels. The first channel is situated at 150 m from the source and the last channel is situated at 6587.5 m. The temporal sampling is 3 ms and the total recording time is 9 s.

Several preprocessings were already done by the provider of the data:

- Multiple removal using 3D SRME and Radon-based methods.
- Butterworth high-pass filter with low cut frequency of 3 Hz.
- Swell noise and linear noise attenuation.

Figure 5-2 displays one shot gather after the preprocessings described above, and the spectrum of the data corresponding to this shot gather. To meet our needs, we have also applied other preprocessings, including:

- Seismic event coherency check. In this step, we check the quality of each shot gather and noisy traces are dropped.
- Picking and direct wave removal. The direct waves are not completely removed from the initial data, as shown in Figure 5-2. As we do not use direct waves for the inversion, we apply a picking on the data and remove the direct waves. For the first-reflection picking we have first tested two methods that are classically used for picking of the first-breaks: 1) the method proposed by [Coppens, 1985], that consists in calculating the ratio of energy of seismogram for two windows, one before the current time point, and one after the current time point. However, with this dataset, this approach fails to detect the first-reflection arrival because of the presence of the residual direct wave (Figure 5-3); and 2) the STA/LTA (Short Term Average / Long Term Average) ratio method [Allen, 1978], that consists in calculating the ratio of two averages of energy between a short-term window and a long-term window. The picking results using different sizes of windows are shown in Figure 5-4. It fails to pick the first reflection and indicates the second reflection.

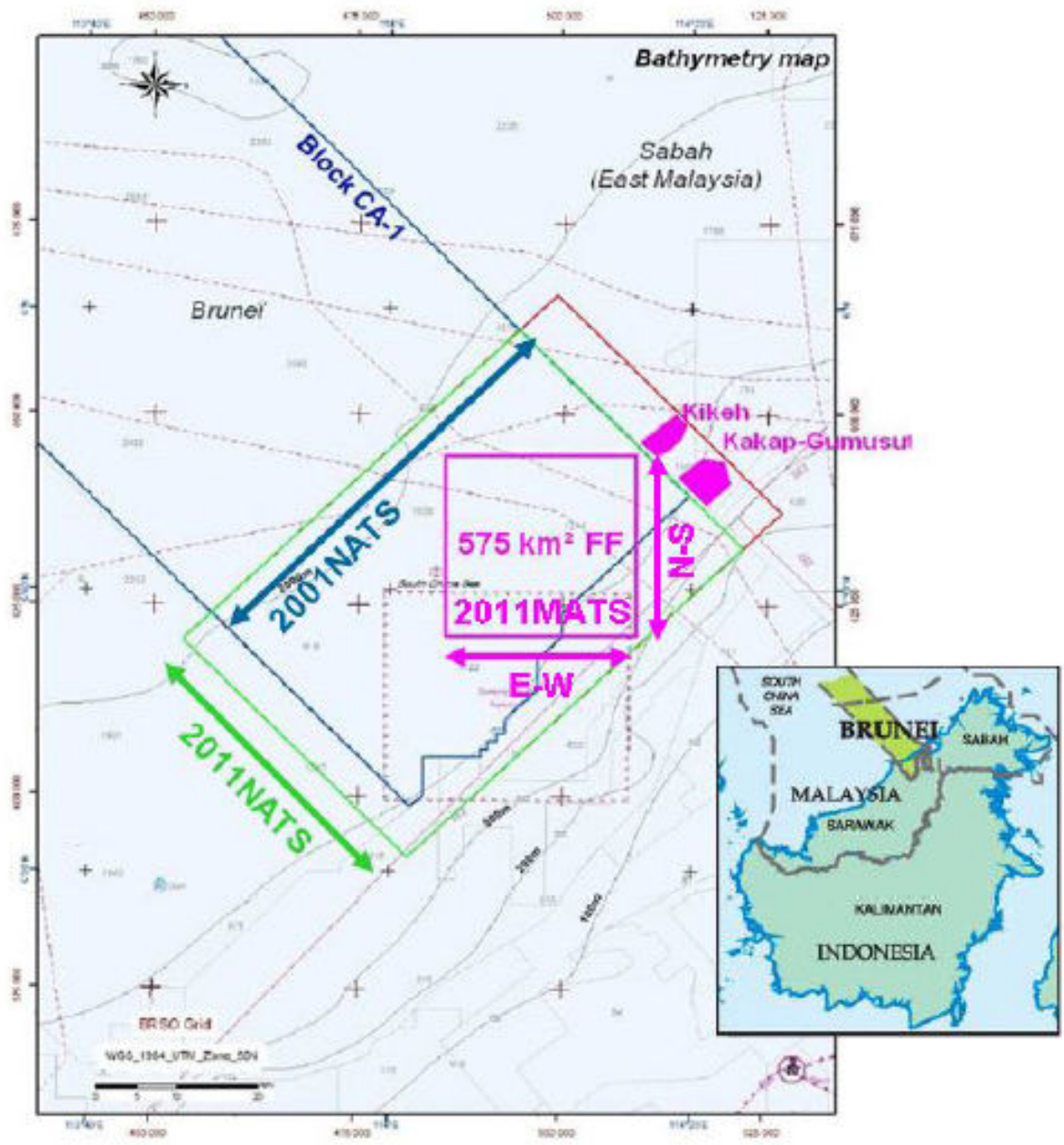


Figure 5-1: Acquisition geometry of the Brunei dataset.

We also tested an algorithm that is widely used in image processing, but less standard in geophysics. The method is based on the contour detection [Arbelaez et al., 2011]. It is similar to Coppens's algorithm. This approach calculates the absolute value of the difference, instead of the ratio of energy of seismogram for two windows. This alternative improves the picking result, as the difference operator is more stable than the ratio operator. The picking result is shown in Figure 5-5. It successfully picks the first reflection. The result after the muting of the direct wave is shown in Figure 5-6 on the left.

- Low pass filtering of the data. As shown in Figure 5-2 on the right, the spectrum of the original data is up to 100 Hz. This frequency is very high for FWI. Moreover, with this frequency, the spatial and the temporal samplings for the wave propagation modeling would be very small and the size of model would exceed our computational capacity. Thus, we apply a low pass filter on the data, by using a $10^{\mathbf{th}}$ order Butterworth filter with high cut frequency of 10 Hz. The filtered shot gather is shown in Figure 5-6 on the right.

- Data resampling. The data is resampled from 3 ms to 2 ms in the time domain, and from 25 m to 12.5 m in the space domain, to satisfy the criteria for the spatial and the temporal samplings for the wavefield propagation.

- Data de-noising. In the lower part of the shot gather (Figure 5-6), we can observe some artefacts that intersects with each other and have slopes different from the slopes of the signal. We filter these artefacts in the f-k domain. The denoised shot gather and its spectrum are displayed in Figure 5-7.

Note that there is no 3D to 2D amplitude correction in the initial data. This correction consists in multiplying the data by $\sqrt{t}$, where t is the recorded time. Since there is a lot of noise in the bottom of the data, rather than amplifying later arrivals in the observed data, we instead correct the calculated data by dividing it by $\sqrt{t}$. This behaves as a preconditioning on the data.

5.3 Seismic inversion

A FWI model is provided by Total using long-offset data (Figure 5-8). The dimensions of the model are 16.7625 km large by 3.0125 km deep. The velocity model building has to contend with challenging geology, including steeply dipping structure and the presence of large shallow gas pockets. These gas pockets exhibited extremely low seismic velocities, and had a negative impact on deeper reflector amplitudes. The objective is to localize these gas traps. When we began to deal with this data set, one of the first tests we wanted to perform was to compare the data modeled on this FWI model with the processed observed data. For this, we need first to estimate the source wavelet.

5.3.1 Estimation of the source wavelet

We use the source-estimation method proposed by [Pratt et al., 1996], and described in Chapter 2. This method was initially applied on direct waves, but as the direct waves are

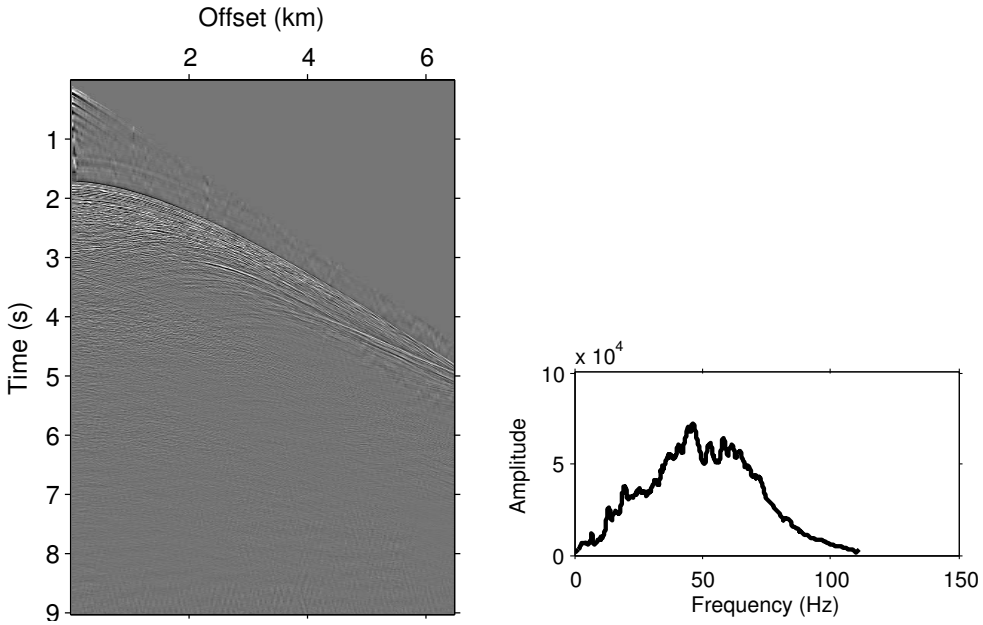


Figure 5-2: One original shot gather (left) and its spectrum (right).

partially removed in our initial data, the residual direct waves are not sufficient for the source estimation (Figure 5-2). Instead we use the first reflection at the water bottom to estimate the source wavelet. We have tested this approach at three source locations: at 2.5 km, 7.5 km and 10 km respectively. The estimated source wavelets are not exactly the same for these three locations. However, we decided to use an average source wavelet. The averaged source wavelet and its spectrum are displayed in Figure 5-9. The bandwidth of the estimated source corresponds to the bandwidth of the data (Figure 5-7). In theory, the source wavelet should be estimated at each iteration of the inversion, along with the velocity update. However, for simplicity, the source wavelet we use remains unchanged during the inversion. This may penalize the inversion result.

5.3.2 Preparatory tests

With the estimated source wavelet, we compute the data using the model in Figure 5-8. First, we investigate the constant density case. Let consider the shot gather for the source situated at 7.5 km. The comparison of the observed data and the calculated data is displayed in Figures 5-10 and 5-11. We note that the variation of the amplitude with offset differs greatly between these two data. Besides, for large offset, the phase is also shifted. As we use the acoustic wave propagation approximation with constant density, the influence of density and other physical parameters of the subsurface on the reflection data amplitude could not be incorporated correctly. Note that for diving waves, the influence of the variation of density does not have much effect on data amplitude, but for reflections, this effect is important.

We try to use a variable density model to compensate the amplitude difference. We rely on the Gardner's formula [Gardner et al., 1974]:

$$\rho = \alpha V^{\beta}. \tag{5.1}$$

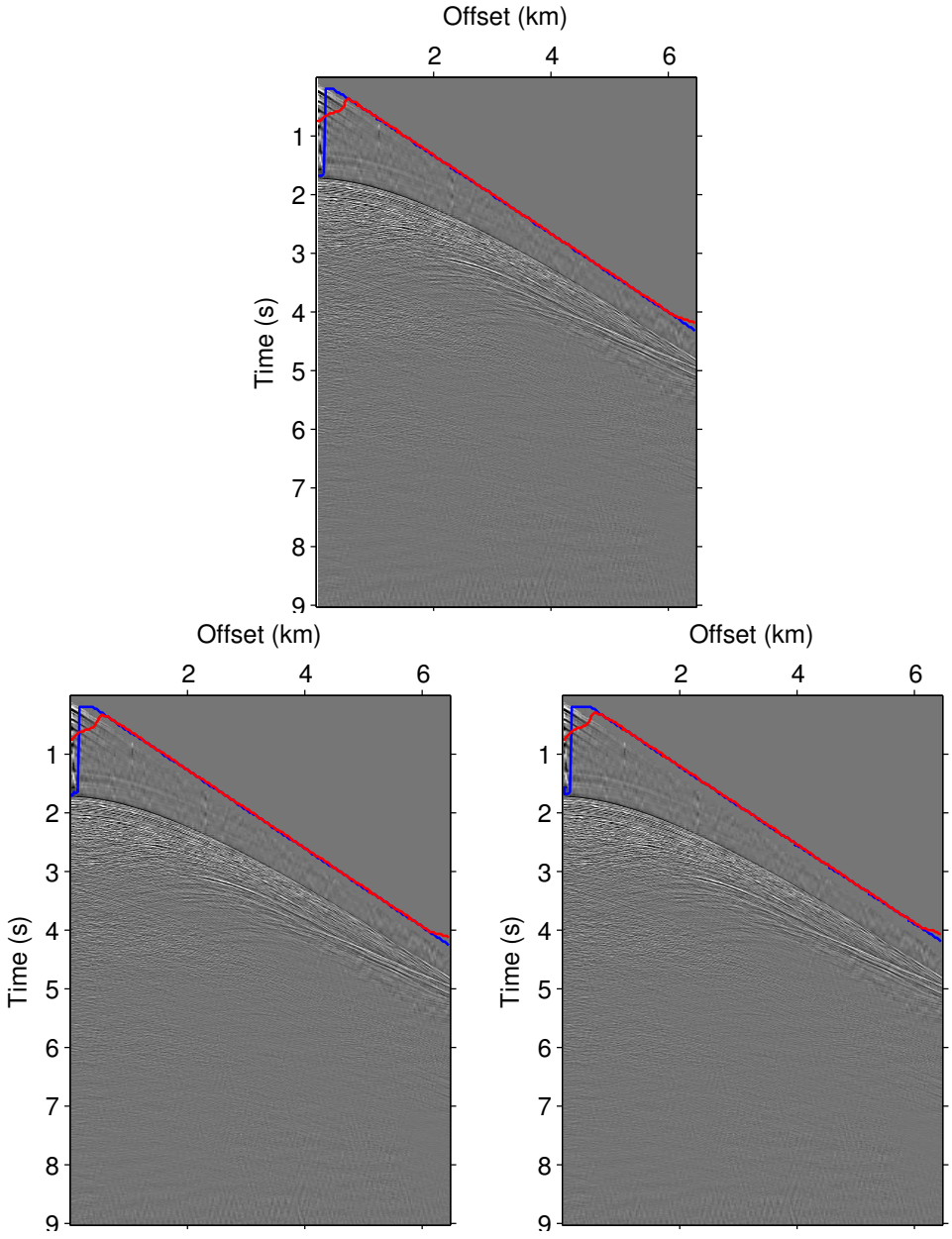


Figure 5-3: Picking results using the Coppen's algorithm (blue) and smoothed versions (red). From top to bottom, the size of windows are respectively 0.09 s, 0.15 s and 0.21 s.

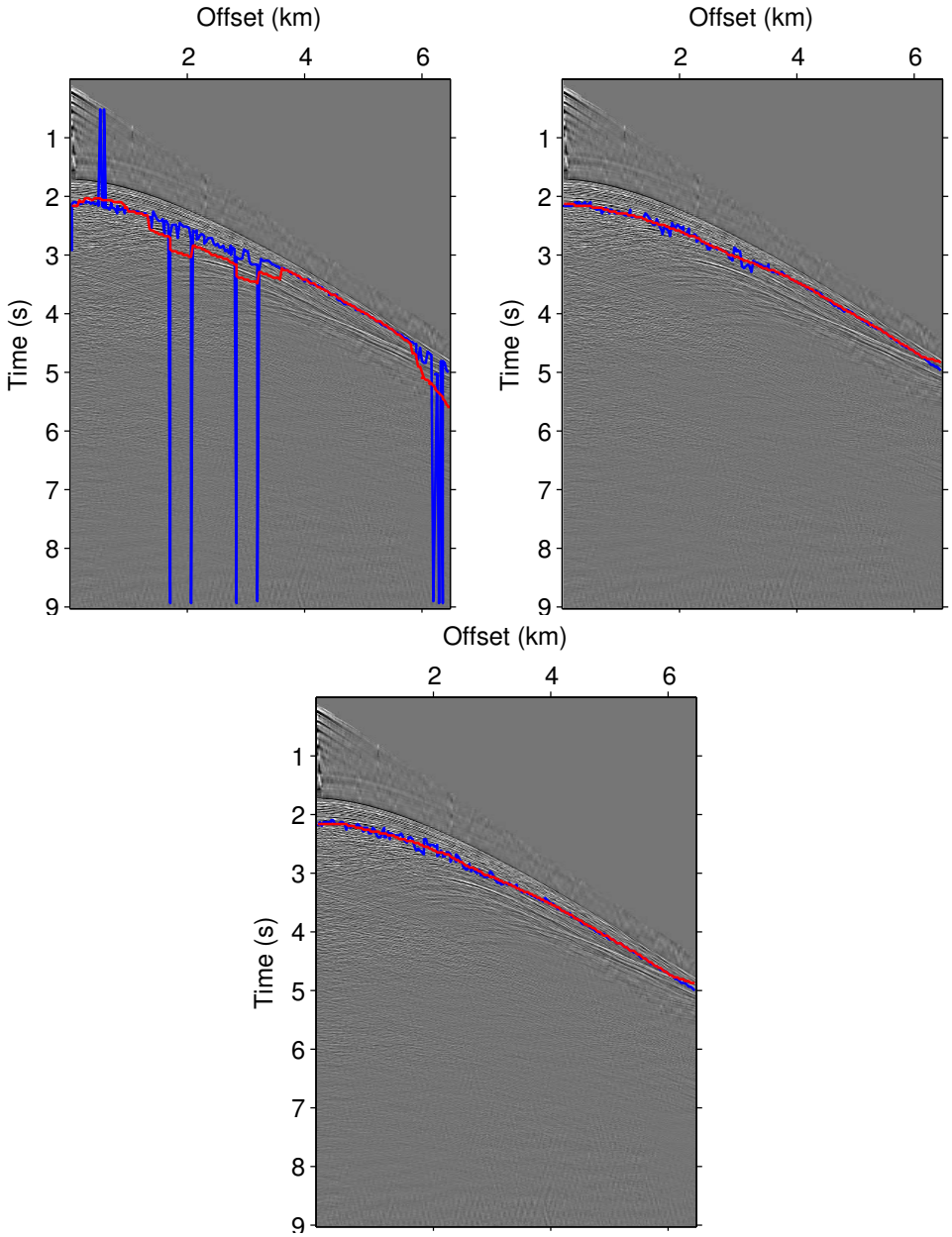


Figure 5-4: Picking results using the STA/LTA algorithm (blue) and smoothed versions (red). From top to bottom, the size of short-term windows are respectively 0.09 s, 0.15 s and 0.21 s, and the size of the long-term windows are respectively 0.9 s, 1.5 s and 2.1 s

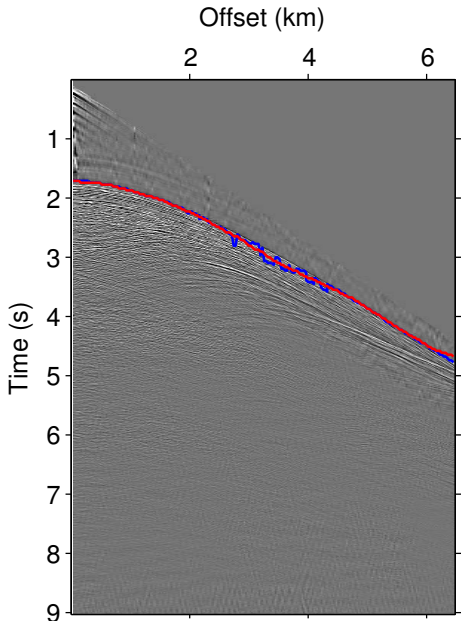


Figure 5-5: Picking results using the difference-based algorithm (blue) and smoothed versions (red). The size of the window is 0.09 s.

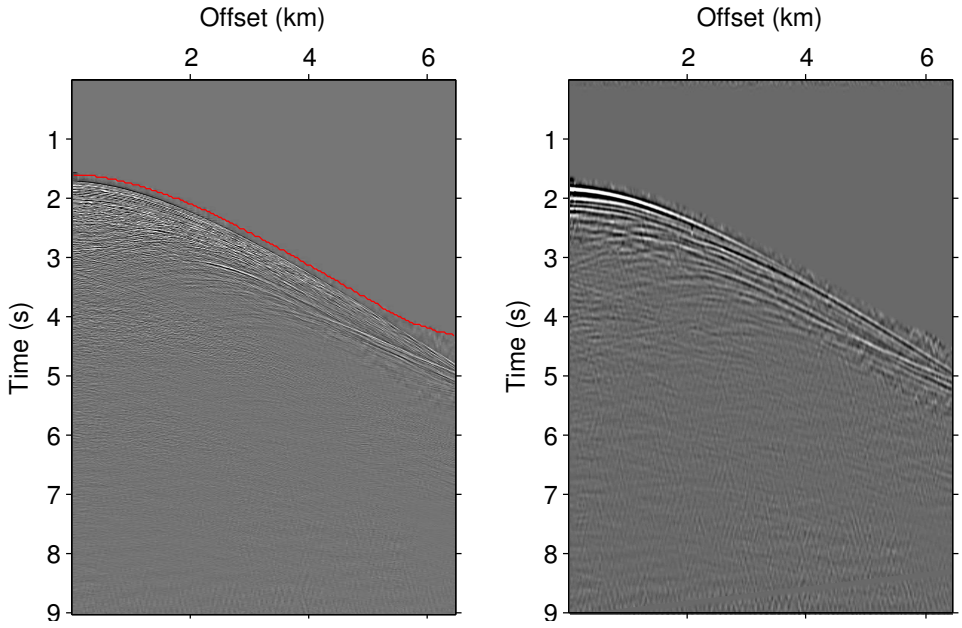


Figure 5-6: The shot gather after removal of direct waves (left) and the shot gather after low-pass filtering with cut frequency of 10 Hz (right).

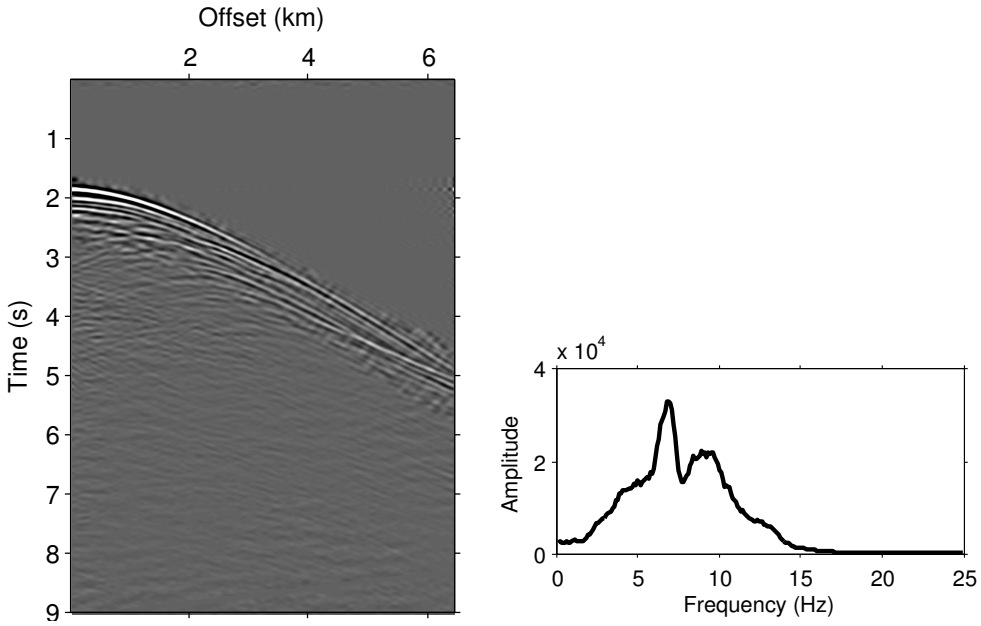


Figure 5-7: The shot gather after denoising in the f-k domain (left) and its spectrum (right).

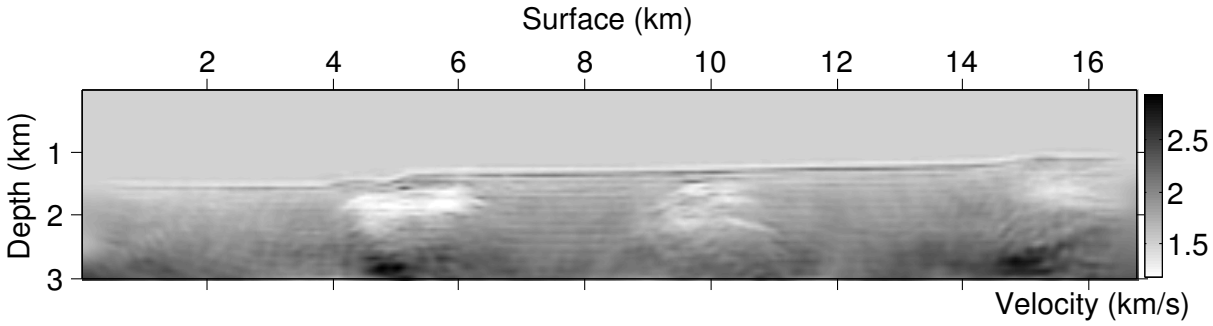


Figure 5-8: FWI model provided by Total.

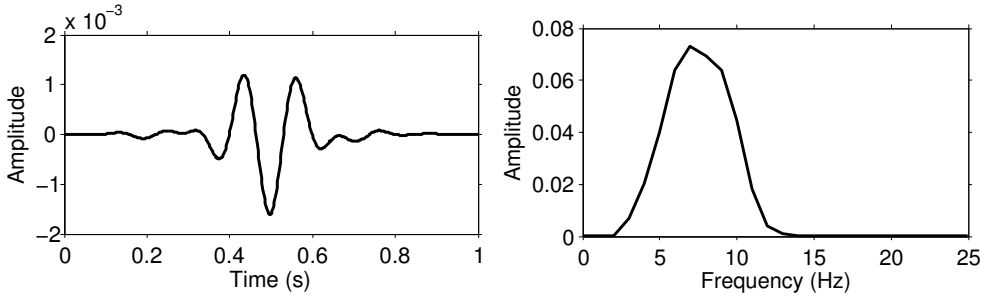


Figure 5-9: Estimated source wavelet in the time domain (left) and in the frequency domain (right).

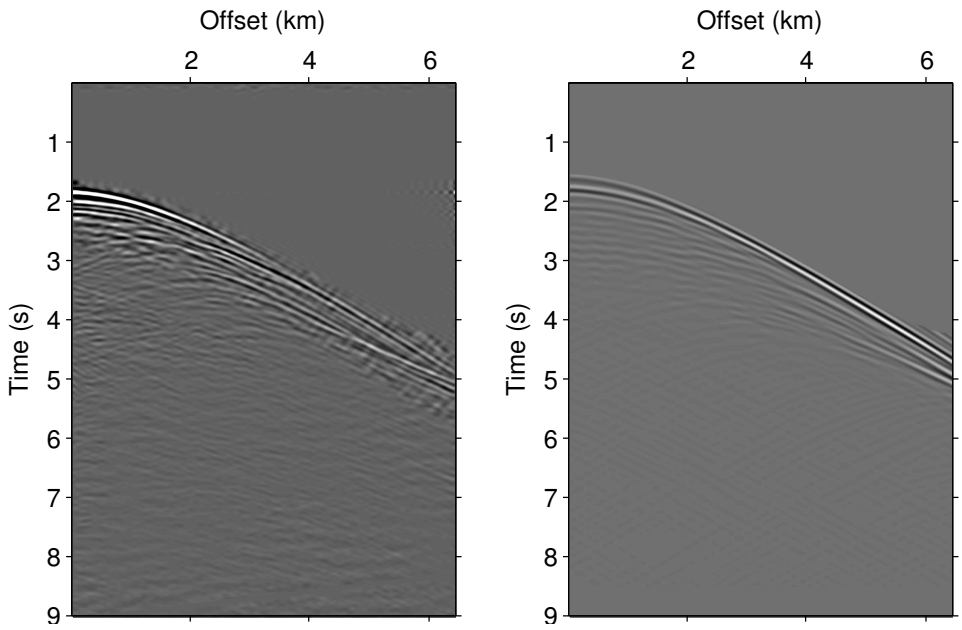


Figure 5-10: Constant density case. The observed data (left) and the calculated data in the model of Figure 5-8 (right).

The Gardner's formula is an empirically derived equation that relates seismic P-wave velocity to the bulk density of the lithology. We tested different choices of α and β and find none of them could match the data amplitudes satisfactorily. We cite one example here. To compensate the amplitude difference, a large density contrast at the water-bottom layer is needed. The velocity of the water here is 1495 m/s. We choose $\alpha = 7.74e-12$ and $\beta = 3.5$. The obtained density model is displayed in Figure 5-12. The source wavelet is re-estimated. We then recompute the shot gather for the source situated at 7.5 km, using the velocity model in Figure 5-8 and the density model in Figure 5-12. The comparison of the calculated and the observed data is shown in Figure 5-13 and 5-14. We observe that the amplitude is better matched than in the constant density case. But the phase shift problem still exists. Besides, due to the density change, the amplitude of the events below the first reflection is attenuated; refer for example to the event at t =5 s and x =6 km. It seems therefore difficult to find a density model that could better explain the real data amplitudes with offset variations.

As the decomposition-based waveform inversion does not need large offset data, we decide to only consider the data in the offset range of [150–2500] m to avoid the large variations in amplitude and phase.

The second preparatory test we carry out is to use the velocity model in Figure 5-8 to get a migrated image. We use the constant density model. The migrated image is displayed in Figure 5-15. We note that this image is not focused, especially in the red circled zone. To analyse this problem, we use data with smaller offset ranges. We test 5 offset ranges: respectively [150–275] m, [275–400] m, [400–525] m, [525–650] m, and [650–775] m. For each offset range, the migrated image is focused (Figure 5-16). But if we compare carefully, we find that these images are shifted horizontally (indicated by red vertical lines in Figure 5-16). Due to this shift, the sum of the five migrated images,

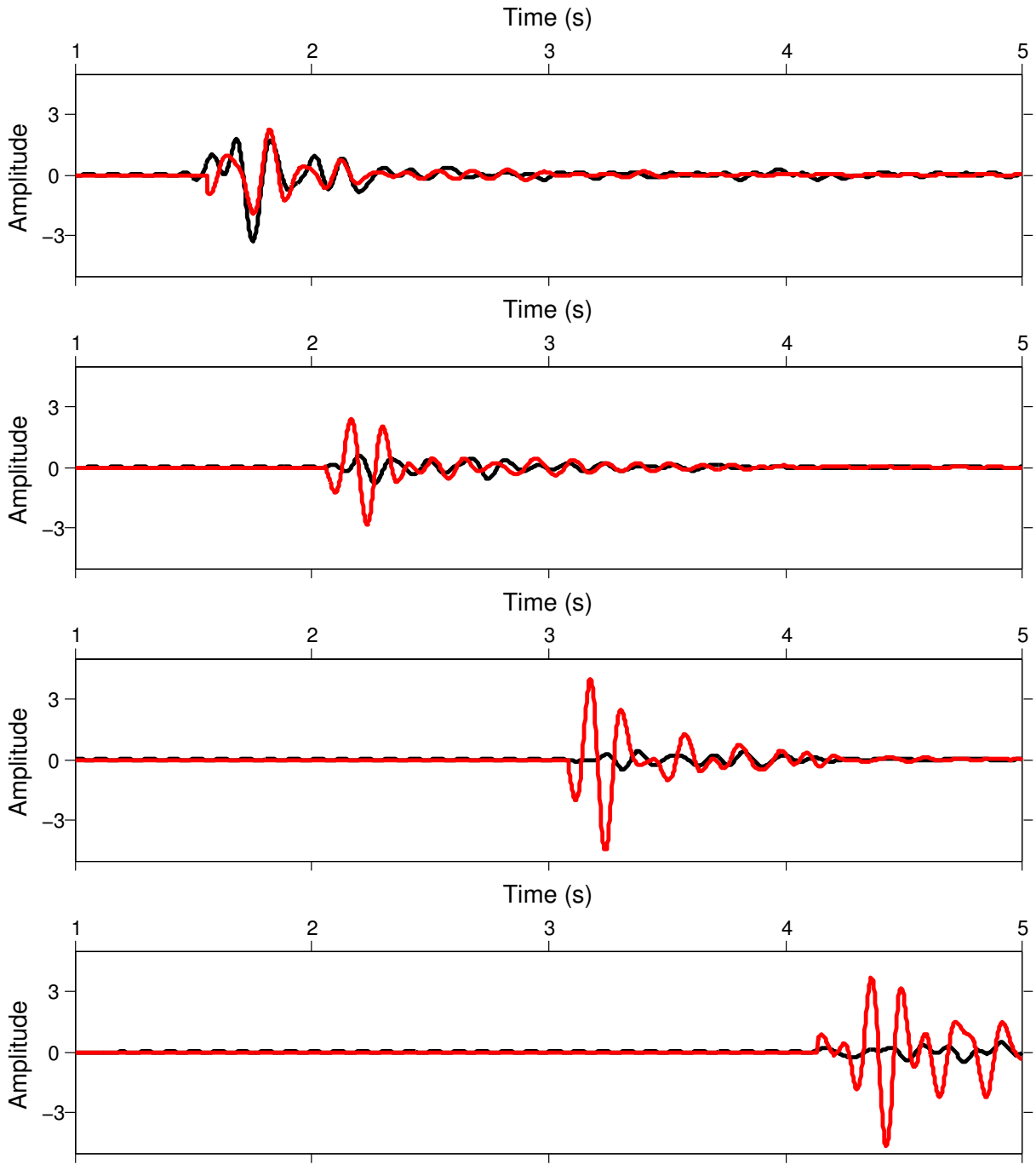


Figure 5-11: Constant density case. Seismic traces of observed (black) and calculated (red) data taken at four offset positions, from top to bottom: 0.15 km, 2 km, 4 km and 6 km respectively.

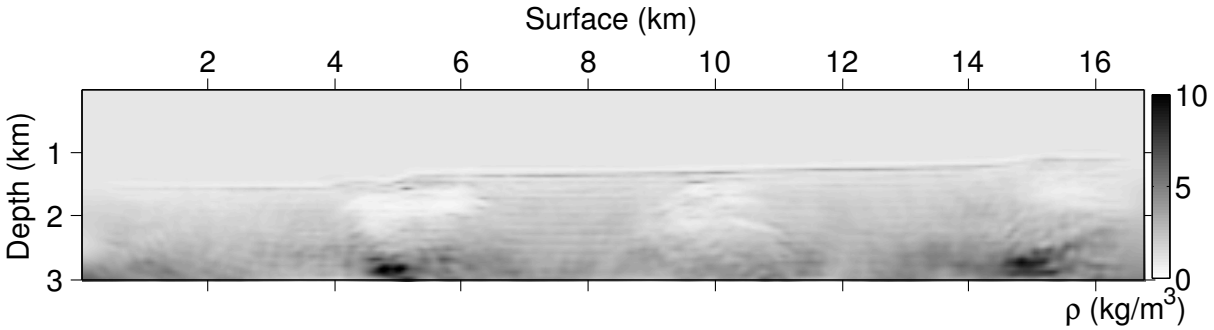


Figure 5-12: Density model obtained with the Gardner's formula in equation 5.1.

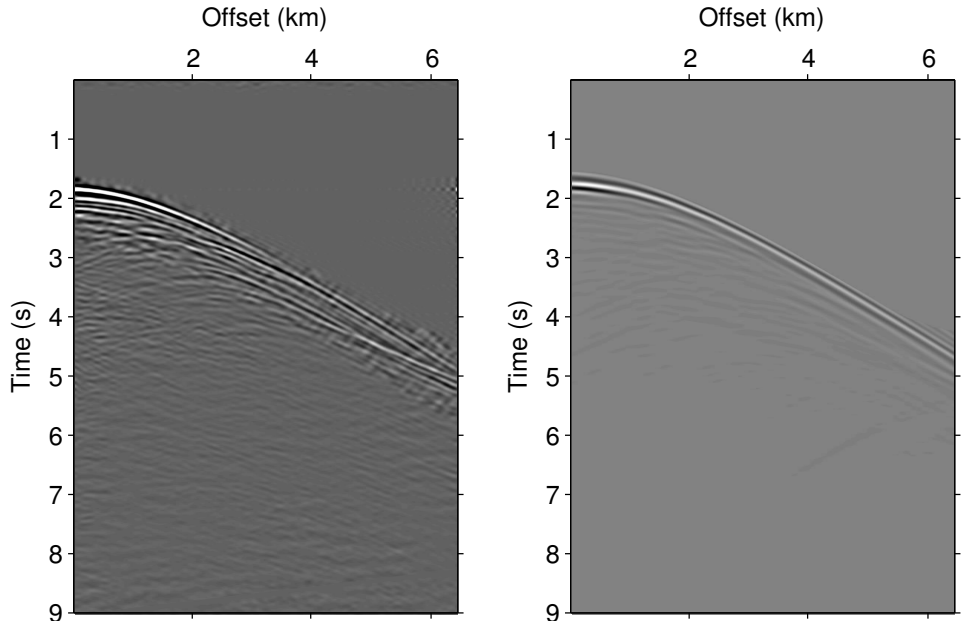


Figure 5-13: Variable density case. The observed data (left) and the calculated data (right) using the density model in Figure 5-12.

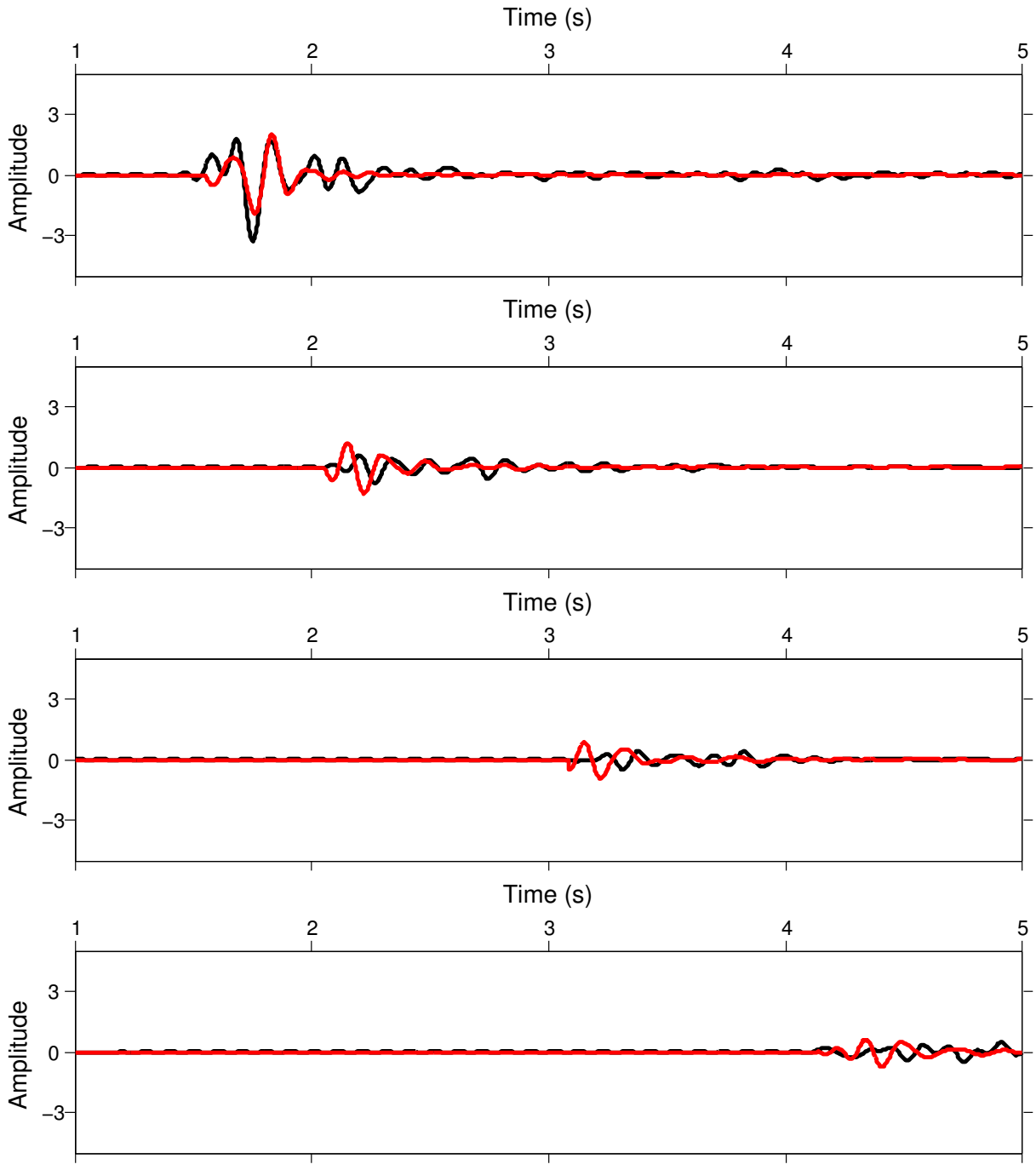


Figure 5-14: Variable density case. Seismic traces of observed (black) and calculated (red) data taken at four offset positions, from top to bottom: 0.15 km, 2 km, 4 km and 6 km respectively.

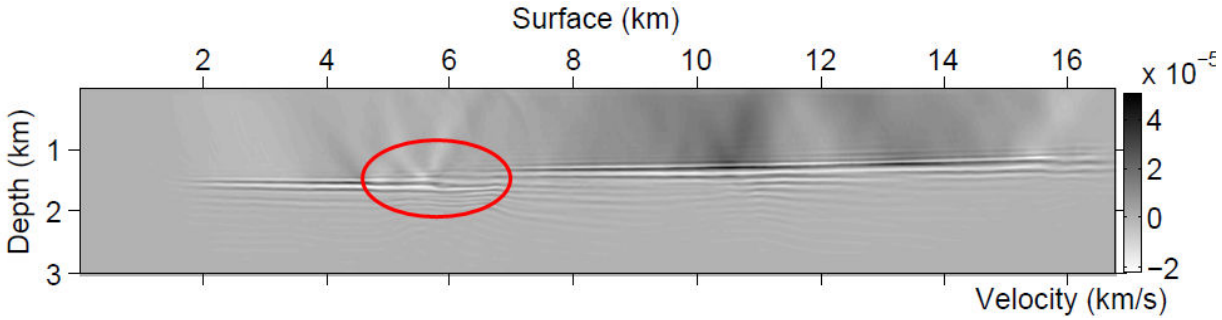


Figure 5-15: The migrated image using data within offset range of [150–2500] m. Note that the image is not focused in the red circled zone.

providing the migrated image for data with offset range of [150–775] m, is still unfocused at $x = 6$ km (see Figure 5-17). Three vertical slices at $x = 3$ km, 6 km, 10 km respectively in these five gradients are shown in Figure 5-18. We can observe the moveout at different offsets. For the following inversion tests, we only consider the right part of the model, starting from $x = 8$ km, to avoid the difficult zone near $x = 6$ km.

5.3.3 Seismic inversion

In this section, we apply FWI and DWI on the data. We consider the constant density case. The initial model (Figure 5-19) we use is the smoothed version of the FWI model in Figure 5-8.

5.3.4 FWI

We apply the conventional FWI on the data in the time domain, with the estimated source wavelet (Figure 5-9). All frequencies are involved at the same time. The sources are spaced 100 m apart and the receivers are spaced 12.5 m apart. The FWI result after 100 iterations is displayed in Figure 5-20. We note that the long-wavelength part of the model is not updated. Besides, due to the use of constant density, the density update is leaked into the velocity update (see the bottom of the water layer).

5.3.5 DWI

We apply the DWI on the same data, with the same source wavelet. We use a nested algorithm. For the short-wavelength update step, we perform 20 iterations of conventional FWI with short-offset data. The offset range is [150–400] m. For the long-wavelength update, we decompose the wavefield and only consider the long-wavelength part in the gradient. We perform 3 global iterations. The long-wavelength gradient at the first iteration is displayed in Figure 5-21. We compare it with the true velocity perturbation in Figure 5-22. The "true model" here means the FWI model in Figure 5-8. The true velocity perturbation thus is the difference between the FWI model and the initial model (Figure 5-19). We can observe that the gas pocket at $x = 2$ km is better retrieved than in the FWI case. But the problem of density still exists as visible below the water-bottom layer.

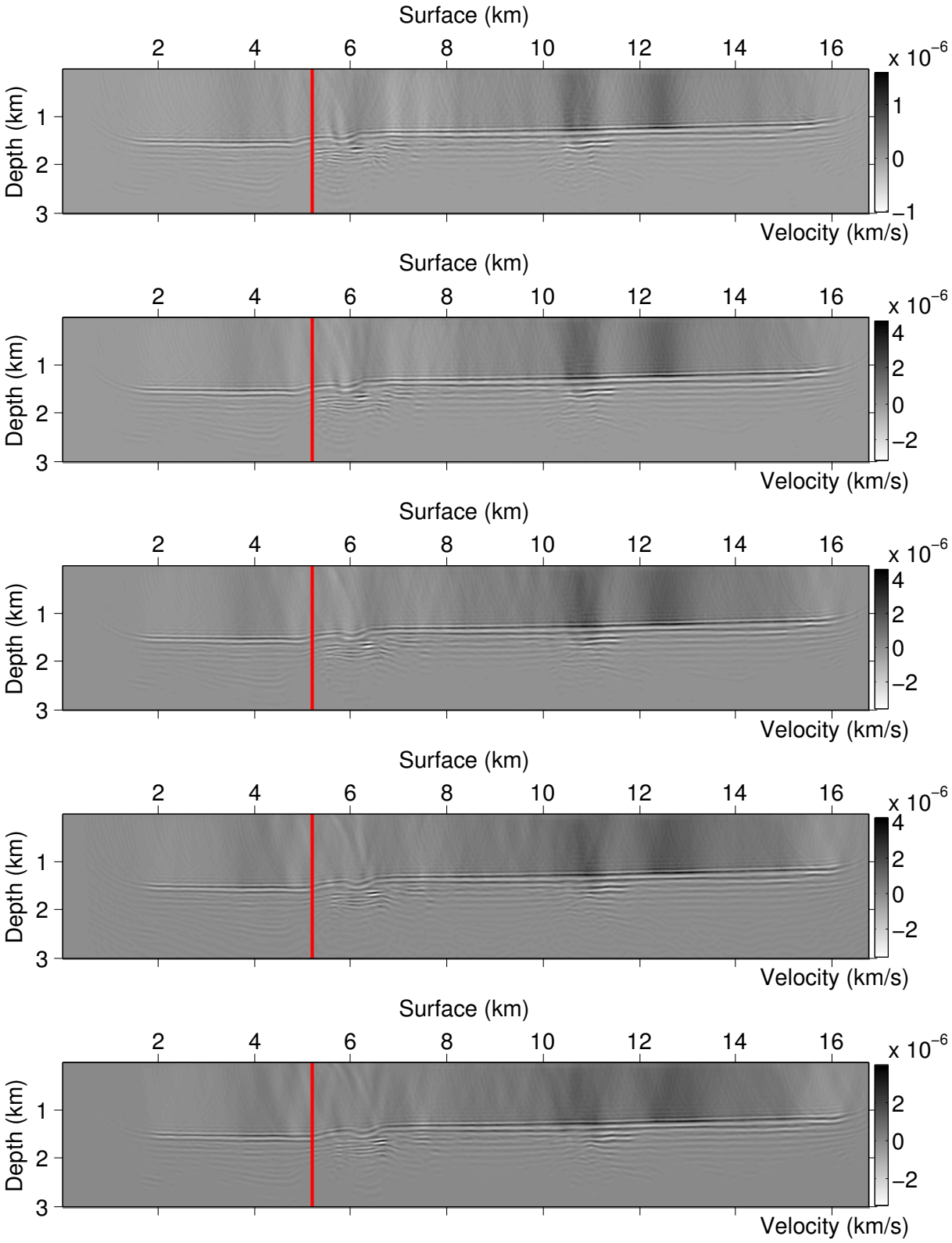


Figure 5-16: From top to bottom: the migrated image using data with offset range of [150–275] m, [275–400] m, [400–525] m, [525–650] m and [650–775] m respectively. Note that from top to bottom, the migrated images shift horizontally. The red vertical lines are fixed at $x = 5.2$ km in these images.

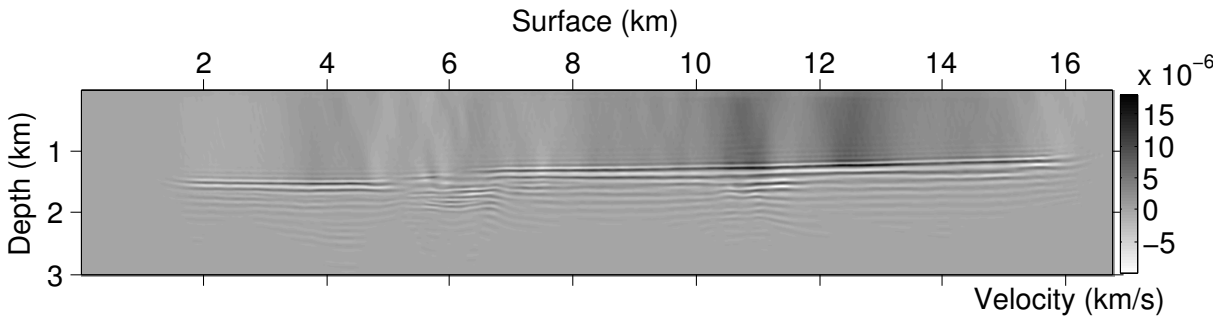


Figure 5-17: The migrated image using data with offset range of [150–775] m.

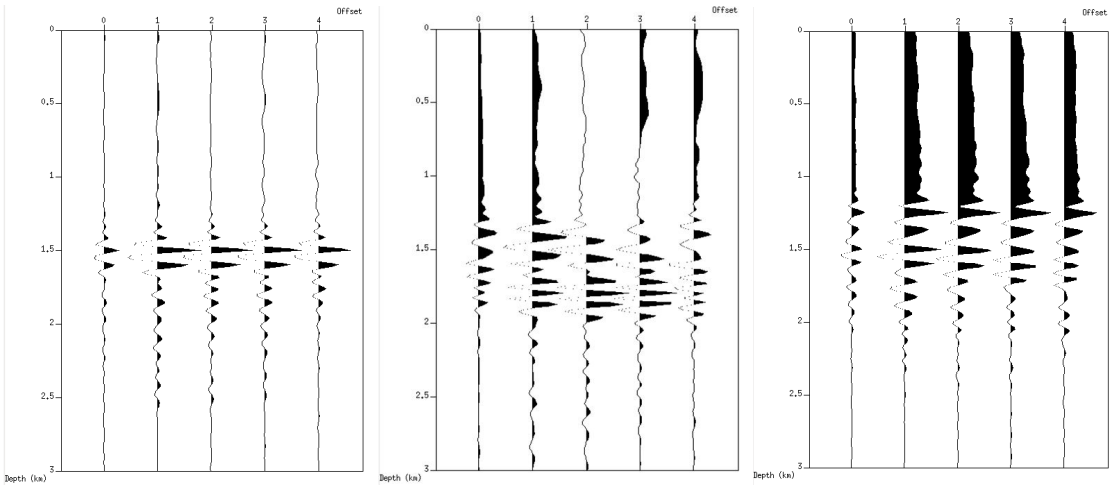


Figure 5-18: Vertical slices taken in Figure 5-16. From left to right, $x = 3$ km, 6 km and 11 km respectively.

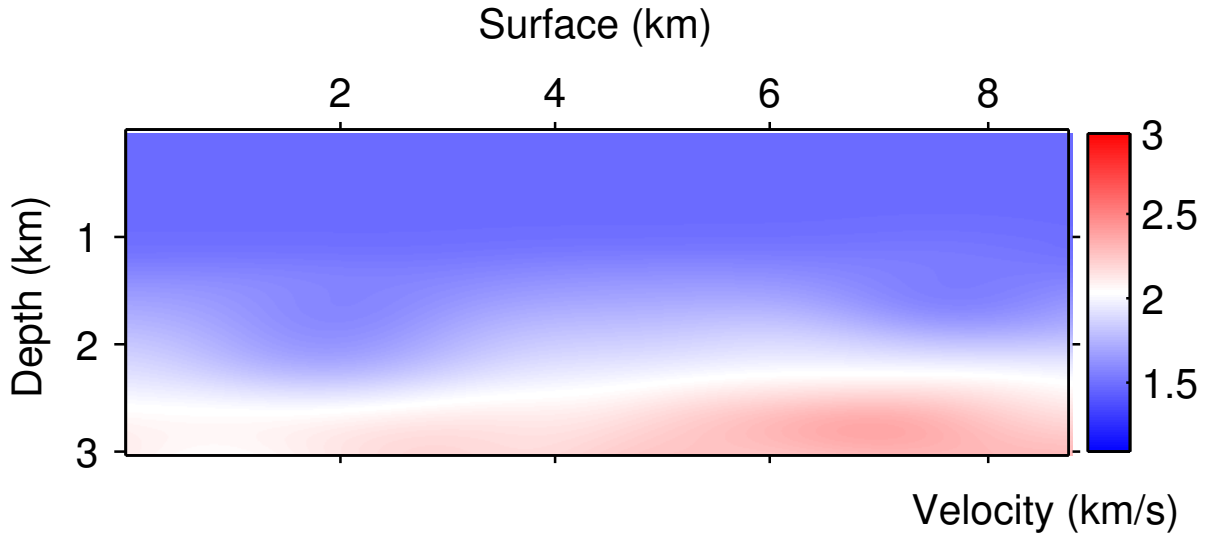


Figure 5-19: The initial velocity model for inversion.

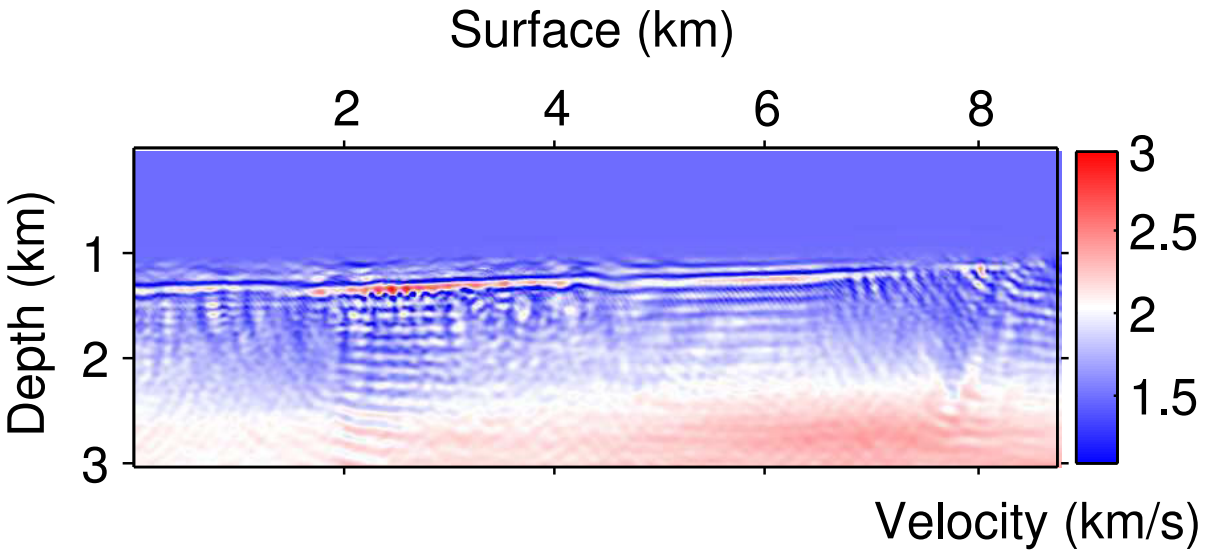


Figure 5-20: FWI result after 100 iterations.

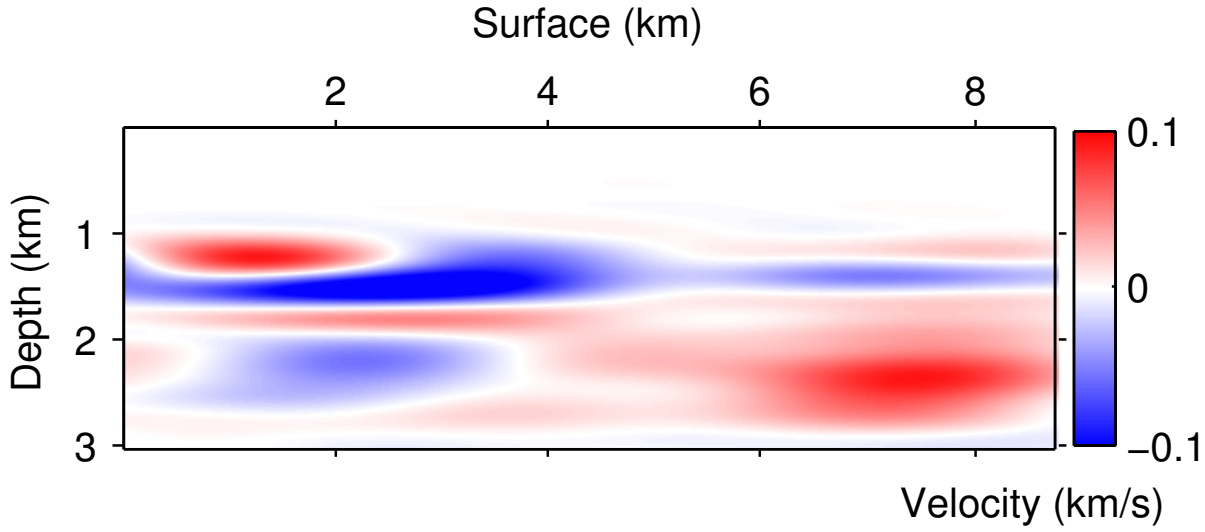


Figure 5-21: The model perturbation of the first global iteration with DWI.

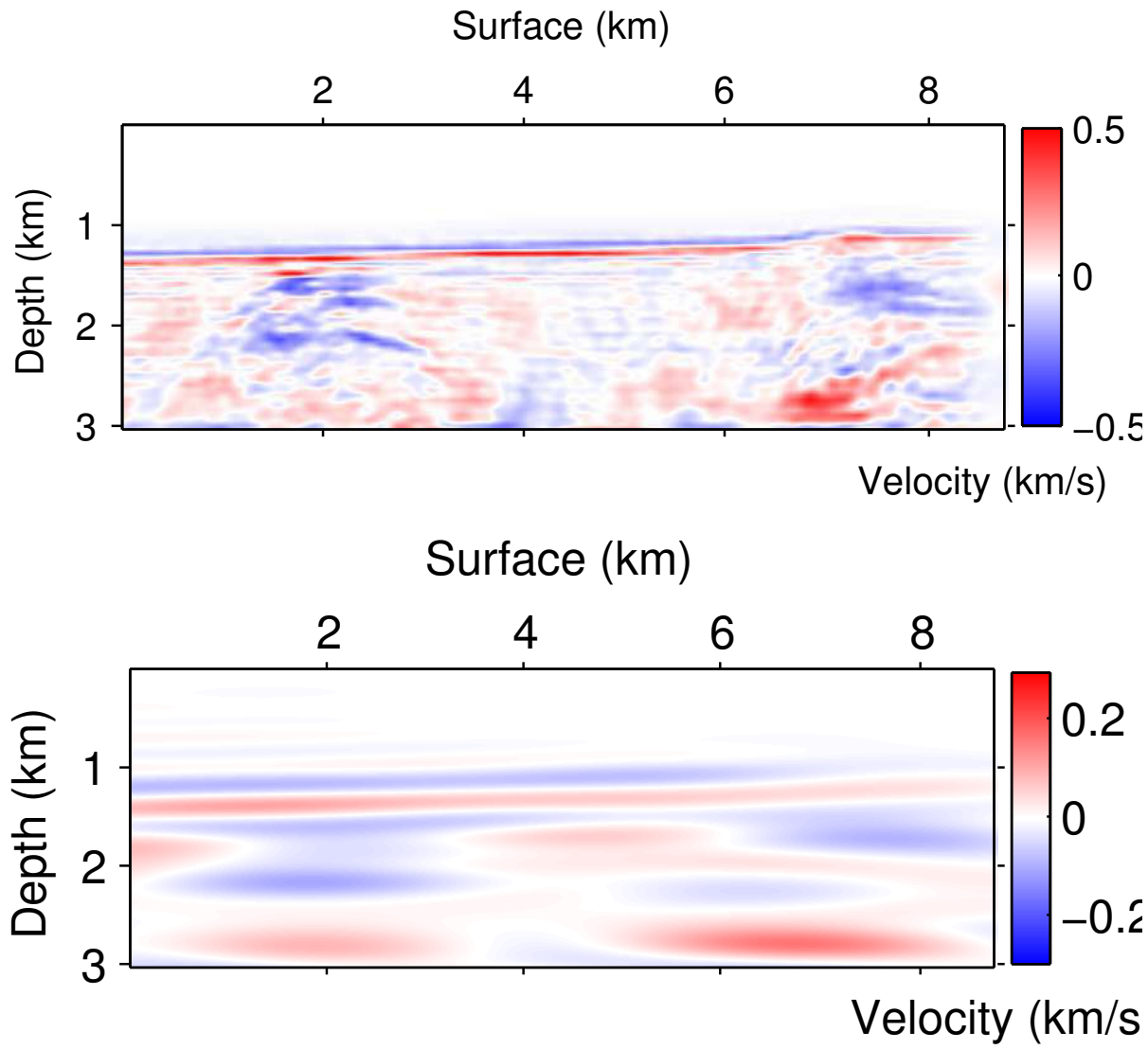


Figure 5-22: (Top) The difference between the FWI model (Figure 5-8) and the initial model (Figure 5-19). (Bottom) The smoothed version of the image on the top.

The long-wavelength velocity model after 3 global iterations is displayed in Figure 5-23 on the top, and the FWI result after 100 iterations, starting from this model, is displayed in Figure 5-23 on the bottom. The result is not yet satisfactory. Compared to conventional FWI, one improvement may be found in the gas pocket at $x = 2$ km, however, the bottom of the water layer pose more problems in this case.

5.4 Discussion and conclusion

In this chapter, we apply the FWI and DWI on the Brunei real data set and show that compared to the conventional FWI, DWI better recovers the gas pocket. However, due to the acoustic approximation with constant density, the data is not correctly simulated, especially for dealing with the data amplitude at the water-bottom layer (Figure 5-24). Attenuation with low quality factor is certainly visible due to the presence of gas pockets [Operto et al., 2013]. Further investigations are needed to incorporate elastic, density and attenuation effect. Besides, we had wanted to consider more frequencies in the data to get a more focused reflectivity in the short-update step of the inversion, which is favorable for the long-wavelength update of the model. However, adding higher frequencies in the data also means a finer spatial and temporal sampling, yielding larger model which is beyond our actual computational capacity.

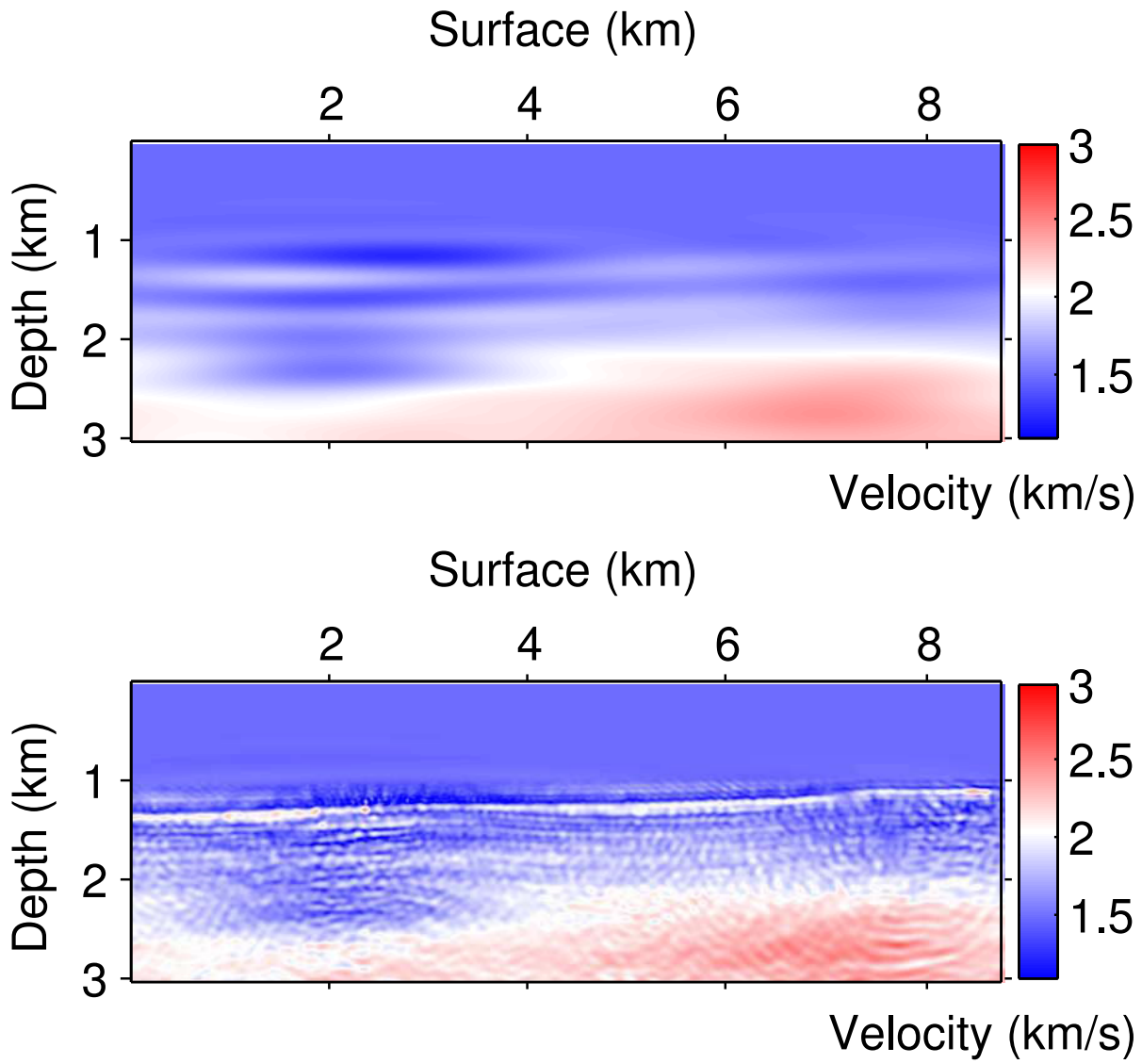


Figure 5-23: Long-wavelength model obtained with DWI (top), and FWI result after 100 iterations, starting from the model on the top (bottom).

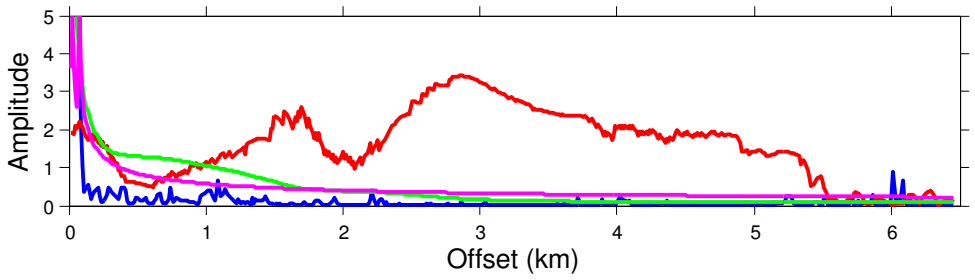


Figure 5-24: Amplitude variation with offset for constant density case. Blue: observed data; red: calculated data; green: the ratio of calculated data and observed data; magenta: $1/\sqrt{t}$ as a reference.

Chapter 6

Conclusions and perspectives

Contents

Résumé du chapitre 6

Ce chapitre est consacré aux conclusions et aux perspectives de la thèse.

L'inversion des formes d'ondes (Full Waveform Inversion, FWI) est largement utilisé pour l'imagerie sismique. Cependant, la fonction objective de FWI est fortement non linéaire par rapport aux paramètres du modèle et présente beaucoup de minima locaux [Bunks et al., 1995]. Ces minima locaux empêchent les techniques basées sur le gradient de trouver le minimum global si le modèle initial est loin de la solution globale. Si un ensemble de données contient principalement des réflexions, les méthodes basées sur le gradient souvent n'arrivent pas à reconstruire les grandes longueurs d'ondes du modèle [Pratt et al., 1996]. Dans cette thèse, nous avons présenté et testé une nouvelle méthode d'inversion basée sur la séparation du gradient après la décomposition du champ d'onde (Decomposition-based Waveform Inversion, DWI). L'objectif est de reconstruire les grandes longueurs d'ondes du modèle de vitesse. La recherche est principalement axée sur trois aspects, comme décrit dans la suite.

Dans le chapitre 2, nous avons revu les formulations classiques de FWI, y compris la fonction objective, les méthodes de modélisation numérique pour la propagation des ondes, le calcul du gradient et du Hessien de la fonction objective, et les méthodes de mise à jour du modèle de vitesse. L'analyse de la résolution du gradient de FWI indique que la résolution dépend des différents types d'ondes et de la configuration d'acquisition. Les ondes plongeantes, les transmissions liée aux réflexions et aux réfractons, et les multiples particulièrement aident à reconstruire les grandes longueurs d'ondes du modèle. Les inversions avec différents paramètres d'acquisition indiquent que les données de grands offsets et de basses fréquences sont favorables pour la mise à jour des grandes longueurs d'ondes du modèle. Les analyses effectuées dans ce chapitre et la littérature nous ont amenés à proposer la nouvelle méthode d'inversion. Pour mieux reconstruire les grandes longueurs d'ondes du modèle, dans le chapitre 3, nous avons proposé une méthode d'inversion basée sur la décomposition. Cette approche consiste à séparer le gradient de FWI en une partie de grandes longueurs d'ondes et une partie de courtes longueurs d'ondes, puis de les inverser séparément et alternativement. Cette séparation du gradient est obtenue par la décomposition des champs d'ondes en une partie descendante et une partie montante. Au lieu de mettre des différents poids dans le gradient [Tang and Lee, 2013], où le poids est difficile à déterminer, nous proposons d'inverser ces deux parties séparément. L'inversion est effectuée avec un algorithme imbriqué. La méthode est illustrée au travers d'un modèle de couche 1D.

Nous avons également comparé DWI avec d'autres méthodes d'inversion basée sur la réflexion. Nous comparons en particulier avec la méthode proposée par [Xu et al., 2012]. La comparaison à travers le modèle de couche 1D indique que les mises à jour du gradient pour ces deux méthodes sont très similaires. La différence réside dans la façon de calculer les champs d'ondes partiels. En outre, comme la méthode de [Xu et al., 2012] est basée sur la modélisation linéarisé, elle suppose que les perturbations de vitesse sont petites et ne considère que les réflexions primaires. En revanche, comme la FWI classique, DWI résout l'équation des ondes non linéarisée, donc les autres types d'ondes pourraient être pris en considération. Comparée à d'autres méthodes qui filtrent le gradient pour obtenir les grandes longueurs d'ondes, DWI récupère les grandes longueurs d'ondes automatiquement. Ceci est un avantage quand on connaît peu sur le modèle. Une autre méthode

similaire est proposé par [Zhou et al., 2015]. Cette approche consiste à combiner la partie tomographique dans le gradient de FWI obtenu par les réflexions avec le gradient obtenu par les ondes plongeantes pour extraire les grandes longueurs d'ondes du modèle. Cette méthode présente l'avantage d'être plus rapide que DWI, mais elle est plus sensible à la réflectivité. Elle fonctionne de manière optimale pour la réflectivité avec une fréquence infinie, mais en pratique, en raison des données de bande limitée, la réflectivité est également limité en fréquence.

La décomposition du gradient est obtenue par la décomposition du champ d'ondes. Nous avons présenté et comparé trois méthodes pour décomposer le champ d'ondes. Ces méthodes sont illustrées à travers un champ d'ondes qui se propage dans le modèle de couche 1D. Les avantages et les inconvénients de chaque méthode sont également détaillée. La méthode par la transformation de Fourier est la plus populaire, et a été largement utilisée pour supprimer des artefacts de haute fréquence dans les images migrées, mais avec cette méthode les basses fréquences sont filtrées dans le résultat final. La méthode de Poynting est la plus rapide car elle peut être calculée au cours de la propagation des ondes, mais elle souffre du problème d'artefacts numériques forts. La méthode de décomposition par les curvelet donne le meilleur résultat car il peut prendre en compte les différentes directions en même temps, avec le coût de calcul le plus élevé des trois. Cependant, contrairement à la méthode de Fourier qui nécessite de garder en mémoire l'ensemble des champs d'ondes, la méthode de curvelet est facilement parallélisable parce que la décomposition est effectuée séparément pour chaque pas de temps.

Dans le chapitre 4, nous appliquons FWI, FWI multi-échelles et DWI sur deux modèles synthétiques et montrons que par rapport à la FWI conventionnel et la FWI multi-échelle, la DWI est plus robuste face à l'absence de basses fréquences dans la source et à des erreurs importantes dans le modèle initial. Nous avons également analysé l'influence de plusieurs paramètres clés lors de l'inversion, comme l'application de l'inversion itérative et non linéaire en utilisant des données de hautes fréquences et zéro offset dans l'étape de mise à jour des courtes longueurs d'ondes. Grâce à cette analyse, nous pouvons conclure que lorsque ces paramètres clés sont tous appliqués dans l'inversion, la mise à jour des grandes longueurs d'ondes est plus proche de la vraie perturbation du modèle de vitesse que dans les cas où les paramètres clés sont partiellement appliqués.

Dans le chapitre 5, nous avons appliqué FWI et DWI sur un jeu de données réelles de Brunei. Nous montrons que, avec l'hypothèse acoustique, densité constante et milieu non atténuant, il est difficile de simuler les formes d'onde dans ce jeu de données réelles, en particulier en ce qui concerne le comportement de l'amplitude avec offset. Avec ces hypothèses limitées, ni FWI ni DWI peuvent fournir un résultat satisfaisant. Cependant, DWI donne un résultat légèrement meilleur pour retrouver les poches de gaz.

A part le problème de l'amplitude, pour DWI, la variation de densité influence également la réflectivité, ce qui est un facteur important pour DWI. Vue dans le chapitre 4, une petite différence dans la réflectivité peut provoquer une grande différence dans le résultat de l'inversion.

D'autres investigations sur ce sujet vont certainement aborder l'utilisation du modèle visco-élastique. En outre, des recherches supplémentaires seraient nécessaires pour analyser l'inversion multi-paramétrique avec la méthode DWI. Modèles synthétiques doivent être validés avant de traiter les données réelles.

Pour DWI, l'inversion itérative est effectuée pour la mise à jour des courtes longueurs

d'onde pour ajuster les données. Cette étape est très coûteuse en temps de calcul. Toutefois, cette étape itérative est importante car elle facilite la mise à jour des grandes longueurs d'onde du modèle. Cette conclusion est compatible avec l'importance d'introduire la migration itérative dans l'analyse de vitesse par migration (Migration-based Velocity Analysis, MVA) [Symes, 2015, Lameloise et al., 2015]. Pour les futures études, nous aimerions mettre en œuvre cette étape dans le domaine pseudo-temps pour éviter la calcul à chaque itération globale [Plessix, 2012, Brossier et al., 2015].

Notons que pour DWI, la longueur du pas pour la mise à jour des grandes longueurs d'ondes est définie au début. Ainsi, la décroissance de la fonction objective n'est pas assurée au cours des itérations. Si nous utilisons l'interpolation quadratique pour déterminer la longueur du pas à chaque itération, nous avons besoin d'effectuer deux problèmes directs supplémentaires, et ici le problème direct correspond à l'inversion itérative, donc rend le calcul très cher. Si cette étape d'inversion itérative pourrait être réalisée dans le domaine pseudo-temps, le coût de calcul pourrait être largement réduit.

Par ailleurs, dans cette thèse, DWI principalement utilise les réflexions. Cependant, en théorie, DWI n'est pas limitée aux réflexions, ce qui est différent des autres méthodes d'inversion basées sur la réflexion, et pourrait intégrer d'autres types d'ondes en même temps. Pour les futures études, nous tenons à combiner ces différents types d'ondes et voir si la combinaison pourrait faciliter l'inversion. [Biondi et al., 2012] introduisent l'offset en sous-sol pour atténuer la non linéarité du problème. Nous aimerions étudier de plus près sur ce point, en introduisant l'offset en sous-sol pour DWI.

Afin d'atténuer la grande dépendance vis-à-vis de l'amplitude des données, nous aimerions tester d'autres fonctions objectives, qui sont moins sensibles à l'amplitude, telles que la fonction objective basée sur la corrélation [van Leeuwen and Mulder, 2008] et fonction objective normalisée.

Enfin, avec le développement des ordinateurs de haute performance, l'application de la méthode de décomposition basée sur le curvelet aux données réelles pourrait devenir bientôt possible. L'élément important est la corrélation dans le domaine des curvelets et non pas en temps pour chacun des directions.

6.1 Conclusions

Full Waveform Inversion (FWI) is widely used for seismic imaging. However, the objective function of FWI is highly nonlinear with respect to the model parameters and has many local minima [Bunks et al., 1995]. These local minima prevent gradient-based techniques from finding the global minimum if the initial model is far from the global solution. If a dataset mainly contains reflections, the gradient-based methods does not recover the long wavelengths of the model [Pratt et al., 1996]. In this thesis, I have presented and tested a new inversion method based on the gradient separation after wavefield decomposition. The objective is to retrieve the long wavelengths of the velocity model. The research mainly focused on three aspects, as described in the following.

6.1.1 Basic FWI theory and resolution analysis

In Chapter 2, we have reviewed the classical formulations of FWI, including the objective function, the methods for the numerical modeling of wavefield propagation, the computation of the gradient and the Hessian of the objective function, and the velocity model update methods. The resolution analysis of the FWI gradient indicates that the resolution depends on the types of waves and on the acquisition setup. The diving waves, reflection-related transmission, refraction-related transmission and multiples particularly help to recover the long-wavelength part of the model. The inversions with different acquisition parameters indicate that long-offset and low-frequency data are favorable for long-wavelength update of the model. The analyses performed in this chapter and the revised literature brought us to propose the DWI inversion method.

6.1.2 The DWI method

To better recover the long-wavelength part of the model, in Chapter 3, we have proposed a Decomposition-based Waveform Inversion (DWI) method. The method consists in separating the gradient of FWI into a long-wavelength part and a short-wavelength part, and then inverting them separately. This separation of gradient is based on the decomposition of the wavefield into their down-going and up-going part. By introducing this wavefield decomposition into the gradient of FWI, a recombination of them gives the long- and short-wavelength parts of the gradient. Instead of setting different weights ([Tang and Lee, 2013]), for which the weight is difficult to determine, we propose to separately invert these two parts. The inversion is performed using a nested algorithm. The method is illustrated through a 1D layer model. The decomposed gradients using the 1D layer model show that one part of the gradient of FWI indicates the positions of reflectors and thus provides the short wavelengths of the model, and the other part indicates the wavepath that connecting the surface and the reflector and provides the long wavelengths of the model.

We have also compared DWI with other reflection-based waveform inversion methods. We compare particularly with the method proposed by [Xu et al., 2012]. The comparison through the 1D layer model indicates that the gradient updates for these two methods are very similar. The difference resides in the way to compute the partial wavefields. Besides, as the method of [Xu et al., 2012] is based on the linearized modeling, it assumes that

the velocity perturbations are small and the data contain only primaries. While similarly to classical FWI, DWI resolves the nonlinear wave equation; therefore refractions, diving waves and multiples could be included in the dataset. Compared to others methods that filter the gradient to get the long wavelengths, DWI retrieves the long-wavelength part automatically. This is an advantage when we know little about the model.

Another similar method is proposed by [Zhou et al., 2015]. This method consists in combining the tomographic part in the FWI gradient using reflection data with the gradient using diving waves to retrieve the long wavelengths of the model. The tomographic part in the FWI gradient with reflection is the same as we use. Instead of decomposing wavefields, the authors first compute the gradient of FWI in the model with reflectivity $m_0 + \delta m$, and then they compute the gradient of FWI in the smooth model m_0, and the difference of these two cancels the migration part in the FWI gradient and provides the tomographic part. This method has the advantage of being faster than the wavefield decomposition-based method, however it is more sensitive to the reflectivity. It works optimally for reflectivity with infinite frequency, but in practice, due to the band-limited data, the reflectivity is also band-limited. This band-limited reflectivity eventually change the kinematics of the model. Thus, the migration part in the FWI gradient calculated in the full model and in the smooth model may be different and the difference of these two gradients may contain residual short wavelengths.

The decomposition of the gradient is based on the wavefield decomposition. We have presented and compared three methods to decompose the wavefield into their one way component. These methods are illustrated through a wavefield that propagates in the 1D layer model. The advantages and the disadvantages of each method are also detailed. The Fourier transform method is the most popular one, and was widely used for suppressing high frequency artefacts in migrated images, but low frequencies are filtered in the final results. The Poynting method is the fastest one as it could be computed during the wave propagation, but it suffers from strong numerical artefacts. The curvelet decomposition method yields the best result as it can take into account different directions at the same time, with the expense of the highest computational cost of the three. However, differently from the 2D Fourier transform method which requires to keep in memory the whole propagated wavefield, the curvelet decomposition method is more easily parallelizable because the decomposition is carried out for each snapshot separately.

6.1.3 Applications to synthetic models and real data

In chapter 4, we apply FWI, multiscale FWI and DWI on two synthetic models and show that compared to the conventional FWI and the multiscale FWI, DWI is more robust with respect to the absence of low frequencies in the source wavelet and to large errors in the initial model. We have also analyzed the influence of several key parameters during the inversion, such as performing the nonlinear iterative inversion in the short-wavelength update step, and using high frequency and zero-offset data in the short-wavelength update step. Thanks to this analysis, we can conclude that when these key parameters are all applied in the inversion, the update of the long-wavelength part is closer to the true velocity perturbation than in cases where the key parameters are partly applied. From the point of view of computational cost, DWI is computational expensive, as iterative inversion is performed at the short-wavelength update step to fit the data. However this iterative step

is important, since it facilitates the long-wavelength update of the model. This conclusion is consistent with introducing iterative migration in MVA [Symes, 2015, Lameloise et al., 2015].

In chapter 5, we have applied FWI and DWI on the Brunei real data set. We show that with the acoustic and the constant density and non attenuative assumption, it is difficult to simulate the waveforms present in this real dataset, especially regarding the behavior of the amplitude with offset. With this limiting assumption, neither FWI nor DWI could provide a satisfactory result. However, DWI yields a slightly better result to retrieve the gas pockets.

6.2 Perspectives

6.2.1 Physics of the earth

For the wave propagation, we assume in this thesis the acoustic approximation, with constant density and no attenuation. This assumption is very limiting when dealing with real datasets, as they are better interpreted by visco-elastic modeling. The inadequacy of the acoustic assumption with constant density and without attenuation to interpret real data and especially to interpret the data amplitude, is observed in Chapter 5. We think that attenuation is a major factor. Further investigations on this topic will certainly address the use of visco-elastic model. Moreover, further research would be needed to analyze multiparameter inversion with the DWI method. Synthetic models need to be validated before dealing with real data.

6.2.2 Role of density

We have investigated the influence of density on the data amplitude in Chapter 5. Apart from the data amplitude issue, for DWI, the variation of density also influence the reflectivity, which is an important factor for DWI. We analyze this effect through a 1D simple model (Figure 6-1). The true velocity for the first and the third layer is 2 km/s and for the second layer is 2.5 km/s. The true density for the first and the third layer is 1000 kg/m^3 and for the second layer is 1500 kg/m^3. The initial velocity is homogeneous at 2 km/s. We have tested three density models. The first one is the true density model. The second one is the homogeneous model at 1000 kg/m^3, and for the third model, the density for the second layer is 2000 kg/m^3 and for the rest of the model is 1000 kg/m^3. The result after running 20 FWI iterations with fixed density parameters are shown in Figure 6-2. We observe that the density error is leaked into the velocity update. For the initial density that is smaller than the true density, the velocity is overestimated to compensate the density error, and for the initial density that is larger than the true density, the velocity is underestimated to correct the overestimation of density. Besides, the sign of the reflectivity is changed. As we have analyzed in chapter 4, DWI is quite sensitive to the reflectivity. A small difference in the reflectivity may cause a large difference in the inversion result. The influence of the density on the inversion, especially on DWI needs further investigations.

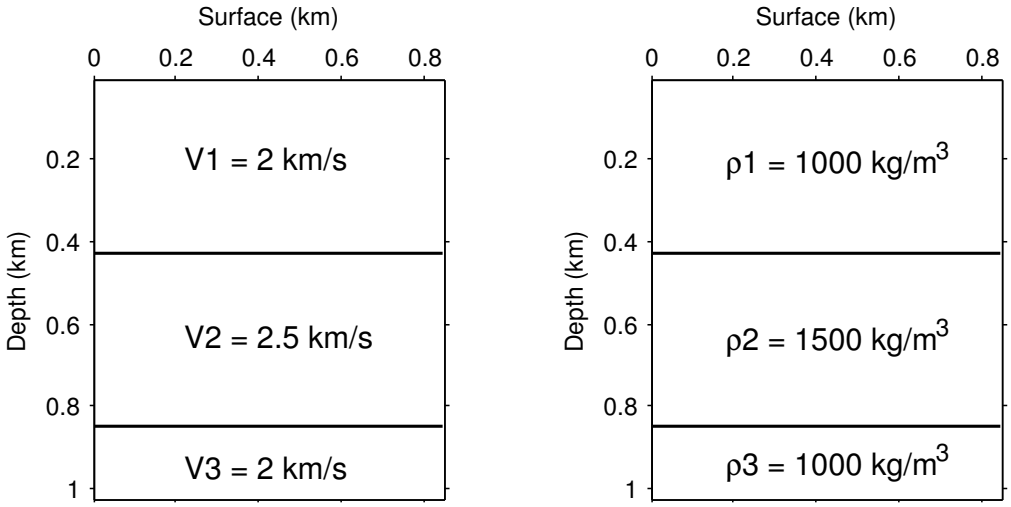


Figure 6-1: True velocity model (left) and the true density model (right).

6.2.3 Inversion techniques

For DWI, iterative inversion is performed at the short-wavelength update step to fit the data. This step is quite computational expensive. However this iterative step is important, since it facilitates the long-wavelength update of the model. For further tests, we would like to implement this step in the pseudo-time domain to avoid the re-computation of the short-wavelength model at each step [Plessix, 2012, Brossier et al., 2015]. It should be noted that for DWI, in the long-wavelength update step, the step length of the gradient descent method is defined at the beginning. Thus the decrease of the objective function is not assured during all the steps of the inversion. If we use the quadratic interpolation to determine the step length at each iteration, we need to carry out two extra direct problems, however here the direct problem corresponds to the iterative inversion, thus this is quite computational expensive. If this iterative inversion step could be performed in the pseudo-time domain, the computational cost could be largely reduced.

Besides, in this thesis, DWI is mainly used for reflections. However, in theory, DWI is not limited to work with reflections only, which is different from other reflection-based waveform inversion methods, and could incorporate other types of waves at the same time, such as diving waves and diffractions. For future research, we would like to combine these different types of waves and to see if the combination could facilitate the inversion.

[Biondi et al., 2012] introduce the subsurface offset to the reflection-based waveform inversion to mitigate the nonlinearity of the problem. We would like to investigate more this point, by introducing the subsurface offset into DWI.

In order to mitigate the problem of data amplitude, we would like to test other objective functions, that are less sensitive to amplitude, such as the correlation-based objective function [van Leeuwen and Mulder, 2008] and normalized objective function. We need to investigate on the feasibility of combining the new objective functions with wavefield decomposition.

Finally, with the development of high performance computers, the application of the curvelet-based decomposition method on real datasets could become soon feasible. To further reduce the computational cost, we would like to test the idea of correlating the

decomposed wavefields in the curvelet domain before the inverse curvelet transform to compute the gradient, rather than performing the inverse curvelet transform on each snapshot and then correlating the whole wavefield in the space domain.

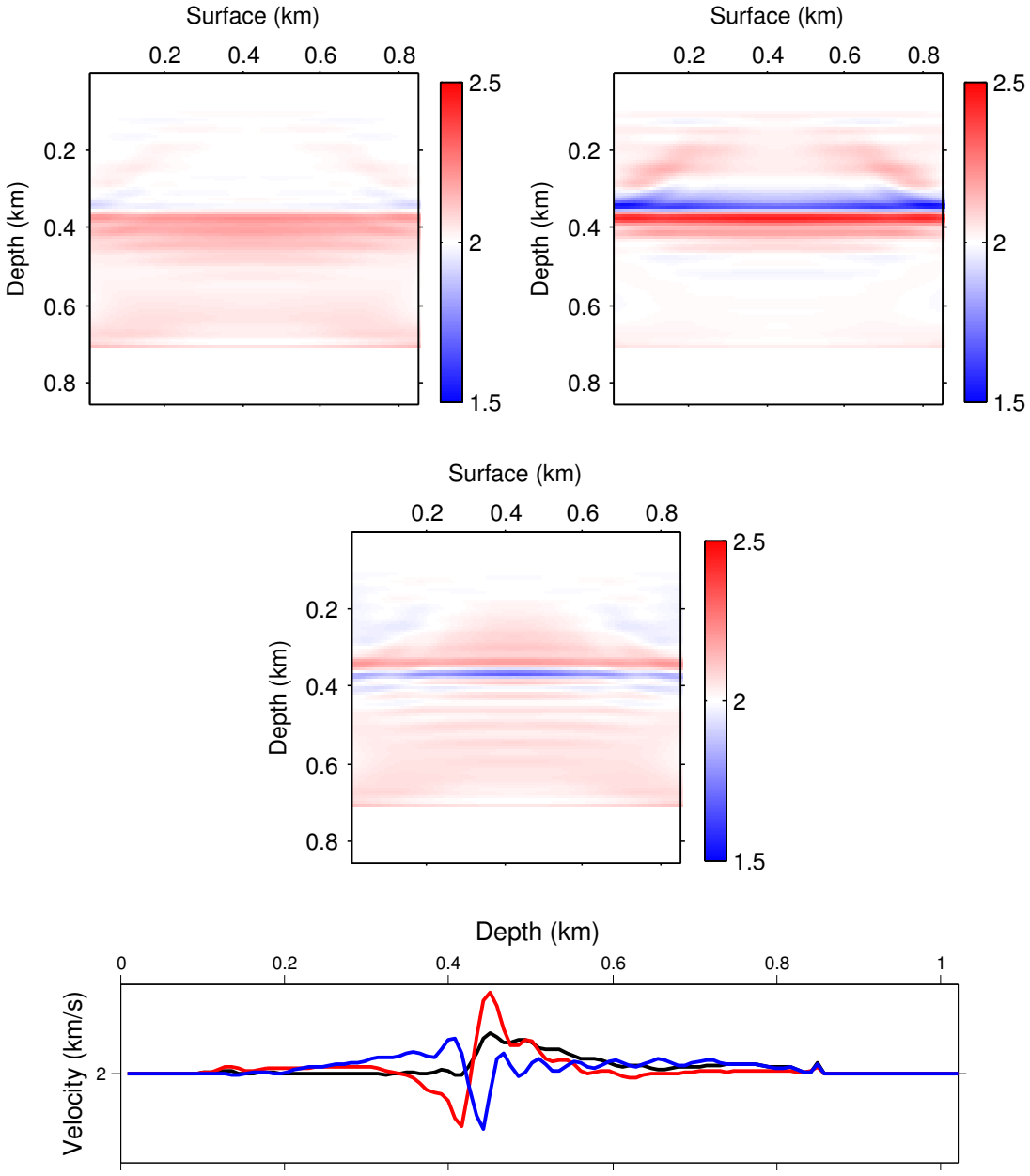


Figure 6-2: FWI results after running 20 iterations using the true density model (top left), smaller density model (top right) and larger density model (middle). Vertical slices taken at $x = 0.425$ km for the above three FWI models (black: true density model; red: smaller density model; blue: larger density model. During the inversion, the density model is fixed.

Bibliography

[Alkhalifah, 2015] Alkhalifah, T. (2015). Scattering-angle based filtering of the waveform inversion gradients. *Geophysical Journal International*, 200(1):363–373. 80

[Alkhalifah and Wu, 2015] Alkhalifah, T. and Wu, Z. (2015). The natural combination of full and image-based waveform inversion. *Geophysical Prospecting*. 34

[Allen, 1978] Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, 68(5):1521–1532. 130

[Amundsen, 1991] Amundsen, L. (1991). Comparison of the least-squares criterion and the Cauchy criterion in frequency-wavenumber inversion. *Geophysics*, 56(12):2027–2035. 33

[Arbelaez et al., 2011] Arbelaez, P., Maire, M., Fowlkes, C., and Malik, J. (2011). Contour detection and hierarchical image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(5):898–916. 128, 132

[Baysal et al., 1984] Baysal, E., Kosloff, D. D., and Sherwood, J. (1984). A two-way nonreflecting wave equation. *Geophysics*, 49(2):132–141. 81

[Baysal et al., 1983] Baysal, E., Kosloff, D. D., and Sherwood, J. W. (1983). Reverse time migration. *Geophysics*, 48(11):1514–1524. 29

[Berenger, 1994] Berenger, J.-P. (1994). A perfectly matched layer for the absorption of electromagnetic waves. *Journal of computational physics*, 114(2):185–200. 44

[Beylkin, 1985] Beylkin, G. (1985). Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform. *Journal of Mathematical Physics*, 26(1):99–108. 29

[Billette and Lambaré, 1998] Billette, F. and Lambaré, G. (1998). Velocity macro-model estimation from seismic reflection data by stereotomography. *Geophysical Journal International*, 135:671–690. 28, 59

[Biondi, 2006] Biondi, B. (2006). *3D seismic imaging*. Number 14. Society of Exploration Geophysicists. 24, 28

[Biondi et al., 2012] Biondi, B., Almomin, A., et al. (2012). Tomographic full waveform inversion (tfwi) by combining full waveform inversion with wave-equation migration velocity anaylisis. In *2012 SEG Annual Meeting*. Society of Exploration Geophysicists. 26, 34, 152, 156

[Bishop et al., 1985] Bishop, T., Bube, K., Cutler, R., Langan, R., Love, P., Resnick, J., Shuey, R., Spindler, D., and Wyld, H. (1985). Tomographic determination of velocity and depth in laterally varying media. *Geophysics*, 50(6):903–923. 28, 59

[Bleistein, 1987] Bleistein, N. (1987). On the imaging of reflectors in the earth. *Geophysics*, 52(7):931–942. 29

[Bozdağ et al., 2011] 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. 33

[Brenders and Pratt, 2007] Brenders, A. J. and Pratt, R. G. (2007). Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model. *Geophysical Journal International*, 168:133–151. 61

[Brossier et al., 2009] Brossier, R., Operto, S., and Virieux, J. (2009). Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion. *Geophysics*, 74(6):WCC105–WCC118. 39, 44, 48, 60

[Brossier et al., 2010] Brossier, R., Operto, S., and Virieux, J. (2010). Which data residual norm for robust elastic frequency-domain full waveform inversion? *Geophysics*, 75(3):R37–R46. 33

[Brossier et al., 2015] Brossier, R., Operto, S., and Virieux, J. (2015). Velocity model building from seismic reflection data by full-waveform inversion. *Geophysical Prospecting*, 63(2):354–367. 26, 34, 35, 36, 55, 80, 126, 152, 156

[Bunks et al., 1995] Bunks, C., Saleck, F., Zaleski, M., and Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473. 25, 31, 32, 39, 59, 60, 101, 104, 150, 153

[Candes et al., 2006] Candes, E., Demanet, L., Donoho, D., and Ying, L. (2006). Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 5(3):861–899. 89

[Candes and Demanet, 2004] Candes, E. J. and Demanet, L. (2004). The curvelet representation of wave propagators is optimally sparse. *arXiv preprint math/0407210*. 90

[Candes et al., 1999] Candes, E. J., Donoho, D. L., et al. (1999). *Curvelets: A surprisingly effective nonadaptive representation for objects with edges*. DTIC Document. 89

[Castellanos et al., 2015] 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):718–742. 59

[Causse, 2002] Causse, E. (2002). Avoiding pitfalls of least-squares inversion by using the energy-flux error criterion. *Geophysical prospecting*, 50(3):317–327. 33

[Cerjan et al., 1985] Cerjan, C., Kosloff, D., Kosloff, R., and Reshef, M. (1985). A nonreflecting boundary condition for discrete acoustic and elastic wave equations. *Geophysics*, 50(4):705–708. 44

[Cerveny et al., 1977] Cerveny, V., Molotkov, I., and Psencik, I. (1977). *Ray method in seismology*. Univerzita Karlova. 42

[Chattopadhyay and McMechan, 2008] Chattopadhyay, S. and McMechan, G. A. (2008). Imaging conditions for prestack reverse-time migration. *Geophysics*, 73(3):S81–S89. 29

[Chauris et al., 2006] Chauris, H. et al. (2006). Seismic imaging in the curvelet domain and its implications for the curvelet design. In *76-th annual international meeting, SEG, expanded abstracts*, volume 25, pages 2406–2409. 15, 88

[Chauris and Nguyen, 2008] Chauris, H. and Nguyen, T. (2008). Seismic demigration/migration in the curvelet domain. *Geophysics*, 73(2):S35–S46. 15, 89, 90

[Chauris and Noble, 2001] Chauris, H. and Noble, M. (2001). Two-dimensional velocity macro model estimation from seismic reflection data by local differential semblance optimization: applications to synthetic and real data sets. *Geophysical Journal International*, 144(1):14–26. 28, 59

[Chauris and Plessix, 2012] Chauris, H. and Plessix, R. (2012). Investigating the differential waveform inversion. In *EAGE Extended Abstracts, Workshops WS06*. 34

[Chavent et al., 1994] Chavent, G., Clément, F., and Gómez, S. (1994). Automatic determination of velocities via migration-based traveltime waveform inversion: A synthetic data example. In *SEG*, pages 1179–1182. Expanded Abstracts. 26, 34, 80

[Chavent and Plessix, 1999] Chavent, G. and Plessix, R.-E. (1999). An optimal true-amplitude least-squares prestack depth-migration operator. *Geophysics*, 64(2):508–515. 48

[Chi et al., 2014] Chi, B., Dong, L., and Liu, Y. (2014). Full waveform inversion method using envelope objective function without low frequency data. *Journal of Applied Geophysics*, 109:36–46. 33

[Choi et al., 2008] Choi, Y., Min, D.-J., and Shin, C. (2008). Two-dimensional waveform inversion of multi-component data in acoustic-elastic coupled media. *Geophysical prospecting*, 56(6):863–881. 38, 42

[Claerbout, 1971] Claerbout, J. (1971). Towards a unified theory of reflector mapping. *Geophysics*, 36(3):467–481. 29, 39, 47, 68

[Claerbout, 1985] Claerbout, J. (1985). *Fundamentals of geophysical data processing*. Pennwell Books, Tulsa, OK. 39, 54

[Claerbout, 1982] Claerbout, J. F. (1982). Imaging the earth's interior. 13, 32

[Clapp, 2005] Clapp, M. L. (2005). *Imaging under salt: illumination compensation by regularized inversion*. PhD thesis, Stanford University. 35

[Clément et al., 2001] Clément, F., Chavent, G., and Gómez, S. (2001). Migration-based traveltime waveform inversion of 2-D simple structures: A synthetic example. *Geophysics*, 66(3):845–860. 109

[Cockburn et al., 2000] Cockburn, B., Karniadakis, G. E., and Shu, C.-W. (2000). *The development of discontinuous Galerkin methods*. Springer. 42

[Crase et al., 1990] Crase, E., Pica, A., Noble, M., McDonald, J., and Tarantola, A. (1990). Robust elastic nonlinear waveform inversion: Application to real data. *Geophysics*, 55(5):527–538. 25, 31, 33, 38, 42, 58

[Devaney, 1982] Devaney, A. (1982). A filtered backpropagation algorithm for diffraction tomography. *Ultrasonic imaging*, 4(4):336–350. 29, 49

[Diaz and Sava, 2012] Diaz, E. and Sava, P. C. (2012). Understanding the reverse time migration backscattering: noise or signal? In *SEG*, pages doi: http://dx.doi.org/10.1190/segam2012–1203.1. Expanded Abstracts. 36

[Donno et al., 2013] Donno, D., Chauris, H., and Calandra, H. (2013). Estimating the background velocity model with the normalized integration method. In *75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013*. 33

[Dumbser and Käser, 2006] Dumbser, M. and Käser, M. (2006). An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes—ii. the three-dimensional isotropic case. *Geophysical Journal International*, 167(1):319–336. 42

[Etgen et al., 2009] Etgen, J., Gray, S. H., and Zhang, Y. (2009). An overview of depth imaging in exploration geophysics. *Geophysics*, 74(6):WCA5–WCA17. 29

[Fichtner, 2010] Fichtner, A. (2010). *Full seismic waveform modelling and inversion*. Springer Science & Business Media. 24, 30, 38, 40

[Fletcher et al., 2005] Fletcher, R. F., Fowler, P., Kitchenside, P., Albertin, U., et al. (2005). Suppressing artifacts in prestack reverse time migration. In *2005 SEG Annual Meeting*. Society of Exploration Geophysicists. 82

[Gardner et al., 1974] Gardner, G., Gardner, L., and Gregory, A. (1974). Formation velocity and density-the diagnostic basics for stratigraphic traps. *Geophysics*, 39(6):770–780. 133

[Gauthier et al., 1986] Gauthier, O., Virieux, J., and Tarantola, A. (1986). Two-dimensional nonlinear inversion of seismic waveforms. *Geophysics*, 51:1387–1403. 25, 31, 58

[Guillaume et al., 2008] Guillaume, P., Lambaré, G., Leblanc, O., Mitouard, P., Le Moigne, J., Montel, J. P., Prescott, T., Siliqi, R., Zhang, X., Zimine, S., et al. (2008). Kinematic invariants: an efficient and flexible approach for velocity model building. In *2008 SEG Annual Meeting*. Society of Exploration Geophysicists. 28, 59

[Guitton and Symes, 2003] Guitton, A. and Symes, W. W. (2003). Robust inversion of seismic data using the Huber norm. *Geophysics*, 68(4):1310–1319. 33

[Ha et al., 2009] Ha, T., Chung, W., and Shin, C. (2009). Waveform inversion using a back-propagation algorithm and a Huber function norm. *Geophysics*, 74(3):R15–R24. 33

[Hennenfent and Herrmann, 2006] Hennenfent, G. and Herrmann, F. J. (2006). Seismic denoising with nonuniformly sampled curvelets. *Computing in Science & Engineering*, 8(3):16–25. 89

[Hicks and Pratt, 2001] Hicks, G. J. and Pratt, R. G. (2001). Reflection waveform inversion using local descent methods: Estimating attenuation and velocity over a gas-sand deposit. *Geophysics*, 66:598–612. 26, 35

[Hu and McMechan, 1987] Hu, L.-Z. and McMechan, G. A. (1987). Wave-field transformations of vertical seismic profiles. *Geophysics*, 52(3):307–321. 82

[Hu et al., 2011] Hu, W., Abubakar, A., Habashy, T., and Liu, J. (2011). Preconditioned non-linear conjugate gradient method for frequency domain full-waveform seismic inversion. *Geophysical Prospecting*, 59(3):477–491. 58

[Huang and Schuster, 2014] Huang, Y. and Schuster, G. T. (2014). Resolution limits for wave equation imaging. *Journal of Applied Geophysics*, 107:137–148. 13, 30, 39, 49, 50, 51, 52, 53

[Jannane et al., 1989] Jannane, M., Beydoun, W., Crase, E., Cao, D., Koren, Z., Landa, E., Mendes, M., Pica, A., Noble, M., Roeth, G., Singh, S., Snieder, S., Tarantola, R., Trezeguet, D., and Xie, M. (1989). Wavelengths of earth structures that can be resolved from seismic reflection data. *Geophysics*, 51(7):906–910. 32, 39, 54

[Jin and Madariaga, 1993] Jin, S. and Madariaga, R. (1993). Background velocity inversion with a genetic algorithm. *Geophysical research letters*, 20(2):93–96. 25, 31

[Jin and Madariaga, 1994] Jin, S. and Madariaga, R. (1994). Nonlinear velocity inversion by a two-step Monte Carlo method. *Geophysics*, 59(4):577–590. 25, 31

[Jin et al., 1992] Jin, S., Madariaga, R., Virieux, J., and Lambaré, G. (1992). Two-dimensional asymptotic iterative elastic inversion. *Geophysical Journal International*, 108(2):575–588. 29

[Kaelin et al., 2006] Kaelin, B., Guitton, A., et al. (2006). Imaging condition for reverse time migration. In *2006 SEG Annual Meeting*. Society of Exploration Geophysicists. 29

[Lailly, 1983] Lailly, P. (1983). The seismic inverse problem as a sequence of before stack migrations. In *Conference on inverse scattering: theory and application*, pages 206–220. Society for Industrial and Applied Mathematics, Philadelphia, PA. 24, 25, 30, 31, 35, 38, 40

[Lambaré et al., 1992] Lambaré, G., Virieux, J., Madariaga, R., and Jin, S. (1992). Iterative asymptotic inversion in the acoustic approximation. *Geophysics*, 57(9):1138–1154. 29, 48

[Lameloise et al., 2015] Lameloise, C.-A., Chauris, H., and Noble, M. (2015). Improving the gradient of the image-domain objective function using quantitative migration for a more robust migration velocity analysis. *Geophysical prospecting*, 63(2):391–404. 29, 72, 96, 126, 152, 155

[Lee and Kim, 2003] Lee, K. H. and Kim, H. J. (2003). Source-independent full-waveform inversion of seismic data. *Geophysics*, 68(6):2010–2015. 45

[Liu et al., 2011] Liu, F., Zhang, G., Morton, S., and Leveille, J. (2011). An effective imaging condition for reverse-time migration using wavefield decomposition. *Geophysics*, 76(1):S29–S39. 36, 80, 82

[Loewenthal et al., 1987] Loewenthal, D., Stoffa, P. L., and Faria, E. L. (1987). Suppressing the unwanted reflections of the full wave equation. *Geophysics*, 52(7):1007–1012. 81

[Lombard and Piraux, 2004] Lombard, B. and Piraux, J. (2004). Numerical treatment of two-dimensional interfaces for acoustic and elastic waves. *Journal of Computational Physics*, 195(1):90–116. 44

[Ma, 2011] Ma, J. (2011). Improved iterative curvelet thresholding for compressed sensing and measurement. *Instrumentation and Measurement, IEEE Transactions on*, 60(1):126–136. 90

[Malinowski et al., 2007] Malinowski, M., Ribodetti, A., and Operto, S. (2007). Multiparameter full-waveform inversion for velocity and attenuation–refining the imaging of a sedimentary basin. In *69th EAGE Conference & Exhibition*. 33

[Marfurt, 1984] Marfurt, K. J. (1984). Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations. *Geophysics*, 49(5):533–549. 25, 32, 38, 42, 44

[Métivier et al., 2012] Métivier, L., Brossier, R., Virieux, J., and Operto, S. (2012). Toward Gauss-Newton and exact Newton optimization for full waveform inversion. In *74th EAGE Conference & Exhibition*. 59

[Métivier et al., 2013] 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. 48, 59

[Miller et al., 1987] Miller, D., Oristaglio, M., and Beylkin, G. (1987). A new slant on seismic imaging: Migration and integral geometry. *Geophysics*, 52(7):943–964. 29, 49

[Moczo et al., 2004] Moczo, P., Kristek, J., and Halada, L. (2004). *The finite-difference method for seismologists*. Comenius University. 25, 32, 38, 42

[Mora, 1987] Mora, P. (1987). Elastic wavefield inversion of reflection and transmission data. *Geophysics*, 53:750–759. 25, 31, 32, 38, 42, 58, 70

[Mora, 1989] Mora, P. (1989). Inversion=migration+tomography. *Geophysics*, 54:1575–1586. 25, 26, 29, 32, 34, 38, 39, 42, 54, 55, 61

[Mulder and Plessix, 2008] Mulder, W. A. and Plessix, R. E. (2008). Exploring some issues in acoustic full waveform inversion. *Geophysical Prospecting*, 56:827–841. 33, 60

[Nemeth et al., 1999] Nemeth, T., Wu, C., and Schuster, G. T. (1999). Least-squares migration of incomplete reflection data. *Geophysics*, 64(1):208–221. 48

[Nocedal, 1980] Nocedal, J. (1980). Updating quasi-Newton matrices with limited storage. *Mathematics of Computation*, 35:773–782. 39, 58, 59, 104

[Operto et al., 2013] Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Metivier, L., and Virieux, J. (2013). A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice. *The Leading Edge*, 32(9):1040–1054. 33, 147

[Operto et al., 2007] Operto, S., Virieux, J., Amestoy, P., L'Excellent, J.-Y., Giraud, L., and Ali, H. B. H. (2007). 3D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study. *Geophysics*, 72(5):SM195–SM211. 43

[Perrone, 2013] Perrone, F. (2013). *Wavefield tomography using image-warping*. PhD thesis, Colorado School of Mines. 16, 98, 100

[Plessix, 2006] Plessix, R. E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167:495–503. 25, 31, 35, 39, 47, 76

[Plessix, 2007] Plessix, R.-E. (2007). A Helmholtz iterative solver for 3D seismic-imaging problems. *Geophysics*, 72(5):SM185–SM194. 42

[Plessix, 2009] Plessix, R.-É. (2009). Three-dimensional frequency-domain full-waveform inversion with an iterative solver. *Geophysics*, 74(6):WCC149–WCC157. 25, 31, 42

[Plessix, 2012] Plessix, R.-É. (2012). A pseudo-time formulation for acoustic full waveform inversion. *Geophysical Journal International*, page ggs056. 126, 152, 156

[Plessix and Mulder, 2004] Plessix, R. E. and Mulder, W. A. (2004). Frequency-domain finite-difference amplitude –preserving migration. *Geophysical Journal International*, 157:975–987. 48, 71, 109

[Polak and Ribiere, 1969] Polak, E. and Ribiere, G. (1969). Note sur la convergence de méthodes de directions conjuguées. *Revue française d'informatique et de recherche opérationnelle, série rouge*, 3(1):35–43. 58

[Pratt, 1999] Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model. *Geophysics*, 64(3):888–901. 25, 31, 38, 45

[Pratt and Chapman, 1992] Pratt, R. G. and Chapman, C. (1992). Traveltime tomography in anisotropic media—ii. application. *Geophysical Journal International*, 109(1):20–37. 28, 59

[Pratt et al., 1998] Pratt, R. G., Shin, C., and Hick, G. (1998). Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. *Geophysical Journal International*, 133(2):341–362. 31, 39, 47, 48, 58, 59

[Pratt et al., 1996] Pratt, R. G., Song, Z. M., Williamson, P., and Warner, M. (1996). Two-dimensional velocity models from wide-angle seismic data by wavefield inversion. *Geophysical Journal International*, 124:323–340. 29, 39, 54, 55, 61, 66, 68, 128, 132, 150, 153

[Pratt and Worthington, 1990] Pratt, R. G. and Worthington, M. (1990). Inverse theory applied to multi-source cross-hole tomography. part 1: acoustic wave-equation method. *Geophysical prospecting*, 38(3):287–310. 25, 31, 32, 38, 42

[Prieux et al., 2013] Prieux, V., Brossier, R., Operto, S., and Virieux, J. (2013). Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field. part 1: imaging compressional wave speed, density and attenuation. *Geophysical Journal International*, page ggt177. 33

[Sava and Biondi, 2004] Sava, P. and Biondi, B. (2004). Wave-equation migration velocity analysis, Part I: Theory. *Geophysical Prospecting*, 52:593–606. 28, 59

[Sava and Fomel, 2006] Sava, P. and Fomel, S. (2006). Time-shift imaging condition in seismic migration. *Geophysics*, 71(6):S209–S217. 29

[Sava and Hill, 2009] Sava, P. and Hill, S. J. (2009). Overview and classification of wavefield seismic imaging methods. *The Leading Edge*, 28(2):170–183. 29

[Sears et al., 2008] Sears, T. J., Singh, S., and Barton, P. (2008). Elastic full waveform inversion of multi-component OBC seismic data. *Geophysical Prospecting*, 56(6):843–862. 61

[Sen and Stoffa, 1991] Sen, M. K. and Stoffa, P. L. (1991). Nonlinear one-dimensional seismic waveform inversion using simulated annealing. *Geophysics*, 56(10):1624–1638. 31

[Shin and Cha, 2008] Shin, C. and Cha, Y. H. (2008). Waveform inversion in the Laplace domain. *Geophysical Journal International*, 173(3):922–931. 33

[Shin et al., 2001] Shin, C., Jang, S., and Min, D.-J. (2001). Improved amplitude preservation for prestack depth migration by inverse scattering theory. *Geophysical prospecting*, 49(5):592–606. 48

[Shin and Min, 2006] Shin, C. and Min, D.-J. (2006). Waveform inversion using a logarithmic wavefield. *Geophysics*, 71(3):R31–R42. 33

[Shipp and Singh, 2002] Shipp, R. M. and Singh, S. C. (2002). Two dimensional full wavefield inversion of wide-aperture marine seismic streamer data. *Geophysical Journal International*, 151:325–344. 30, 31, 64

[Simoncelli et al., 1992] Simoncelli, E. P., Freeman, W. T., Adelson, E. H., and Heeger, D. J. (1992). Shiftable multiscale transforms. *Information Theory, IEEE Transactions on*, 38(2):587–607. 90

[Sirgue, 2003] Sirgue, L. (2003). *Inversion de la forme d'onde dans le domain fréquentiel de données sismiques grand offset*. PhD thesis, Université Paris XI. 13, 39, 47, 54, 59, 60, 61, 64, 109

[Sirgue, 2006] Sirgue, L. (2006). The importance of low frequency and large offset in waveform inversion. In *EAGE*, page A037. Extended Abstracts. 30, 31, 55, 64

[Sirgue et al., 2010] Sirgue, L., Barkved, O., Dellinger, J., Etgen, J., Albertin, U., and Kommedal, J. (2010). Thematic set: Full waveform inversion: The next leap forward in imaging at Valhall. *First Break*, 28(4). 25, 30, 31

[Sirgue and Pratt, 2004] Sirgue, L. and Pratt, R. G. (2004). Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248. 25, 31, 32, 38, 42, 49, 60

[Snieder et al., 1989] Snieder, R., Xie, M., Pica, A., and Tarantola, A. (1989). Retrieving both the impedance contrast and background velocity: A global strategy for the seismic reflection problem. *Geophysics*, 54:991–1000. 26, 35

[Staal et al., 2012] Staal, X., Verschuur, D., et al. (2012). Velocity estimation using internal multiples. *SEG, Soc. Expl. Geophys., Expanded abstracts, Las Vegas*, pages 1–5. 55

[Symes, 2015] Symes, W. (2015). Algorithmic aspects of extended waveform inversion. In *77th EAGE Conference and Exhibition-Workshops*. 29, 72, 96, 126, 152, 155

[Symes, 2008] Symes, W. W. (2008). Migration velocity analysis and waveform inversion. *Geophysical Prospecting*, 56:765–790. 80

[Symes and Carazzone, 1991] Symes, W. W. and Carazzone, J. (1991). Velocity inversion by differential semblance optimization. *Geophysics*, 56:654–663. 28, 59, 120

[Tang and Lee, 2013] Tang, Y. and Lee, S. (2013). Tomographically enhanced full wavefield inversion. In *SEG*, pages 1037–1041. Expanded Abstracts. 26, 34, 35, 55, 70, 80, 81, 150, 153

[Tarantola, 1984] Tarantola, A. (1984). Linearized inversion of seismic reflection data. *Geophysical Prospecting*, 32:998–1015. 24, 25, 30, 31, 32, 35, 38, 40, 42, 45, 46

[Tarantola, 1986] Tarantola, A. (1986). A strategy for nonlinear elastic inversion of seismic reflection data. *Geophysics*, 51(10):1893–1903. 25, 31

[Tarantola, 2005] Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM. 33, 40, 47, 58

[Valenciano et al., 2003] Valenciano, A., Biondi, B., et al. (2003). 2-D deconvolution imaging condition for shot-profile migration. In *73rd Ann. Internat. Mtg., Soc. of Expl. Geophys., Expanded Abstracts*, pages 1059–1062. 29

[Valenciano, 2008] Valenciano, A. A. (2008). *Imaging by wave-equation inversion*. PhD thesis, Stanford University. 35

[van Leeuwen and Mulder, 2008] van Leeuwen, T. and Mulder, W. (2008). Velocity analysis based on data correlation. *Geophysical Journal International*, 56(6):791–803. 34, 152, 156

[van Leeuwen and Mulder, 2010] van Leeuwen, T. and Mulder, W. (2010). A comparison of seismic velocity inversion methods for layered acoustics. *Inverse Problems*, 26:1–21. 124

[Versteeg, 1993] Versteeg, R. J. (1993). Sensitivity of prestack depth migration to the velocity model. *Geophysics*, 58(6):873–882. 30

[Vigh et al., 2009] Vigh, D., Starr, E. W., and Kapoor, J. (2009). Developing earth models with full waveform inversion. *The Leading Edge*, 28(4):432–435. 58

[Virieux, 1986] Virieux, J. (1986). P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method. *Geophysics*, 51(4):889–901. 25, 32, 38, 42, 43

[Virieux and Operto, 2009] Virieux, J. and Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):wcc127–wcc152. 14, 24, 26, 30, 32, 33, 38, 40, 46, 48, 58, 61

[Virieux et al., 2009] Virieux, J., Operto, S., Ben-Hadj-Ali, H., Brossier, R., Etienne, V., Sourbier, F., Giraud, L., and Haidar, A. (2009). Seismic wave modeling for seismic imaging. *The Leading Edge*, 28(5):538–544. 42

[Wang et al., 2013] Wang, F., Chauris, H., Donno, D., and Calandra, H. (2013). Taking advantage of wavefield decomposition in full waveform inversion. In *EAGE*, page Tu0708. Extended Abstracts. 26, 34, 35, 36, 55, 81

[Wang et al., 2015a] Wang, F., Donno, D., Chauris, H., Calandra, H., and Audebert, F. (2015a). Full waveform inversion based on wavefield decomposition. In *SEG summer research workshop*. SEG. 81

[Wang et al., 2015b] Wang, H., Singh, S. C., Audebert, F., and Calandra, H. (2015b). Inversion of seismic refraction and reflection data for building long-wavelength velocity models. *Geophysics*, 80(2):R81–R93. 26, 34, 35, 80

[Whitmore et al., 1983] Whitmore, N. et al. (1983). Iterative depth migration by backward time propagation. In *1983 SEG Annual Meeting*. Society of Exploration Geophysicists. 29, 81

[Woodward et al., 2008] Woodward, M. J., Nichols, D., Zdraveva, O., Whitfield, P., and Johns, T. (2008). A decade of tomography. *Geophysics*, 73(5):VE5–VE11. 28, 59

[Wu et al., 2014] Wu, R.-S., Luo, J., and Wu, B. (2014). Seismic envelope inversion and modulation signal model. *Geophysics*, 79(3):WA13–WA24. 33

[Xu et al., 2012] Xu, S., Wang, D., Chen, F., Zhang, Y., and Lambaré, G. (2012). Full waveform inversion for reflected seismic data. In *EAGE*, page W024. Extended Abstracts. 26, 34, 35, 55, 66, 73, 80, 120, 150, 153

[Yoon and Marfurt, 2006] Yoon, K. and Marfurt, K. J. (2006). Reverse-time migration using the Poynting vector. *Geophysics*, 37:102–107. 36, 85

[Zhang et al., 2007] Zhang, W., Youn, S., Doan, Q. T., et al. (2007). Understanding reservoir architectures and steam-chamber growth at Christina Lake, Alberta, by using 4D seismic and crosswell seismic imaging. *SPE Reservoir Evaluation & Engineering*, 10(05):446–452. 28

[Zhang et al., 2006] Zhang, Y., Xu, S., Zhang, G., et al. (2006). Imaging complex salt bodies with turning-wave one-way wave equation. In *2006 SEG Annual Meeting*. Society of Exploration Geophysicists. 81

[Zhang et al., 2005] Zhang, Y., Zhang, G., and Bleistein, N. (2005). Theory of true-amplitude one-way wave equations and true-amplitude common-shot migration. *Geophysics*, 70(4):E1–E10. 35

[Zhou and Greenhalgh, 2003] Zhou, B. and Greenhalgh, S. A. (2003). Crosshole seismic inversion with normalized full-waveform amplitude data. *Geophysics*, 68(4):1320–1330. 45

[Zhou et al., 1995] Zhou, C., Cai, W., Luo, Y., Schuster, G. T., and Hassanzadeh, S. (1995). Acoustic wave-equation traveltime and waveform inversion of crosshole seismic data. *Geophysics*, 60(3):765–773. 35, 68

[Zhou et al., 2015] Zhou, W., Brossier, R., Operto, S., and Virieux, J. (2015). Combining diving and reflected waves for velocity model building by waveform inversion. *Geophysical Journal International*, 202:1535–1554. 26, 34, 80, 151, 154

L'inversion des formes d'ondes par décomposition des champs d'ondes

RESUME : L'inversion des formes d'ondes (FWI) est une procédure d'imagerie sismique pour imager le sous-sol de la Terre. FWI est résolue comme un problème d'optimisation. En fonction du contenu en fréquence des données, la fonction objective de FWI peut être fortement non linéaire. Pour des données associées des réflexions, ce problème empêche notamment les méthodes basées sur le gradient de retrouver les grandes longueurs d'onde du modèle de vitesse.

Dans cette thèse, nous proposons une variante de FWI basée sur la séparation des champs d'ondes, typiquement en champs montants et descendants, pour atténuer la non-linéarité du problème. Il consiste à décomposer le gradient de FWI en une partie de courte longueur d'onde et une partie de grande longueur d'onde après décomposition des champs d'ondes. L'inversion est effectuée d'une manière alternée entre ces deux parties. Nous appliquons cette méthode à plusieurs études de cas et montrons que la nouvelle approache est plus robuste en particulier pour la construction du modèle de grande longueur d'onde.

Mots clés : Inversion des formes d'ondes, décomposition des champs d'ondes, imagerie sismique.

Waveform inversion based on wavefield decomposition

ABSTRACT: Full Waveform Inversion (FWI) is a seismic imaging procedure to image the subsurface of the Earth. FWI is resolved as an optimization problem . Depending on the frequency content of the data, the objective function of FWI may be highly nonlinear. If a data set mainly contains reflections, this problem particularly prevents the gradient-based methods from recovering the long wavelengths of the velocity model.

In this thesis, I propose a variant of FWI based on the wavefield separation, typically between up- and down- going waves, to mitigate the nonlinearity of the problem. The new method consists of decomposing the gradient of FWI into a short-wavelength part and a long-wavelength part after wavefield decomposition. The inversion is performed in an alternating fashion between these two parts. We apply this method to several case studies and show that the new method is more robust especially for constructing the long-wavelength model.

Keywords : Waveform inversion, wavefield decomposition, seismic imaging.

MINES ParisTech

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY