Contributions to Seismic Full Waveform Inversion for Time Harmonic Wave Equations: Stability Estimates, Convergence Analysis, Numerical Experiments involving Large Scale Optimization Algorithms

Florian Faucher

Ph.D. Dissertation

Applied Mathematics

INRIA Bordeaux
Sud-Ouest, Équipe-Projet
Magique 3D

Université de Pau et des pays de l'Adour, Laboratoire de Mathématiques et de leurs Applications – UMR CNRS 5142

École Doctorale 211: Sciences Exactes et leurs applications

Contributions to Seismic Full Waveform Inversion for Time Harmonic Wave Equations: Stability Estimates, Convergence Analysis, Numerical Experiments involving Large Scale Optimization Algorithms.

Florian Faucher – November 29th 2017

Thesis advisors: Mme. Hélène Barucq & Mr. Henri Calandra

Reviewers

Mr. Samuli Siltanen	Professor, University of Helsinki
Mr. Jean Virieux	Emeritus Professor

Thesis committee

Mme. Hélène Barucq	Senior Research Scientist, Inria	Advisor
Mme. Elena Beretta	Professor, Politecnico di Milano	Examiner
Mr. Henri Calandra	Senior Research Engineer, TOTAL	Advisor
Mr. Guy Chavent	Emeritus Professor	Examiner
Mr. Maarten V. de Hoop	Simons Chair, Rice University	Examiner
Mr. Andreas Fichtner	Professor, ETH Zürich	Examiner
Mr. Jean Virieux	Emeritus Professor	Reviewer

Ph.D. Dissertation

Mathématiques Appliquées

INRIA Bordeaux
Sud-Ouest, Équipe-Projet
Magique 3D

Université de Pau et des pays de l'Adour, Laboratoire de Mathématiques et de leurs Applications – UMR CNRS 5142

École Doctorale 211: Sciences Exactes et leurs applications

Contribution à l'imagerie sismique par inversion des formes d'onde pour les équations d'ondes harmoniques: estimation de stabilité, analyse de convergence, expériences numériques utilisant des algorithmes d'optimisation à grande échelle.

Florian Faucher – 29 Novembre 2017

Directeurs de thèse: Mme. Hélène Barucq & Mr. Henri Calandra

Rapporteurs

Mr. Samuli Siltanen	Professor, University of Helsinki
Mr. Jean Virieux	Professeur Émérite

Jury de thèse

Mme. Hélène Barucq	Senior Research Scientist, Inria	Advisor
Mme. Elena Beretta	Professor, Politecnico di Milano	Examiner
Mr. Henri Calandra	Senior Research Engineer, TOTAL	Advisor
Mr. Guy Chavent	Emeritus Professor	Examiner
Mr. Maarten V. de Hoop	Simons Chair, Rice University	Examiner
Mr. Andreas Fichtner	Professor, ETH Zürich	Examiner
Mr. Jean Virieux	Emeritus Professor	Reviewer

"It is better to have your head in the clouds, and know where you are... than to breathe the clearer atmosphere below them, and think that you are in paradise."

Henry David Thoreau

Acknowledgement

My deepest gratitude goes to all the people who have helped and guided me throughout these three years. First of all, my advisor, Hélène Barucq, who has been kind enough to warmly welcome me among the team Magique 3D. Her great knowledge was a steady support, and she has been able to endure me during these three years, which is a proof of great composure. The confidence you place in our work is reassuring, and your availability for discussion has been widely advantageous. I also wish to deeply thank Henri Calandra and Maarten V. de Hoop, who trusted me from the very beginning. I am still puzzled by this somehow blind trust and the freedom you gave me. At the same time you all have always been encouraging, and available for discussions. I feel lucky to have been able to have the assistance of you three, and benefit from your eminent expertise.

Essential collaborations have been part of this work, and I wish to acknowledge my profound sympathy for Giovanni Alessandrini, Elena Beretta, Guy Chavent and Otmar Scherzer. You have been of tremendous help with your renowned scientific expertise. I am impressed by your immense patience and devotion. It has been my privilege to interact with you, and benefit from your knowledge. Your extraordinary kindness makes few words relevant. I have a dear respect for you, and I thank you. I also wish to acknowledge Andreas Fichtner, with whom I had the pleasure to interact and learn from each comments.

Research is sometimes ungrateful and always arduous. Yet the atmosphere in Magique 3D is wonderful and positively impact the everyday work. Once again, I acknowledge Hélène, for her sense of friendship and devotion for the team, and the meals she prepares for us with affection. The team is primarily of great scientific value, and I thank Julien Diaz for the fruitful discussions and help he provided me, and the patience he always has, despite my sometimes exaggerate insistence. I also include all members of the team, for the great discussions and time we had, and the constant support and friendship they offer, in addition to their smartness. I thank Aurélien, Ella, Juliette, Justine, Mamadou, Marc, Sébastien, Théo, Victor, Vincent and, Izar and Marie to have shared the office with constant malice. I also acknowledge the fruitful discussions with Jia, whose abilities and reactivity are valuable.

Consequently, this production is marked by the people I have interacted with. Besides, I wish to mention the authors from courses and books that I have used. There are certainly too many, but some are emphasized in the bibliography. The quality in diffusing mathematics, geophysics, and sciences in general is fundamental and sometimes as important as

the research itself.

I finally want to thank Ha for all that she taught me. Your mentality is certainly far from the fancy researcher, and it may not be the trend, but you bear the magnificent essence of research. It means laborious and respectful, full of modesty, integrity, and humility. You show great consideration for the problem and its foundations before everything. You validate a profound comprehension before assessing anything, and honestly carry your work without pretension. You are more than a model and can be assured of my sincere admiration.

Surely I will preciously retain the time you all have given me; it is to me the most valuable gift one can offer. Therefore, my final word goes to Jean Virieux and Samuli Siltanen, who have kindly accepted to review this manuscript; your expertise is of great value, and I am very grateful of the precious comments you gave me.

Abstract

In this project, we investigate the recovery of subsurface Earth parameters. We consider the seismic imaging as a large scale iterative minimization problem, and deploy the Full Waveform Inversion (FWI) method, for which several aspects must be treated. The reconstruction is based on the wave equations because the characteristics of the measurements indicate the nature of the medium in which the waves propagate. First, the natural heterogeneity and anisotropy of the Earth require numerical methods that are adapted and efficient to solve the wave propagation problem. In this study, we have decided to work with the harmonic formulation, i.e., in the frequency domain. Therefore, we detail the mathematical equations involved and the numerical discretization used to solve the wave equations in large scale situations.

The inverse problem is then established in order to frame the seismic imaging. It is a nonlinear and ill-posed inverse problem by nature, due to the limited available data, and the complexity of the subsurface characterization. However, we obtain a conditional Lipschitz-type stability in the case of piecewise constant model representation. We derive the lower and upper bound for the underlying stability constant, which allows us to quantify the stability with frequency and scale. It is of great use for the underlying optimization algorithm involved to solve the seismic problem. We review the foundations of iterative optimization techniques and provide the different methods that we have used in this project. The Newton method, due to the numerical cost of inverting the Hessian, may not always be accessible. We propose some comparisons to identify the benefits of using the Hessian, in order to study what would be an appropriate procedure regarding the accuracy and time. We study the convergence of the iterative minimization method, depending on different aspects such as the geometry of the subsurface, the frequency, and the parametrization. In particular, we quantify the frequency progression, from the point of view of optimization, by showing how the size of the basin of attraction evolves with frequency.

Following the convergence and stability analysis of the problem, the iterative minimization algorithm is conducted via a multi-level scheme where frequency and scale progress simultaneously. We perform a collection of experiments, including acoustic and elastic media, in two and three dimensions. The perspectives of attenuation and anisotropic reconstructions are also introduced. Finally, we study the case of Cauchy data, motivated by the dual sensors devices that are developed in the geophysical industry. We derive a novel cost function, which arises from the stability analysis of the problem. It allows elegant perspectives where no prior information on the acquisition set is required.

Abstract

Dans ce projet, nous étudions la reconstruction de milieux terrestres souterrains. L'imagerie sismique est traitée avec un problème de minimisation itérative à grande échelle, et nous utilisons la méthode de l'inversion des formes d'ondes (Full Waveform Inversion, FWI method). La reconstruction est basée sur des mesures d'ondes sismiques, car ces ondes sont caractérisées par le milieu dans lequel elles se propagent. Tout d'abord, nous présentons les méthodes numériques qui sont nécessaires pour prendre en compte l'hétérogénéité et l'anisotropie de la Terre. Ici, nous travaillons avec les solutions harmoniques des équations des ondes, donc dans le domaine fréquentiel. Nous détaillons les équations et l'approche numérique mises en place pour résoudre le problème d'onde.

Le problème inverse est établi afin de reconstruire les propriétés du milieu. Il s'agit d'un problème non-linéaire et mal posé, pour lequel nous disposons de peu de données. Cependant, nous pouvons montrer une stabilité de type Lipschitz pour le problème inverse associé avec l'équation de Helmholtz, en considérant des modèles représentés par des constantes par morceaux. Nous explicitons la borne inférieure et supérieure pour la constante de stabilité, qui nous permet d'obtenir une caractérisation de la stabilité en fonction de la fréquence et de l'échelle. Nous revoyons ensuite le problème de minimisation associé à la reconstruction en sismique. La méthode de Newton apparaît comme naturelle, mais peut être difficilement accessible, dû au coup de calcul de la Hessienne. Nous présentons une comparaison des méthodes pour proposer un compromis entre temps de calcul et précision. Nous étudions la convergence de l'algorithme, en fonction de la géométrie du sous-sol, la fréquence et la paramétrisation. Cela nous permet en particulier de quantifier la progression en fréquence, en estimant la taille du rayon de convergence de l'espace des solutions admissibles.

A partir de l'étude de la stabilité et de la convergence, l'algorithme de minimisation itérative est conduit en faisant progresser la fréquence et l'échelle simultanément. Nous présentons des exemples en deux et trois dimensions, et illustrons l'incorporation d'atténuation et la considération de milieux anisotropes. Finalement, nous étudions le cas de reconstruction avec accès aux données de Cauchy, motivé par les *dual sensors* développés en sismique. Cela nous permet de définir une nouvelle fonction coût, qui permet de prometteuses perspectives avec un besoin minimal quant aux informations sur l'acquisition.

Contents

General introduction

The study of the Earth is a challenging research, where many fields interact. It has been considered for centuries, and it remains a very active subject of interest. The nature of the Earth subsurface is yet broadly well established. The different layers are defined as follow: the continental crust, the upper mantle, the lower mantle, the outer core, and the inner core. However, depending on the resolution (scale) of accuracy, the information is relatively partial. As an illustration, we have very few knowledge on the ocean floor, which represents more than 70% of the Earth surface. The seismology, which aims the study of earthquakes and the propagation of waves on Earth is also arduous and suffers from the minimal information and the scale of the problem; for instance, the prediction of earthquakes remains out of reach. On the other hand, the understanding of Earth is valuable for different reasons. On the planetary scale, it can allow the prediction of potentially dangerous events such as earthquakes and tsunamis. On a more local scale, it can help to identify structures or materials of interest in the subsurface (seismic prospecting). Overall, it is scientifically crucial to characterize precisely the very place where humankind lives. The methods employed to study the Earth depends on the scale investigated. In this project, we focus on the recovery subsurface region of several ten kilometers, by means of quantitative reconstructions of the subsurface properties. It is a *large scale inverse problem*.

The understanding of Earth usually relies on the propagation of waves. The waves are influenced by the medium in which they propagate so that the analysis of this propagation reveals information on the zone. In particular, we refer to *seismic waves* for the mechanical waves that propagate in the Earth. Therefore, the first step is to define the behavior of waves in terms of mathematical equations: these are the *wave equations*. They are extracted from the pioneering work of Newton, and allow the different situations to be accounted for (fluid, solid, anisotropy, etc). It defines, in seismic, the *direct* (or *forward*) problem, which refers to the resolution of the waves propagation problem, from the knowledge of the medium properties (wave speed, density, etc).

The forward problem invokes the wave equations, and necessitates the knowledge of

the material properties, as well as the initial source perturbation, to accurately predict the waves path. On the contrary, the inverse problem aims at recovering the subsurface properties, using the observation of waves at the surface. Those measurements can result from physical phenomena (e.g., earthquakes) or can result from a human made source. It is the latest in our context of seismic prospecting. More generally, the essence of inverse problems is the recovery of information which we cannot access. The medical imaging is probably a well-known example, where the interior of the human body (bone, organ, tumor) is explored from measurements at the skin level. Many applications involve inverse problems; we can mention, for instance, the study of celestial objects, archeology, helioseismology, sonar, etc. The methods to recover the investigated parameters may differ depending on the situations, but the concept is the same: the discovery of the invisible, the unknown. It usually consists in a nonlinear ill-posed problem, which requires precautious treatments at both mathematical and numerical levels.

In our project, the inverse wave problem is processed by casting it into an iterative optimization problem. This vision was developed in the 1980's, following the pioneering work of Tarantola. The method, relatively restricted at its early ages due to the computational limitation, has been widely popularized since then. Indeed, because of the intrinsic heterogeneity of the Earth, and the scale of the area of interest, the computational burden grows rapidly. That is why the improvement of the computational power, and the High Performance Computing (HPC) framework, have allowed significant advances regarding large scale problems in general, and the seismic reconstruction in particular.

We follow the time-harmonic formulation of the waves and consequently work in the frequency domain. The wave problem naturally occurs in the time domain, but the frequency approach has some interesting impacts: the numerical boundary conditions are easier to handle, the incorporation of attenuation is natural, the frequency decomposition is required for the stability of the inverse problem, etc. The drawback of the method is certainly the computational time and memory required to solve large linear systems (by means of matrix factorization); yet, the research on novel solvers is an active field, and also benefits from the constant increase of computational power. At the end of the day, the two formulations have their pros and cons, and both can be justified.

We have given a brief overview of the concepts involved in this project. We do not elaborate because every chapter is composed of its own introduction, where more details and the appropriate references are given on the specific aspect investigated therein.

Plan of the manuscript

The first two chapters are dedicated to the forward problem, which is not really the heart of the project, but a fundamental key to apprehending it. We review the propagation of waves in terrestrial media in Chapter 1, and introduce the underlying *wave equations*. We define their time-harmonic formulation, which is the one employed throughout the manuscript, and depict the different possible situations: acoustic, elastic, anisotropic. Then, we provide the numerical methods to solve the wave propagation problems in Chapter 2. We review the discretization schemes that have been used in our work, with the finite differences, and the continuous and discontinuous Galerkin approaches.

We introduce the inverse problem in Chapter 3, where we examine the stability of the inverse problem associated with the Helmholtz equation. Providing a piecewise constant representation of the model, a Lipschitz-type stability is obtained, and we give the characterization of the stability constant, with its lower and upper bounds. The analytical proof is confronted with numerical situations to demonstrate the sharpness of the bounds. The stability depending on the frequency and scale gives the first brick in our reconstruction algorithm.

The inverse problem is cast as an iterative minimization problem in Chapter 4, following the Full Waveform Inversion method, and the pioneering work of Tarantola. Newton types methods appear naturally as the best candidates to conduct the minimization, but they can suffer from numerical restriction (time and memory). We detail the methods that have been used in our software, and the computation of the gradient and Hessian via adjoint state methods. The Hessian is approximated using a Conjugate Gradient algorithm to limit the computational requirement. We further compare the efficiency of using gradient based methods or incorporating Hessian information. Depending on the situation, the computational burden involved in the large scale optimization problems we face is of importance and restrict the possibilities.

The convergence of the iterative scheme is probed in Chapter 5. We designed estimations of the size of the basin of attraction depending on the frequency and the geometry, to quantify the nonlinearity of the problem with respect to these aspects. Using the radius of curvature, we also identify the maximum distance between the data and the attainable set. We finally inquire the effect of parametrizations during the reconstruction.

Following the stability and convergence analysis of the seismic inverse problem, we deploy a multi-level iterative algorithm for the minimization, where frequency and scale evolve simultaneously. In Chapter 6, we proceed to experimental tests, with two and three-

dimensional domains, using acoustic and elastic media. We introduce the possibility for the consideration of attenuation and anisotropy as well.

Finally, in Chapter 7, we investigate the seismic reconstruction from Cauchy data, motivated by the measurements obtained from dual sensors, in marine seismic campaigns. We illustrate how it is revealed to be promising, by the definition of a novel cost function, adequate for this type of measurements. It allows minimal information on the acquisition set, and new perspectives to reduce the numerical cost.

Related productions

During the three years of this thesis, one article has been published and two are currently in preparation,

- **In preparation,** "Convergence Properties for Numerical Acoustic and Elastic Full Waveform Inversion in the Frequency Domain" with Hélène Barucq, Guy Chavent and Henri Calandra.
  The essence of this paper is in Chapter 5.
- **In preparation,** "Inverse problem for the Helmholtz equation with Cauchy data: reconstruction with conditional well-posedness driven iterative regularization" with Giovanni Alessandrini, Maarten V. de Hoop, Romina Gaburro, Eva Sincich and Hélène Barucq.
  The essence of this paper is in Chapter 7.
- "Inverse boundary value problem for the Helmholtz equation: quantitative conditional Lipschitz stability estimates" with Elena Beretta, Maarten V. de Hoop and Otmar Scherzer. *SIAM Journal on Mathematical Analysis* 48, no. 6 (2016): 3962-3983.
  This paper is reproduced in the Appendix A.

We also mention the proceedings, expanded abstracts and conferences attended in relation to this work,

- "Stability and convergence for seismic reconstruction" with Hélène Barucq, Henri Calandra, Guy Chavent and Maarten V. de Hoop. *Johann Radon Institut for Computational and Applied Mathematics (RICAM), Linz, Austria, July 2017.*
- "Stability and convergence for seismic reconstruction using full waveform inversion" with Hélène Barucq, Henri Calandra, Guy Chavent and Maarten V. de Hoop. Workshop on Computational Inverse Problems for Partial Differential Equations. *Oberwolfach Research Institute for Mathematics, Germany, May 2017.*

– “Acoustic Full Waveform Inversion from Cauchy data via conditional well-posedness driven iterative regularization" with Giovanni Alessandrini, Maarten V. de Hoop, Romina Gaburro and Eva Sincich. *Proceedings of the Project Review, Geo Mathematical Imaging Group, Houston, USA, 2017.*

– “Elastic full-waveform inversion with surface and body waves" with Jia Shi, Maarten V. De Hoop and Henri Calandra. *SEG Technical Program Expanded Abstracts 2016, pp. 1120-1124. Society of Exploration Geophysicists, 2016.*

– “Full waveform inversion for elastic medium using quantitative Lipschitz stability estimates" with Hélène Barucq, Henri Calandra, Maarten V. de Hoop and Jia Shi. *7th EAGE Saint Petersburg International Conference and Exhibition, Saint-Petersburg, Russia, 2016.*

– “Inverse Problem in the Frequency Domain for Subsurface Reconstruction". *Workshop DIP – Depth Imaging Partnership (INRIA-TOTAL), Houston, USA, 2016*

– “Elastic isotropic full waveform inversion via quantitative stability estimates" with Hélène Barucq, Henri Calandra, Maarten V. de Hoop and Jia Shi. *PANACM (Pan-American Congress On Computational Mechanics), Buenos Aires, Argentina, 2015.*

– “Stability estimates for full waveform inversion" *Workshop DIP – Depth Imaging Partnership (INRIA-TOTAL), Pau, France, 2015.*

– “Multi-level elastic full waveform inversion in isotropic media via quantitative Lipschitz stability estimates" with Jia Shi, Henri Calandra and Maarten V. de Hoop. *Proceedings of the Project Review, Geo-Mathematical Imaging Group, Chicago, USA, 2014.*

– “Multi-level, multi-frequency, 3D Full Waveform Inversion", *MATHIAS – TOTAL Symposium on Mathematics, Paris, France, 2014.*

Introduction Générale

L'étude de la Terre demeure un secteur de recherche actif, où beaucoup de domaines interragissent. Cela fait par ailleurs des siècles que la Terre est un sujet d'étude. La nature de la structure souterraine est assez bien définie, avec la croûte, le manteau supérieur, le manteau inférieur, le noyau externe et le noyau interne. Cependant, selon la précision souhaitée, l'information se révèle partielle. Par exemple, nous avons très peu de connaissances sur le fond des océans, qui représente pourtant 70% de la surface terrestre. La sismologie, qui étudie les tremblements de terre ou la propagation d'ondes dans la Terre, est également complexe à cause de la faible quantité d'information et la taille du système à résoudre; ainsi, la prédiction des tremblements de terre reste inaccessible. L'étude de la Terre est cependant très importante pour différentes raisons. À l'échelle planétaire, cela pourrait permettre la prédiction de phénomènes dangereux (comme les tramblements de terre et les tsunamis). D'un point de vue plus local, cela peut permettre d'identifier des ressources d'intérêt dans le sous-sol. Il est par ailleurs scientifiquement crucial de comprendre précisément l'endroit où nous vivons. Les méthodes à employer pour étudier la Terre dépendent de la résoluton souhaitée. Dans notre projet, nous nous intéressons la reconstruction de mileux souterrains dont la taille est de quelques dizaines de kilomètres, avec une reconstruction quantitative des propriétés du milieu. Il s'agit d'un problème inverse à grande échelle.

La compréhension de la Terre se repose généralement sur la propagation d'ondes. Les ondes sont influencées par le milieu dans lequel elles se propagent et l'analyse de leur propagation donne ainsi des informations sur le domaine. En particulier, nous parlons d'*ondes sismiques* pour les ondes mécaniques qui se propagent dans la Terre. De fait, la première étape est de définir le comportement des ondes en termes d'équations mathématiques: ce sont les *équations des ondes*. Elles dérivent des lois de Newton, et permettent de prendre en compte différents cas (acoustique, élastique, etc). Nous parlons alors de problème *direct* pour la résolution de ces équations, qui nécessitent la connaissance des propriétés du milieu (vitesse, densité, etc).

A l'opposé, le problème inverse vise la reconstruction de milieux souterrains, à par-

tir d'observations d'ondes à la surface. Ces mesures peuvent résulter de phénomènes naturels (tremblement de terre) ou provoqués par l'homme. C'est cette dernière situation qui prévaut pour la prospection sismique. De manière générale, l'essence du problème inverse est la reconstruction d'information pour lesquelles nous n'avons pas d'accès direct. L'imagerie médicale est certainement l'exemple le plus évident, où l'intérieur du corps humain (os, organe, tumeur) est exploré à partir de mesures d'ondes au niveau de la peau. De nombreuses applications sont basées sur des problèmes inverses: l'étude d'objets célestes, l'archéologie, l'héliosismologie, le sonar, etc. Les méthodes permettant d'identifier les paramètres d'intérêt diffèrent selon la situation, mais le concept reste le même: la découverte de l'inconnu, de l'invisible. De plus, il s'agit généralement d'un problème non-linéaire et mal posé, qui demande des précautions de traitements, tant aux niveaux mathématiques, que numériques.

Dans notre projet, le problème inverse est représenté par un problème d'optimisation. Cette vision, développée dans les années 1980, découle du travail de Tarantola. La méthode, relativement restreinte d'un point de vue numérique à son établissement, s'est depuis popularisée avec l'avènement du calcul haute performance, et l'accroissement des puissances de calcul à disposition.

Nous nous intéressons à la propagation d'ondes harmoniques, et travaillons de fait dans le domaine fréquentiel. Alors que le phénomène se relie naturellement au temps, l'approche fréquentielle présente certains avantages tels que la facilité de traitement des conditions aux limites ou la prise en compte de l'atténuation. La décomposition en fréquence est également fondamentale pour la stabilité du problème. L'inconvénient majeur est le coût de calcul, en temps et mémoire, pour la création et la factorisation de la matrice dans le cas de grands problèmes 3D. De manière générale, les deux approches, temporelle et fréquentielle, ont chacune leurs avantages et inconvénients, et peuvent toutes deux se justifier.

Plan du manuscrit

Les deux premiers chapitres se concentrent sur le problème direct, qui, s'il n'est pas réellement le coeur du projet, est une brique fondamentale pour le traiter. Nous présentons la propagation des ondes en milieux terrestres au Chapitre 1 et introduisons les équations des ondes. Nous définissons leur formulation harmonique, qui est employée dans notre étude. Ensuite, nous déployons les méthodes numériques pour résoudre la propagation d'ondes au Chapitre 2. Les techniques de discrétisation qui sont utilisées dans notre travail sont introduites, il s'agit des différences finies, et des méthodes de Galerkin continues et discontinues.

Le problème inverse est présenté au Chapitre 3, où nous étudions en particulier la stabilité du problème inverse associé à l'équation de Helmholtz. Avec une représentation du modèle par des constantes par morceaux, la stabilité de type Lipschitz est obtenue, et nous caractérisons la borne inférieure et supérieure de la constante de stabilité. La théorie est comparée à des expérimentations numériques et nous illustrons notamment la précision des bornes. La stabilité dépend de la fréquence et de l'échelle, et constitue la première brique de notre algorithme de reconstruction.

Nous présentons le problème de minimisation itérative associé à l'imagerie sismique au Chapitre 4, avec la méthode d'inversion des formes d'ondes (FWI, Full Waveform Inversion). La méthode de Newton apparaît naturellement comme la meilleure option, mais peut générer des coûts numériques prohibitifs. Nous présentons les différentes techniques qui ont été introduites dans notre software, avec le calcul du gradient et de la Hessienne basé sur l'état adjoint. L'inversion de la Hessienne est en fait approchée en utilisant un algorithme de gradient conjugué afin de réduire le coût numérique. Nous présentons également des expériences numériques afin d'illustrer le problème d'optimisation à grande échelle, et les possibles bénéfices d'utiliser la Hessienne.

La convergence de la méthode itérative est détaillée au Chapitre 5. Nous estimons la taille du bassin d'attraction en fonction de la fréquence et de la géométrie du milieu souterrain. Nous proposons également l'estimation du rayon de courbure, qui identifie la distance maximale autorisée avec l'ensemble des possibles. Finalement, nous détaillons l'effet de la paramétrisation durant la reconstruction.

À partir de l'analyse de stabilité et de convergence, nous deployons un algorithme de minimisation multi-niveaux, où la fréquence et l'échelle évoluent simultanément. Au Chapitre 6, nous présentons des tests numériques, en deux et trois dimensions, avec des milieux acoustiques et élastiques. Nous présentons également les possibilités dans le cas de l'atténuation et l'anisotropie.

Enfin, au Chapitre 7, nous présentons les perspectives de reconstruction avec des données de Cauchy, motivés par le développement du 'dual sensors' en sismique marine. Cela nous permet de promouvoir une nouvelle fonction coût, adéquate à ce type de mesures. Ainsi, une information minimale est requise quant aux données d'acquisition, et le coût numérique s'en trouve réduit.

Production bibliographique

Durant les trois ans de cette thèse, un article a été publié, et deux sont actuellement en préparation :

- **In preparation,** "Convergence Properties for Numerical Acoustic and Elastic Full Waveform Inversion in the Frequency Domain" with Hélène Barucq, Guy Chavent and Henri Calandra.
  Ce papier est détaillé au Chaptire 5.
- **In preparation,** "Inverse problem for the Helmholtz equation with Cauchy data: reconstruction with conditional well-posedness driven iterative regularization" with Giovanni Alessandrini, Maarten V. de Hoop, Romina Gaburro, Eva Sincich and Hélène Barucq.
  Ce papier est détaillé au Chaptire 7.
- "Inverse boundary value problem for the Helmholtz equation: quantitative conditional Lipschitz stability estimates" with Elena Beretta, Maarten V. de Hoop and Otmar Scherzer. *SIAM Journal on Mathematical Analysis* 48, no. 6 (2016): 3962-3983.
  Ce papier est reproduit en Appendice A.

Nous présentons aussi les conférences, et les (expanded) abstracts en relation,

- "Stability and convergence for seismic reconstruction" with Hélène Barucq, Henri Calandra, Guy Chavent and Maarten V. de Hoop. *Johann Radon Institut for Computational and Applied Mathematics (RICAM), Linz, Austria, July 2017.*
- "Stability and convergence for seismic reconstruction using full waveform inversion" with Hélène Barucq, Henri Calandra, Guy Chavent and Maarten V. de Hoop. Workshop on Computational Inverse Problems for Partial Differential Equations. *Oberwolfach Research Institute for Mathematics, Germany, May 2017.*
- "Acoustic Full Waveform Inversion from Cauchy data via conditional well-posedness driven iterative regularization" with Giovanni Alessandrini, Maarten V. de Hoop, Romina Gaburro and Eva Sincich. *Proceedings of the Project Review, Geo Mathematical Imaging Group, Houston, USA, 2017.*
- "Elastic full-waveform inversion with surface and body waves" with Jia Shi, Maarten V. De Hoop and Henri Calandra. *SEG Technical Program Expanded Abstracts 2016, pp. 1120-1124. Society of Exploration Geophysicists, 2016.*
- "Full waveform inversion for elastic medium using quantitative Lipschitz stability estimates" with Hélène Barucq, Henri Calandra, Maarten V. de Hoop and Jia Shi. *7th*

EAGE Saint Petersburg International Conference and Exhibition, Saint-Petersburg, Russia, 2016.

– "Inverse Problem in the Frequency Domain for Subsurface Reconstruction". *Workshop DIP – Depth Imaging Partnership (INRIA-TOTAL), Houston, USA, 2016*

– "Elastic isotropic full waveform inversion via quantitative stability estimates" with Hélène Barucq, Henri Calandra, Maarten V. de Hoop and Jia Shi. *PANACM (Pan-American Congress On Computational Mechanics), Buenos Aires, Argentina, 2015.*

– "Stability estimates for full waveform inversion" *Workshop DIP – Depth Imaging Partnership (INRIA-TOTAL), Pau, France, 2015.*

– "Multi-level elastic full waveform inversion in isotropic media via quantitative Lipschitz stability estimates" with Jia Shi, Henri Calandra and Maarten V. de Hoop. *Proceedings of the Project Review, Geo-Mathematical Imaging Group, Chicago, USA, 2014.*

– "Multi-level, multi-frequency, 3D Full Waveform Inversion", *MATHIAS – TOTAL Symposium on Mathematics, Paris, France, 2014.*

Chapter 1

The wave equations

Abstract

The mathematical framework to study the propagation of seismic waves is discussed in this chapter. The different equations are introduced following the laws of classical mechanics. Seismic waves require a physical material for their propagation and the nature of this material affects the formulation of the equations. For instance, the state of the matter, fluid or solid, gives insight into the form of propagation, as does the condition of isotropy for example. The simplifications of the wave equations arise from those properties and are required for a proper understanding of the problem. The propagation of wave naturally occurs in the time domain, and the harmonic formulation follows from solutions having special forms. The wave equation is transformed to expose the frequency variable, which is possibly complex. It is the formulation of choice for the rest of this manuscript. The notion of attenuation in the medium is easily incorporated in this situation, and introduces a complex-valued velocity.

Contents

Seismic exploration, and more generally the study of Earth, relies mainly on waves. Waves, as traveling energy, contain fundamental information of the medium in which they propagate. Similarly to any traveler, the wave is affected by the path it takes. However, opposite to the traveler's free will, the wave movement is deterministic and decided by the medium; for example, the speed of the wave is fixed by the medium in which it propagates. That is why the understanding of Earth first necessitates the understanding of wave propagation. It is further required to predict and anticipate some potentially dangerous effects carried by the waves propagation such as tsunami or earthquake.

There exist different types of propagating waves, and their nature has to be specified. The most natural one to think of is the mechanical wave, which we will consider throughout this manuscript. The mechanical waves need matter to propagate, more precisely the oscillation of this matter defines the propagation as the resulting transport of energy. It is not the only possibility when referring to wave, as illustrated with the case of electromagnetic waves. Mechanical and electromagnetic waves differ in the type of energy they create: electromagnetic waves induce a coupling of an electrical and a magnetic stimulations while it is a physical displacement (and pressure) in the case of mechanical waves. Another important difference lies in that electromagnetic waves do not specially need a medium to propagate; instead they can travel in the vacuum and through entire galaxies while mechanical waves fundamentally need a material for the support of information. When a material is used as the support of information, the wave energy suffers from attenuation at increasing distance from the source (which is not the case for electromagnetic waves in vacuum).

All types of wave have in common that they come into existence via an initial perturbation: this is what we will refer to as the *source*. For instance, when a hammer hits a wood panel, it generates a mechanical wave starting from the collision point. Alternatively, the nuclear reactions continuously happening in the solar core produce electromagnetic waves (radiation) that travel through space. As we mentioned, the speed of the waves is decided by the material in which they propagate. In the case of vacuum, the electromagnetic waves travel at the speed of light (approximately 3×10^8 $\mathrm{m\,s^{-1}}$). Inside materials, the motion of waves is slower, for example, mechanical waves in water usually reach the speed of about 1500 $\mathrm{m\,s^{-1}}$. They travel slightly faster in solid, about 5100 $\mathrm{m\,s^{-1}}$ in iron and about 3500 $\mathrm{m\,s^{-1}}$ in wood. Yet it is nothing compared to the speed in the vacuum of electromagnetic waves.

In our project we focus on the waves propagating through the Earth: seismic waves. In Section 1.1 we review the classification of the seismic waves and illustrate their influence onto the medium of propagation. In Section 1.2 the successive historical developments concerning the study of waves are reviewed. It is interesting to notice that waves are originally strongly connoted to the notion of the visible (etymologically speaking as well),

yet it is now widely used to identify the unseen, the unknown.

From the mathematical point of view, waves propagation is governed by the so-called *wave equations*, which give the framework from the knowledge of the geophysical properties (density, velocity, etc). The plural form reveals the different possibilities in the formulation, possibilities arising from the nature of the medium of interest. In particular, the propagation of waves in fluid and solid does not focus on the same quantity, pressure for the fluid and displacement (or velocity) for the solid. It further makes the distinction between acoustic (fluid) and elastic (solid) propagations. Behind this distinction is the underlying notion of Lagrangian or Eulerian approach, see Aki and Richards (2002). The general wave equations are derived from the law of classical mechanics and successive simplifications can be obtained by including assumptions on the medium of interest. Section 1.3 introduces the different steps to formulate the mathematical equations that coincide with wave propagation. We introduce the word *wavefield* to identify the general solution of the wave equation, independently of the situation. The mathematical framework is extended to identify simplified equations depending on the medium characteristics (isotropy, acoustic, etc) in Section 1.4. The isotropic characterization of the medium is established when the waves propagate similarly in all directions; this can be pictured by some consistency in the material. It is unfortunately rarely the case in the nature but it provides important modification to simplify the mathematical equations and consequently the numerical problem.

The time dependent nature of the propagation is a instinctive aspect in our mind. Indeed the phenomenon has a point of departure (the excitation generated by the source), and the waves propagate for 'some time'. This thinking is altered by the harmonic (frequency domain) formulation of the wave equations, which utilizes a special solution of the equation to replace the time variable by the frequency. More details and the reformulated equations will be explained in Section 1.6. This is the formulation selected for the rest of the manuscript. We also review some additional components in Sections 1.7 and 1.8 such as the incorporation of attenuation, the boundary conditions, and we give some initial material regarding the analytical study of the mathematical problem and the consequent Partial Differential Equation (PDE).

1.1 Seismic waves

The seismic waves propagating through a medium are divided upon two classes: the *surface waves*, which, as the name indicates, refer to waves that propagate at the surface of the material; and the *body waves*, for the component that propagates inside, through the medium. Let us picture the common image of a rock thrown to the water, the surface waves are the

visible circles appearing at the water surface while the body waves are the underwater ones, likely to be unseen by the observer. These two main families can be further divided into subclasses, following the general structure presented in Figure 1.1.

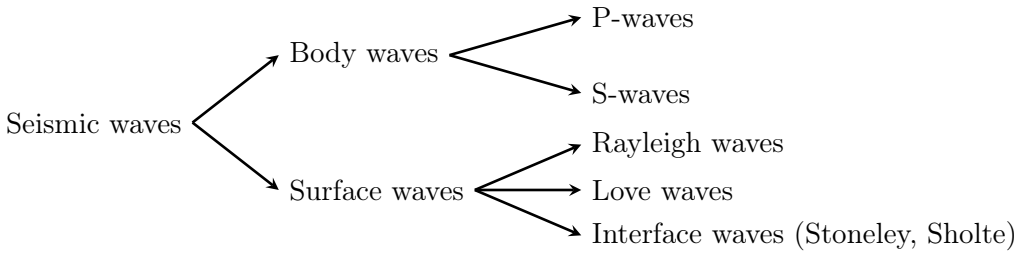
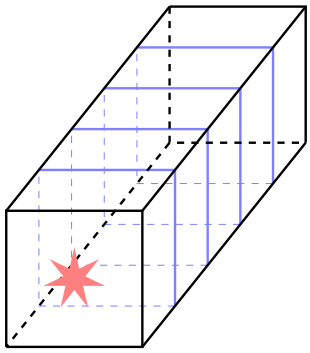


Figure 1.1: The different classes encompassed by the seismic waves.

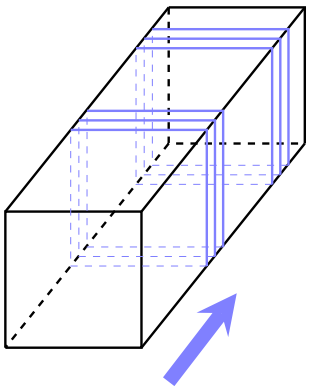
The body waves are separated into two subclasses: the P-waves (also referred to as *primary* or *compressional* waves) and the S-waves (also referred to as *secondary* or *shear* waves). The P-waves appear via the compression and dilatation of the medium along the longitudinal axis, perpendicular to the surface where the initial perturbation occurs. The S-waves correspond to the transverse perturbations of the medium: parallel to the surface. The latest can further be separated into horizontal and vertical displacement, the SH-and SV-waves respectively. Figure 1.2 illustrates the decomposition of those two waves: the starting medium at rest is pictured on the top left, the P-wave is illustrated in the central top and the S-wave on the top right and bottom left figures (SH and SV-waves respectively). The speed of propagation, the *wave speed*, also differs for both types. P-waves are faster than S-waves, with a factor of about 8/5. S-waves are generally more difficult to detect and only appear in a solid medium, i.e. not in fluids. This is because the shear force does not exist in fluid (namely fluids cannot be seen as rigid).

The surface wave travels slower than the body waves, they can also be separated in different families. The Love waves (named after Augustus Edward Hough Love, 1863–1940) result from the shear (S-) waves and correspond to the lateral (perpendicular to the direction of propagation) surface motion. The Rayleigh waves (named after John William Strutt Rayleigh, 1842–1919) propagate in the longitudinal direction (parallel to the propagation) and refers to the rolling displacement along the surface, see Figure 1.2 (at the bottom). There are finally the waves appearing at the interface between two domains, for instance between solid and fluid (the Scholte waves) or between two solid sections (the Stoneley waves). We refer to Aki and Richards (2002) for complete descriptions.

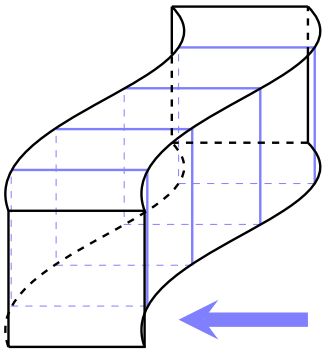
The source of seismic waves, that initiates the oscillations of matter, can be natural or human made. For example, the source is natural for an earthquake, where the waves travel from the origin, the hypocenter, through the subsurface medium and possibly reach the surface. Similarly, a volcano eruption generates powerful waves. We can also think of



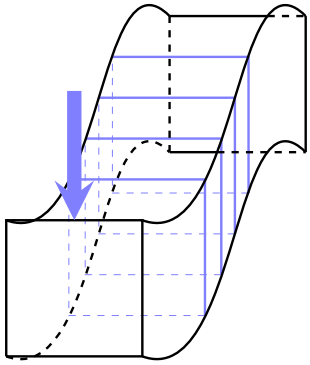
(a) Initial medium



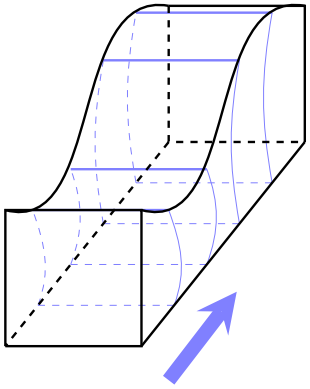
(b) Illustration of the (body) P-wave



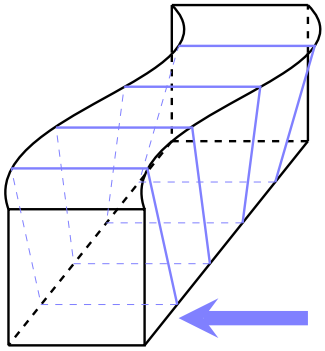
(c) Illustration of the (body) SH-wave



(d) Illustration of the (body) SV-wave



(e) Illustration of the (surface) Rayleigh wave



(f) Illustration of the (surface) Love wave

Figure 1.2: Illustration of the decomposition of a seismic wave into body (1.2(b), 1.2(c) and 1.2(d)) and surface (1.2(e) and 1.2(f)) waves. Figure 1.2(a) corresponds to the medium without perturbation, where the source is located at the center of the front face, marked with the red star. It is where the wave is generated from. The body wave is decomposed into the P- (1.2(b)), SH- (1.2(c)) and SV- (1.2(d)) waves. The P-wave component propagates along the longitudinal axis of the medium, normal to the front surface and introduces inner compression and dilatation of the medium as shown by the displacement of the blue sections. The S-wave corresponds to the part of the wave which is parallel to the surface and introduces the oscillations marked by the blue sections. Surface wave can further be decomposed into the Rayleigh (1.2(e)) and Love (1.2(f)) waves. The Rayleigh wave is along the longitudinal axis and the Love wave perpendicular. The blue arrows represents the direction of propagation of the wave components.

tides, resulting from the gravitational force. On the other hand, for the scientific study of subsurface material properties of Earth, the source is often artificially created by humans. In this case it can be an explosion (literally dynamite placed in the area) that initiates the process. Another possibility on land is the use of vibroseis: large truck that applies heavy constraint on the Earth surface to create the initial vibration. For marine source, ultrasound can also be used via air gun, in this case the wavefield propagates through water

and eventually reaches the bottom rocks, where the propagation continues. The cost of a seismic exploration campaign is high, especially considering the size of the area of interest, commonly several tens of kilometers, and the cost of the machines used (vibroseis, boat, receivers, sources).

Contrary to natural phenomenon like earthquakes where the source is relatively deep in the subsurface, artificial sources position is often limited to surface or near surface locations, due to understandable access restriction. The will of not altering the media (i.e. we disregard borehole) is also important and such processes are called *non-intrusive*. Limiting the source to the surface area obviously restricts the study of the deepest subsurface structures because the waves may not have enough energy to reach them. This is of crucial importance for the inverse problem that is addressed in the sequel.

1.2 Historical background

The study of waves comes from a very ancient interest. It has originally been strongly related to the phenomena appearing in water and the oscillating movement than can be easily seen at the surface by any observer. The etymology of the word reflects this original meaning, that has later been extended. The Latin root, '*unda*', which has given the origin of many designations in current languages ('*onde*' in French, '*onda*' in Spanish and Italian), is also related to the general notion of movement in liquids. Alternatively, the English root is directly related to the notion of oscillation (as well as the German word '*welle*'). We can find similar significance for the Greek word '$\kappa\acute{\upsilon}\mu a$' and its ancient root '$\kappa\acute{\upsilon}\omega$', which literally means *swell*. Here we see that the concept of wave is directly related to the phenomenon of oscillation, especially in liquids and as what can be seen. Later this relation to water and liquid has been widely extended to other kind of traveling information, such as the previously mentioned electromagnetic waves.

The understanding of waves has been initially thought of in the ancient Greece, especially because of the properties of music. Pythagoras (circa 570–495 BC) related the notion of harmony with the sounds produced by different strings of proportional length (as well as sounds depending on the weight of a blacksmith hammer). Such experiments clearly make appear the concept of frequency and period. We can remark that this concept is still applied in the music where an *octave* can be seen as doubling (or halving) the size of the string (hence introducing a factor two in the frequency), and more generally a rational ratio is applied to link the musical notes. In ancient Greece, Aristotle (384–322 BC) defined the notion of vibration and period for traveling waves. Later the Italian scientist Galileo Galilei (1564–1642) discovered the law of the vibrating string, as well as the French Marin

Mersenne (1588–1648) independently. Eventually, we will see that the work of Isaac Newton (1642–1727) in the seventeenth century allowed the development of the mathematical equations to describe the propagation of waves from physical acoustics (the term *acoustic* was initially devoted to the music and sounds, and was introduced by Joseph Sauveur, 1653–1716). Later, Jean le Rond D'Alembert (1717–1783), Daniel Bernoulli (1700–1782) and Leonhard Euler (1707–1783) developed the equation corresponding with a vibrating string. This work was then extended to find analytical solutions as well as approximate solutions, and to consider other types of geometries (membrane, plate) with the contribution of many scientists in the eighteenth and nineteenth centuries such as Joseph-Louis Lagrange (1736–1813), Sophie Germain (1776–1831, plate vibration), Claude-Louis Navier (1785–1836, elasticity in mechanics), Gustav Kirchhoff (1824–1887) or John William Strutt Rayleigh (1842–1919, author of *The theory of sound*, 1896), among others. Regarding the propagation of waves in plates, the work of Horace Lamb (1849–1934) has also contributed in the increase of understanding, with the incorporation of the eponymous *Lamb waves*. The work of George Green (1793–1841) and its contribution *An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism* (1828) is also revealed to produce crucial advances for the wave problems, with the subsequent Green's functions. This long history of wave studies is obviously made of trials, mistakes and corrections to eventually reach the current state of the art, but it is important to keep in mind those progressive steps. More details regarding the historical review can be found, for example, in Aki and Richards (2002); Dahlen and Tromp (1998); Rao (2007).

1.3 Formulation of the wave equations

The wave equations give a mathematical representation of the propagation of waves in a medium. We consider mechanical waves, which means that they require a physical material (opposed to electromagnetic waves that can travel in vacuum). It is easy to comprehend mechanical waves from the resulting oscillations of the matter they produce, where *matter* here naturally refers to the medium in which they propagate. Many images can be used to illustrate this oscillating behavior, such as the traditional thinking of throwing an object into a fluid, which induces the circular oscillations visible at the surface. We have also thought of an earthquake and the vibrations that can be experienced in the near area around the epicenter, created by the wave propagation. More generally we are interested in the resulting effect of a physical perturbation through an area of interest.

The wave equations are obtained from the classical (Newtonian) mechanics (by opposition to the quantum mechanics) and arise naturally from the revolutionary work of Isaac Newton (1642–1727) in *Philosophiae naturalis principia mathematica* (1687) whom intro-

duced the underlying *Newton's laws*, which are the essential point of departure in defining the wave equations. The formulation of the wave equation can be found in many books and courses. Here we especially refer to the work of Aki and Richards (2002); Slawinski (2010); Tsvankin and Grechka (2011). The reader is warmly encouraged towards these productions for additional details with precise geophysical context in the framework of linear elasticity.

1.3.1 Physical quantities

It is important to distinguish the Lagrangian and the Eulerian approaches to describe the wave phenomena. The Lagrangian approach focuses on the particles itself and it is the framework for the elastic propagation (which occurs in solid material) where the displacement is considered. Indeed by displacement is referred the displacement of the medium particles. The Eulerian approach focuses on an area of space instead, whatever particles appear to be in the considered region. In the context of acoustic media, i.e. fluid, it is the natural approach when investigating the pressure field. We refer to Aki and Richards (2002) for further discussion, noting that this distinction is not critical in the seismic context.

Hence we have two quantities of interest for the wave propagation. The displacement, denoted $\boldsymbol{u}$ is given in meter (m) and usually accounts for elastic propagation. The pressure p is given in Pascal (Pa) and is of interest for propagation in acoustic, i.e. fluid media. The bold notation for the displacement indicates a vector, with size given by the dimensions, see below.

1.3.2 Newton's second law

The Newton's second law (Newton, 1687) gives a fundamental relation between the forces applied to a body and the acceleration of this body. Namely, the sum of the different forces equals the acceleration multiplied by the mass, under the important assumption that the mass remains unchanged in the considered body. This assumption is always imposed for the geophysical situations we investigate, and appears natural. Indeed the geophysical time scale is much longer than the duration of the phenomena we study.

Let us consider a body Ω as a three-dimensional domain of space: $\Omega \subset \mathbb{R}^3$. We further denote the boundary of this domain by Γ. The spacial position is represented in the Cartesian coordinates by the variable $\mathbf{x} = (x, y, z)$ and t stands for the time variable. We introduce $\boldsymbol{a}(\mathbf{x}, t)$ the vector acceleration of this body, encompassing the components for the three directions in the Cartesian system of coordinates. We assume that n forces

$\boldsymbol{F_i}(\mathbf{x}, t)$, where $i \in \{1, \dots, n\}$ (with similar vectorial notation as for the acceleration) are applied to the body. The mass of the body, m, is defined to be invariable, as required by the assumption. The Newton's second law states that

$$\sum_i \boldsymbol{F_i}(\mathbf{x}, t) = m\boldsymbol{a}(\mathbf{x}, t). \tag{1.1}$$

To further study the forces apply on the domain of interest, the body is considered as the addition of infinitesimally small particles $\mathrm{d}\Omega$. The associated density is denoted by $\rho(\mathbf{x})$ and defined by

$$\rho(\mathbf{x}) = \frac{m}{\mathrm{d}\Omega}.$$

We see that the density does not depend on the time, which is a consequence of the invariable mass assumption.

The acceleration is recast as the second derivative of the vector displacement $\boldsymbol{u}(\mathbf{x}, t)$ such that

$$\boldsymbol{a}(\mathbf{x}, t) = \frac{\partial^2 \boldsymbol{u}(\mathbf{x}, t)}{\partial t^2}.$$

Remark 1.1. *We can briefly recall the standard units using the International System of Units (SI) for the different quantities:*

- *the mass is given in kilogram:* kg*;*
- *the displacement is given in meter:* m*;*
- *the acceleration is given in meter per second squared:* $\mathrm{m\,s^{-2}}$*;*
- *the force is given in Newton:* N*, which is equivalent to kilogram meter per second squared:* $1\ \mathrm{kg\,m\,s^{-2}} = 1\ \mathrm{N}$*;*
- *the density is given in kilogram per meter cubed:* $\mathrm{kg\,m^{-3}}$ *or kilogram per meter squared in two dimensions,* $\mathrm{kg\,m^{-2}}$*.*

1.3.3 Decomposition of the forces

The different forces that apply on the body are separated in two types, the ones that apply in the interior of Ω ('inside'), that we denote $\boldsymbol{F_b}(\mathbf{x}, t)$ (for body force) and the forces onto the surface Γ, say $\boldsymbol{F_s}(\mathbf{x}, t)$ (for surface force). The body force applies on all particles of the domain $\mathrm{d}\Omega$ meanwhile the surface force only concerns the particles of the boundary of Ω, i.e. the ones that are on Γ. For the interior, we separate the gravitation (the weight of the object) from other forces so that

$$\mathrm{d}\boldsymbol{F_b}(\mathbf{x}, t) = \rho(\mathbf{x})\boldsymbol{g}(\mathbf{x}, t)\,\mathrm{d}\Omega + \rho(\mathbf{x})\boldsymbol{f}(\mathbf{x}, t)\,\mathrm{d}\Omega,$$

where $\mathrm{d}\boldsymbol{F}$ indicates the force on the infinitesimal domain, $\boldsymbol{g}(\mathbf{x}, t)$ is the acceleration due to gravity, with typical value approximated as 9.81 $\mathrm{m\,s^{-2}}$ (with small variation depending on the location of the body on Earth). The sign of the gravitation can also be adjusted depending on the convention of the orientation of the axis. Standardly with the Cartesian basis it is only existing for the direction $-\overrightarrow{z}$. Non gravity forces are taken into account by $\boldsymbol{f}(\mathbf{x}, t)$.

The surface forces are given as the composition of the different constraints applying in the three Cartesian directions. We introduce the normal unit vector at the surface $\boldsymbol{\nu} = (\overrightarrow{\nu_x}, \overrightarrow{\nu_y}, \overrightarrow{\nu_z})$. In every direction, the corresponding surface force is the sum of the contribution on each normal directions to this surface. This is illustrated Figure 1.3, where we only represent the component for the y-axis on a selected part of the surface and where we use the notation σ_{ij} to identify the forces.

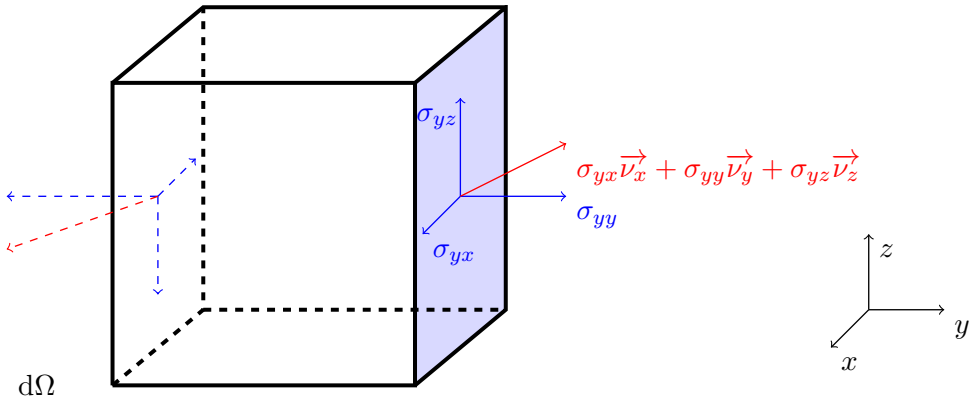


Figure 1.3: Illustration of the decomposition of the surface force along the y-axis considering a cubical shape of a boundary particle $\mathrm{d}\Omega$, using the Cartesian system of coordinates. The total force along the (x, z) (blue) plane surface is represented by the red arrow. It is decomposed onto the three normal axes, marked by the blue arrows. The normal force for the considered surface is $\sigma_{yy}(\mathbf{x}, t)$, parallel forces contributing are $\sigma_{yx}(\mathbf{x}, t)$ and $\sigma_{yz}(\mathbf{x}, t)$, where σ_{yx} and σ_{yz} are perpendicular. At equilibrium, the forces applied on surfaces of both sides are balanced, this is represented with dashed arrows on the opposite left surface.

The surface force is then given over the boundary Γ by,

$$\mathrm{d}\boldsymbol{F_s}(\mathbf{x}, t) = \underline{\sigma}(\mathbf{x}, t)\boldsymbol{n}\,\mathrm{d}\Gamma.$$

We have introduced $\underline{\sigma}(\mathbf{x}, t)$, which is the (Cauchy) stress tensor of order two (named after Augustin Louis Cauchy, 1789–1857, who introduced the mathematical formulation of stress, see (Slawinski, 2010, Section 2.3) for further discussion) and can be represented by the following three by three matrix

$$\underline{\sigma}(\mathbf{x}, t) = \begin{bmatrix} \sigma_{xx}(\mathbf{x}, t) & \sigma_{xy}(\mathbf{x}, t) & \sigma_{xz}(\mathbf{x}, t) \\ \sigma_{yx}(\mathbf{x}, t) & \sigma_{yy}(\mathbf{x}, t) & \sigma_{yz}(\mathbf{x}, t) \\ \sigma_{zx}(\mathbf{x}, t) & \sigma_{zy}(\mathbf{x}, t) & \sigma_{zz}(\mathbf{x}, t) \end{bmatrix}. \tag{1.2}$$

Remark. *An order d tensor in dimension n can be seen as a multi-dimensional array (similar to a hypermatrix, generalization of matrices). More precisely the dimensions of the underlying array is n^d. For example, the three-dimensional order two stress tensor $\underline{\sigma}(\boldsymbol{x}, t)$ is represented with an array of size 3×3, thus a squared matrix, as in the Equation (1.2).*

The state of equilibrium of the considered medium implies that the sum of the momenta applied to the body equates to zero. Using the conservation of mass, (Slawinski, 2010, Section 2.7), it gives that $\underline{\sigma}(\mathbf{x}, t)$ is symmetric so that $\sigma_{ij} = \sigma_{ji}$. Consequently, there are only six remaining independent coefficients instead of the initial nine coefficients such that

$$\underline{\sigma}(\mathbf{x}, t) = \begin{bmatrix} \sigma_{xx}(\mathbf{x}, t) & \sigma_{xy}(\mathbf{x}, t) & \sigma_{xz}(\mathbf{x}, t) \\ \sigma_{xy}(\mathbf{x}, t) & \sigma_{yy}(\mathbf{x}, t) & \sigma_{yz}(\mathbf{x}, t) \\ \sigma_{xz}(\mathbf{x}, t) & \sigma_{yz}(\mathbf{x}, t) & \sigma_{zz}(\mathbf{x}, t) \end{bmatrix}.$$

1.3.4 General wave equation

We have developed the concepts to differentiate surface and interior forces in the domain of interest. The definitions of the forces are incorporated in the main Newton's second law, Equation (1.1). By integrating over the domain, it leads to

$$\int_\Gamma \underline{\sigma}(\mathbf{x}, t)\boldsymbol{n} \, \mathrm{d}\Gamma + \int_\Omega \rho(\mathbf{x})\boldsymbol{g}(\mathbf{x}, t) + \rho(\mathbf{x})\boldsymbol{f}(\mathbf{x}, t) \, \mathrm{d}\Omega = \int_\Omega \rho(\mathbf{x}) \frac{\partial^2 \boldsymbol{u}(\mathbf{x}, t)}{\partial t^2} \, \mathrm{d}\Omega. \tag{1.3}$$

Remark 1.2. *The integral notation in the above equation should be denoted in three dimensions by three integral signs for the volume ($\iiint_\Omega$) to be strictly correct. Similarly the further boundary integrals should have two signs ($\iint_\Gamma$). We implicitly consider this repetition and always use only one sign to simplify the expressions because it is naturally implicitly induced by the index of the integral, Γ or Ω.*

The use of the Ostrogradsky's theorem (which is also referred to as the divergence, Gauss, or Green–Ostrogradsky theorem) requires assumptions on Ω to be compact and $\underline{\sigma}(\mathbf{x}, t)$ to be differentiable. Then this theorem (named after Mikhaïl Ostrogradski, 1801–1862) states that

$$\int_\Omega \nabla \cdot \underline{\sigma}(\mathbf{x}, t) \, \mathrm{d}\Omega = \int_\Gamma \underline{\sigma}(\mathbf{x}, t).\boldsymbol{n} \, \mathrm{d}\Gamma,$$

where $(\nabla\cdot)$ stands for the divergence. Following the use of the theorem, the Equation (1.3) can be rewritten equivalently

$$\int_\Omega \rho(\mathbf{x}) \frac{\partial^2 \boldsymbol{u}(\mathbf{x}, t)}{\partial t^2} - \nabla \cdot \underline{\sigma}(\mathbf{x}, t) - \rho(\mathbf{x})\boldsymbol{g}(\mathbf{x}, t) - \rho(\mathbf{x})\boldsymbol{f}(\mathbf{x}, t) \, \mathrm{d}\Omega = 0.$$

From this formulation, the common assumption is to neglect the gravitational force and consequently remove the term $(\rho(\mathbf{x})\boldsymbol{g}(\mathbf{x},t))$. This can be justify by the weight of the object particles, easily negligible as compared to the forces considered in geophysics (earthquake, air gun, etc). Only remains the other interior forces, $\boldsymbol{f}$, which can possibly incorporate the source of the phenomenon.

Because the equation is true for any arbitrary volume Ω, the equality with zero can be established for the inner part of the integrals, it verifies

$$\rho(\mathbf{x})\frac{\partial^2 \boldsymbol{u}(\mathbf{x},t)}{\partial t^2} = \nabla \cdot \underline{\sigma}(\mathbf{x},t) + \rho(\mathbf{x})\boldsymbol{f}(\mathbf{x},t). \tag{1.4}$$

This defines the general wave equation in terms of the displacement. It can be seen as the mathematical representation of the propagation of waves and where the only assumptions are to neglect the gravitational force, the conservation of mass and state of equilibrium. It represents the general relation between the displacement $\boldsymbol{u}$ and the constraint $\underline{\sigma}$ applied to the domain of interest. In the case where no body forces are included, then $\boldsymbol{f} = 0$ and the wave equation becomes

$$\rho(\mathbf{x})\frac{\partial^2 \boldsymbol{u}(\mathbf{x},t)}{\partial t^2} = \nabla \cdot \underline{\sigma}(\mathbf{x},t). \tag{1.5}$$

Here the wave equation is expressed depending on the displacement $\boldsymbol{u}$ but it is straightforward to obtain similar formulation using the velocity $\boldsymbol{v}(\mathbf{x},t) = \partial_t \boldsymbol{u}(\mathbf{x},t)$ or, less commonly in literature, the acceleration as it has been defined above.

1.3.5 Constitutive law for linear elasticity

The resolution of the wave equation (1.4) signifies the recovery of the displacement field $\boldsymbol{u}$. In fact the stress tensor $\underline{\sigma}$ contains the properties that characterize the medium, as well as the strain coefficients. Robert Hooke (1635–1703) first noticed the linear relation between the force applied to a string and its length. The extension of this relation was later found for materials in the context of linear elasticity, meaning linear relation between the stress and the strain.

Hence the *Hooke's law* gives the coefficients of the stress tensor $\underline{\sigma}$ as a linear combination of stiffness (or elasticity) coefficients C (which represents the resistance of the body to deformation) and strain coefficients ϵ (which are related to the deformation of the body). In three dimensions we have

$$\sigma_{i,j}(\mathbf{x},t) = \sum_k \sum_l C_{i,j,k,l}(\mathbf{x})\epsilon_{k,l}(\mathbf{x},t), \tag{1.6}$$

where $(i,j,k,l) = \{x,y,z\}$ stand for the directions. The stiffness tensor $\underline{\underline{C}}(\mathbf{x})$ (or elastic-

ity tensor) encompasses the coefficients $C_{i,j,k,l}(\mathbf{x})$ which are representative of the material properties. Those coefficients do not depend on the time in general, assuming the relatively small duration (compared to the geophysical time scale) of the event we analyze, as we already mentioned for the density. Therefor, we do not consider any time dependency in the stiffness tensor, it only depends on the space variable. It is a tensor of order four (i.e. a three by three by three by three array for three-dimensional cases). $\underline{\epsilon}(\mathbf{x},t)$ is the strain tensor, it is of order two with coefficients depending on the displacement such that

$$\epsilon_{i,j}(\mathbf{x},t) = \frac{1}{2}\big(\partial_j u_i(\mathbf{x},t) + \partial_i u_j(\mathbf{x},t)\big). \tag{1.7}$$

Considering the matrix representation for the tensor $\underline{\epsilon}(\mathbf{x},t)$, which is clearly symmetric, the resulting three by three matrix is written as

$$\underline{\epsilon}(\mathbf{x},t) = \begin{bmatrix} \epsilon_{xx}(\mathbf{x},t) & \epsilon_{xy}(\mathbf{x},t) & \epsilon_{xz}(\mathbf{x},t) \\ \epsilon_{xy}(\mathbf{x},t) & \epsilon_{yy}(\mathbf{x},t) & \epsilon_{yz}(\mathbf{x},t) \\ \epsilon_{xz}(\mathbf{x},t) & \epsilon_{yz}(\mathbf{x},t) & \epsilon_{zz}(\mathbf{x},t) \end{bmatrix}.$$

The stiffness tensor $\underline{\underline{C}}(\mathbf{x})$, being of order four, has originally $3^4 = 81$ different coefficients. The symmetry of the stress tensor $\underline{\sigma}$ gives $\sigma_{i,j} = \sigma_{j,i}$ and involves that $C_{i,j,k,l} = C_{j,i,k,l}$, reducing the number of independent coefficients to fifty-four (54). Using the symmetry of the strain tensor $\underline{\epsilon}$, one can show that $C_{i,j,k,l} = C_{i,j,l,k}$. At this point the stiffness tensor has thirty-six (36) different coefficients and can in particular be pictured as a six by six matrix.

In order to visualize the formulation in a clearer fashion, the Voigt notation (named after Woldemar Voigt, 1850–1919) is introduced. It allows us to represent the Equation (1.6) in a matrix vector multiplication way. First the six independent coefficients of the tensors $\underline{\sigma}(\mathbf{x},t)$ and $\underline{\epsilon}(\mathbf{x},t)$ are concatenated and represented in a vector, such that

$$\underline{\sigma}(\mathbf{x},t) = (\sigma_{xx}(\mathbf{x},t), \sigma_{yy}(\mathbf{x},t), \sigma_{zz}(\mathbf{x},t), \sigma_{yz}(\mathbf{x},t), \sigma_{xz}(\mathbf{x},t), \sigma_{xy}(\mathbf{x},t))^T,$$

where ^{T} stands for the transposed, and similarly for $\underline{\epsilon}(\mathbf{x},t)$. Following the same idea, the thirty six coefficients of the stiffness tensor $C(\mathbf{x})$ are replaced by a six by six matrix and

the Equation (1.6) is written as

$$\begin{pmatrix} \sigma_{xx}(\mathbf{x},t) \\ \sigma_{yy}(\mathbf{x},t) \\ \sigma_{zz}(\mathbf{x},t) \\ \sigma_{yz}(\mathbf{x},t) \\ \sigma_{xz}(\mathbf{x},t) \\ \sigma_{xy}(\mathbf{x},t) \end{pmatrix} = \begin{pmatrix} C_{11}(\mathbf{x}) & C_{12}(\mathbf{x}) & C_{13}(\mathbf{x}) & C_{14}(\mathbf{x}) & C_{15}(\mathbf{x}) & C_{16}(\mathbf{x}) \\ C_{21}(\mathbf{x}) & C_{22}(\mathbf{x}) & C_{23}(\mathbf{x}) & C_{24}(\mathbf{x}) & C_{25}(\mathbf{x}) & C_{26}(\mathbf{x}) \\ C_{31}(\mathbf{x}) & C_{32}(\mathbf{x}) & C_{33}(\mathbf{x}) & C_{34}(\mathbf{x}) & C_{35}(\mathbf{x}) & C_{36}(\mathbf{x}) \\ C_{41}(\mathbf{x}) & C_{42}(\mathbf{x}) & C_{43}(\mathbf{x}) & C_{44}(\mathbf{x}) & C_{45}(\mathbf{x}) & C_{46}(\mathbf{x}) \\ C_{51}(\mathbf{x}) & C_{52}(\mathbf{x}) & C_{53}(\mathbf{x}) & C_{54}(\mathbf{x}) & C_{55}(\mathbf{x}) & C_{56}(\mathbf{x}) \\ C_{61}(\mathbf{x}) & C_{62}(\mathbf{x}) & C_{63}(\mathbf{x}) & C_{64}(\mathbf{x}) & C_{65}(\mathbf{x}) & C_{66}(\mathbf{x}) \end{pmatrix} \begin{pmatrix} \epsilon_{xx}(\mathbf{x},t) \\ \epsilon_{yy}(\mathbf{x},t) \\ \epsilon_{zz}(\mathbf{x},t) \\ 2\epsilon_{yz}(\mathbf{x},t) \\ 2\epsilon_{xz}(\mathbf{x},t) \\ 2\epsilon_{xy}(\mathbf{x},t) \end{pmatrix}. \tag{1.8}$$

This formulation allows a clear understanding of how the properties of the media, contained in the stiffness coefficients $C(\mathbf{x})$, impact the wave equations following a modification of the stress tensor. Eventually, one can derive the formulation and using the arbitrariness of the order of differentiation and Schwarz's theorem (Slawinski, 2010, Section 4.2), it implies that $C_{i,j,k,l} = C_{k,l,i,j}$. This reduces the thirty six independent coefficients of the stiffness tensor to twenty one. Hence the formulation (1.8) can be rewritten where $\underline{\underline{C}}$ is represented with a symmetric matrix such that

$$\begin{pmatrix} \sigma_{xx}(\mathbf{x},t) \\ \sigma_{yy}(\mathbf{x},t) \\ \sigma_{zz}(\mathbf{x},t) \\ \sigma_{yz}(\mathbf{x},t) \\ \sigma_{xz}(\mathbf{x},t) \\ \sigma_{xy}(\mathbf{x},t) \end{pmatrix} = \begin{pmatrix} C_{11}(\mathbf{x}) & C_{12}(\mathbf{x}) & C_{13}(\mathbf{x}) & C_{14}(\mathbf{x}) & C_{15}(\mathbf{x}) & C_{16}(\mathbf{x}) \\ C_{12}(\mathbf{x}) & C_{22}(\mathbf{x}) & C_{23}(\mathbf{x}) & C_{24}(\mathbf{x}) & C_{25}(\mathbf{x}) & C_{26}(\mathbf{x}) \\ C_{13}(\mathbf{x}) & C_{23}(\mathbf{x}) & C_{33}(\mathbf{x}) & C_{34}(\mathbf{x}) & C_{35}(\mathbf{x}) & C_{36}(\mathbf{x}) \\ C_{14}(\mathbf{x}) & C_{24}(\mathbf{x}) & C_{34}(\mathbf{x}) & C_{44}(\mathbf{x}) & C_{45}(\mathbf{x}) & C_{46}(\mathbf{x}) \\ C_{15}(\mathbf{x}) & C_{25}(\mathbf{x}) & C_{35}(\mathbf{x}) & C_{45}(\mathbf{x}) & C_{55}(\mathbf{x}) & C_{56}(\mathbf{x}) \\ C_{16}(\mathbf{x}) & C_{26}(\mathbf{x}) & C_{36}(\mathbf{x}) & C_{46}(\mathbf{x}) & C_{56}(\mathbf{x}) & C_{66}(\mathbf{x}) \end{pmatrix} \begin{pmatrix} \epsilon_{xx}(\mathbf{x},t) \\ \epsilon_{yy}(\mathbf{x},t) \\ \epsilon_{zz}(\mathbf{x},t) \\ 2\epsilon_{yz}(\mathbf{x},t) \\ 2\epsilon_{xz}(\mathbf{x},t) \\ 2\epsilon_{xy}(\mathbf{x},t) \end{pmatrix}. \tag{1.9}$$

We utilize this visualization in the following Section 1.4 where we introduce the different types of medium. It results in further reduction of the number of independent coefficients in the stiffness tensor which leads to fundamental simplifications in the general wave equation formulation.

1.4 The wave equation depending on the media

The propagation of waves in a medium with time can be represented by an hyperbolic partial differential equation of order two, where the vector field displacement $\boldsymbol{u}(\mathbf{x},t)$ is solution of the general Equation (1.4). This formulation makes use of the stress tensor $\underline{\sigma}$, which itself is defined with the strain tensor $\underline{\epsilon}$, defined with the displacement field. The use of the stress tensor, see Equation (1.9), results in equations which can be complicated to deal with due to the number of parameters in the stiffness tensor $\underline{\underline{C}}$, which represents the material properties. Yet they can be simplified by assuming some special characterizations

in the nature of the medium. For instance symmetry axis in the material reduces the number of stiffness coefficients. Those simplifications deriving from assumptions on the medium are important for the resolution of the problem. Despite the general anisotropy of the Earth, they are considered in seismic to give fundamental reduction of the complexity of the equations and allow efficient numerical resolution.

Several classes of media can be found in the literature, for example monoclinic, trigonal, cubic, etc. Each of them has specific properties which involve dependencies on the stiffness tensor coefficients. The objective is to reduce the original twenty one independent coefficients. Details concerning such media can be found in Bos et al. (2004); Felício Fuck and Tsvankin (2009); Tsvankin and Grechka (2011); Slawinski (2010); Tsvankin (2012). Here we present the main characterizations that are used in geophysical applications. The notion of isotropy in the medium is one of the major property to reduce the number of stiffness coefficients. Isotropy refers to similar properties of the material in every directions. Alternatively this characteristic can be limited to a single direction, using a horizontal, vertical, or oriented axis of symmetry. These are referred to as Horizontal, Vertical and Titled Transverse Isotropy (HTI, VTI and TTI) respectively. A different aspect is the consideration of homogeneity in the medium parameters, meaning that the parameters do not depend on the space but are simply constant over the domain. This is independent of the type of media and can indistinctly holds in the context of isotropy or anisotropy. We will not use this assumption in applications but it allows simple analytical solutions of the wave equations via the Green's functions, see Subsection 1.8.3. In Section 1.5 we study the acoustic wave propagation, which refers to the propagation of waves in fluid.

One way to derive the stiffness tensor associated with specific class of media is to introduce matrices to represent the symmetry. The interested reader is referred to (Slawinski, 2010, Chapter 5) for the procedure, here we will simply review the resulting tensors.

Remark 1.3. *In order to clarify the equations, the time and space dependencies* $(\boldsymbol{x}, t)$ *is omitted in the following subsections. Hence we denote* $\underline{\underline{C}} = \underline{\underline{C}}(\boldsymbol{x})$, $\underline{\epsilon} = \underline{\epsilon}(\boldsymbol{x}, t)$ *and* $\underline{\sigma} = \underline{\sigma}(\boldsymbol{x}, t)$. *The dependencies are usually reminded when introducing new quantities.*

1.4.1 Orthotropic medium

Orthotropic media (also referred to as orthorhombic media, Tsvankin (1997)) have three orthogonal symmetric axis. In this case, the stiffness tensor is reduced to nine independent

coefficients with

$$\underline{\underline{C}}_{\text{orthotropic}} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{12} & C_{22} & C_{23} & 0 & 0 & 0 \\ C_{13} & C_{23} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} \end{pmatrix}.$$

This representation already gives a significant simplification by introducing twelve zero coefficients. Yet the number of coefficients can further be reduced.

1.4.2 Titled transverse isotropy

Transverse isotropy signifies that there is one plan of isotropy in the medium. It has been studied in Thomsen (1986) to show that it reshapes the stiffness tensor so that it is reduced to five independent coefficients. When assuming a vertical plane of isotropy, it gives

$$\underline{\underline{C}}^{\text{VTI}} = \begin{pmatrix} C_{11} & C_{11}-2C_{66} & C_{13} & 0 & 0 & 0 \\ C_{11}-2C_{66} & C_{11} & C_{13} & 0 & 0 & 0 \\ C_{13} & C_{13} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} \end{pmatrix}. \tag{1.10}$$

From those five different coefficients, the Thomsen's parameters, (Thomsen, 1986), are defined and represent dimensionless quantities useful in geophysics: $\delta(\mathbf{x})$, $\epsilon(\mathbf{x})$ and $\gamma(\mathbf{x})$ with following relation with the stiffness coefficients,

$$\begin{aligned} \delta(\mathbf{x}) &= \frac{(C_{13}(\mathbf{x}) + C_{44}(\mathbf{x}))^2 - (C_{33}(\mathbf{x}) - C_{44}(\mathbf{x}))^2}{2C_{33}(\mathbf{x})(C_{33}(\mathbf{x}) - C_{44}(\mathbf{x}))}, \\ \epsilon(\mathbf{x}) &= \frac{C_{11}(\mathbf{x}) - C_{33}(\mathbf{x})}{2C_{33}(\mathbf{x})}, \\ \gamma(\mathbf{x}) &= \frac{C_{66}(\mathbf{x}) - C_{44}(\mathbf{x})}{2C_{44}(\mathbf{x})}. \end{aligned} \tag{1.11}$$

The introduction of those three coefficients allows the distinction between the isotropic and anisotropic nature of the material. We shall see in Subsection 1.4.3 that the Thomsen's parameter are equal to zeros when reducing to the isotropic case. Furthermore, the special case when $\epsilon = \delta$ is called *elliptical anisotropy* and the *weak anisotropy* is defined when those parameters have relatively small values (< 0.2 in Thomsen (1986)).

It is natural to extend the formulation for any oriented axis of isotropy by including two angles which represent the inclination of the plane, those angles are denoted θ and ϕ and introduce the so-called Titled Transverse Isotropy (TTI) characterization. Let us introduce the associated rotation matrix R^{TTI} such that (Zhang et al., 2011),

$$R^{TTI} = \begin{pmatrix} \cos(\theta)\cos(\phi) & \cos(\theta)\sin(\phi) & -\sin(\theta) \\ -\sin(\phi) & \cos(\phi) & 0 \\ \sin(\theta)\cos(\phi) & \sin(\theta)\sin(\phi) & \cos(\theta) \end{pmatrix}.$$

Then the coefficients C^{TTI} of the TTI stiffness tensor are given from the VTI coefficients by

$$C^{TTI}_{ijkl} = \sum_{p=1}^{3}\sum_{q=1}^{3}\sum_{r=1}^{3}\sum_{s=1}^{3} R^{TTI}_{pi} R^{TTI}_{qj} R^{TTI}_{rk} R^{TTI}_{sl} C^{VTI}_{pqrs}. \tag{1.12}$$

We see that TTI media eventually give a full (symmetric) stiffness tensor. However, the number of actual independent coefficients is limited to seven. The rotation matrix R can further be written in different ways depending on the situation, e.g. Iversen and Pšenčík (2007). Note that the idea of defining TTI with a rotation of the VTI tensor using an underlying rotation matrix is also sometimes referred to as the *Bond transform*, with reference to the work of Bond (1943) for crystals physics.

Some special cases of TTI media can be considered, depending on the angles of inclination for the isotropic plane. We see that in the case where $\theta = \phi = 0$ we retrieve the VTI case as defined in Equation (1.10). If the plane of symmetry is perpendicular to the VTI one, we obtain the HTI medium, which gives similar non-zero pattern with

$$\underline{\underline{C}}^{\text{HTI}} = \begin{pmatrix} C_{11} & C_{13} & C_{13} & 0 & 0 & 0 \\ C_{13} & C_{33} & C_{33}-2C_{44} & 0 & 0 & 0 \\ C_{13} & C_{33}-2C_{44} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{55} \end{pmatrix}.$$

1.4.3 Elastic isotropy

The consideration of isotropic material is now regarded. In this case, the number of independent coefficients in the stiffness tensor $\underline{\underline{C}}(\mathbf{x})$ is reduced to two. Only $C_{33}(\mathbf{x})$ and $C_{44}(\mathbf{x})$

are independent and the Equation (1.8) becomes, with the use of the Voigt notation,

$$\begin{pmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{yz} \\ \sigma_{xz} \\ \sigma_{xy} \end{pmatrix} = \begin{pmatrix} C_{33} & C_{33}-2C_{44} & C_{33}-2C_{44} & 0 & 0 & 0 \\ C_{33}-2C_{44} & C_{33} & C_{33}-2C_{44} & 0 & 0 & 0 \\ C_{33}-2C_{44} & C_{33}-2C_{44} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{44} \end{pmatrix} \begin{pmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \epsilon_{zz} \\ 2\epsilon_{yz} \\ 2\epsilon_{xz} \\ 2\epsilon_{xy} \end{pmatrix}. \tag{1.13}$$

Hence, compared to the VTI, the isotropic case gives the following relation between the stiffness coefficients,

$$\begin{aligned} C_{11}(\mathbf{x}) &= C_{33}(\mathbf{x}), \\ C_{66}(\mathbf{x}) &= C_{44}(\mathbf{x}), \\ C_{13}(\mathbf{x}) &= C_{33}(\mathbf{x}) - 2C_{44}(\mathbf{x}). \end{aligned}$$

There is no notion of angle as we assume here similar properties in every direction. It is straightforward to notice from this definition that the Thomsen's parameter δ, ϵ and γ, Equation (1.11), are equal to zero for isotropic media. That is why those parameters are representative of the anisotropic nature of the material and we recall that Thomsen (1986) also refers to weak anisotropy when these coefficients are sufficiently small.

The two remaining coefficients $C_{33}(\mathbf{x})$ and $C_{44}(\mathbf{x})$ give the definition of the Lamé parameters (named after Gabriel Lamé, 1795–1890) $\lambda(\mathbf{x})$ and $\mu(\mathbf{x})$ such that

$$\begin{aligned} C_{33}(\mathbf{x}) &= \lambda(\mathbf{x}) + 2\mu(\mathbf{x}), \\ C_{44}(\mathbf{x}) &= \mu(\mathbf{x}). \end{aligned}$$

$\lambda(\mathbf{x})$ is simply referred to as the first Lamé parameter (or coefficient) and $\mu(\mathbf{x})$ is the *shear modulus* (or second Lamé parameter). The unit of those quantities is the Pascal (Pa, i.e. $\mathrm{kg\,m^{-1}\,s^{-2}}$). Other coefficients can also be defined, the most popular being the bulk modulus $\kappa(\mathbf{x})$ and the Poisson's ratio $\nu(\mathbf{x})$ (named after Siméon Poisson, 1781–1840), they are expressed with respect to the Lamé parameters in the case of elastic material by

$$\begin{aligned} \nu(\mathbf{x}) &= \frac{\lambda(\mathbf{x})}{2(\lambda(\mathbf{x}) + \mu(\mathbf{x}))}, \\ \kappa(\mathbf{x}) &= \lambda(\mathbf{x}) + \frac{2}{3}\mu(\mathbf{x}). \end{aligned} \tag{1.14}$$

The bulk modulus is expressed in Pascal while the Poisson's ratio is a dimensionless coefficient. From its definition and due to the positiveness of the quantities, the Poisson's ratio varies between 0 and 0.5. The bulk modulus also represents the invert of the compressibility. We introduce the P-wave speed $c_p(\mathbf{x})$, also referred to as compressional or primary wave speed, and the S-wave speed $c_s(\mathbf{x})$, also shear or secondary wave speed, which incorporate

the density such that,

$$
\begin{aligned}
c_p(\mathbf{x}) &= \sqrt{\frac{\lambda(\mathbf{x}) + 2\mu(\mathbf{x})}{\rho(\mathbf{x})}}, \\
c_s(\mathbf{x}) &= \sqrt{\frac{\mu(\mathbf{x})}{\rho(\mathbf{x})}}.
\end{aligned}
\tag{1.15}
$$

They are expressed in $\mathrm{m\,s^{-1}}$ and correspond to the wave velocity in the medium. The P-wave speed naturally represents the velocity of the P-waves while the S-wave speed is the velocity of the shear waves (see Figure 1.2), see the later Remark 1.4. Clearly because λ and μ are positive, the P-wave speed has higher values than the S-wave speed and the P-waves propagate faster than the S-waves.

We can rewrite the Equation (1.13) replacing the stiffness coefficients with the Lamé parameters expression,

$$
\begin{pmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{yz} \\ \sigma_{xz} \\ \sigma_{xy} \end{pmatrix} = \begin{pmatrix} \lambda+2\mu & \lambda & \lambda & 0 & 0 & 0 \\ \lambda & \lambda+2\mu & \lambda & 0 & 0 & 0 \\ \lambda & \lambda & \lambda+2\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu \end{pmatrix} \begin{pmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \epsilon_{zz} \\ 2\epsilon_{yz} \\ 2\epsilon_{xz} \\ 2\epsilon_{xy} \end{pmatrix}
$$

$$
= \begin{pmatrix} \lambda(\epsilon_{xx}+\epsilon_{yy}+\epsilon_{zz}) + 2\mu\epsilon_{xx} \\ \lambda(\epsilon_{xx}+\epsilon_{yy}+\epsilon_{zz}) + 2\mu\epsilon_{yy} \\ \lambda(\epsilon_{xx}+\epsilon_{yy}+\epsilon_{zz}) + 2\mu\epsilon_{zz} \\ 2\mu\epsilon_{yz} \\ 2\mu\epsilon_{xz} \\ 2\mu\epsilon_{xy} \end{pmatrix},
$$

where the space and time dependencies are omitted. Here the Voigt notation can be substituted back to the original matrix representation for the second order tensors,

$$
\begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{pmatrix} = \lambda \begin{pmatrix} \epsilon_{xx}+\epsilon_{yy}+\epsilon_{zz} & 0 & 0 \\ 0 & \epsilon_{xx}+\epsilon_{yy}+\epsilon_{zz} & 0 \\ 0 & 0 & \epsilon_{xx}+\epsilon_{yy}+\epsilon_{zz} \end{pmatrix}
$$

$$
+2\mu \begin{pmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{xy} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{xz} & \epsilon_{yz} & \epsilon_{zz} \end{pmatrix}.
$$

Finally, the stress tensor for elastic isotropic media follows to be

$$\underline{\sigma} = \lambda Tr(\underline{\epsilon}) I_d + 2\mu\underline{\epsilon},$$

where $Tr(\cdot)$ stands for the trace operator of the matrix (the sum of the diagonal coefficients) and I_d is the three by three identity matrix. This formulation is injected into the wave equation where the body forces have been ignored for simplicity, Equation (1.5), to give,

$$\rho\frac{\partial^2 \boldsymbol{u}}{\partial t^2} = \nabla \cdot (\lambda Tr(\underline{\epsilon}) I_d) + \nabla \cdot (2\mu\underline{\epsilon}). \tag{1.16}$$

The wave equation can finally be reformulated by replacing the strain tensor $\underline{\epsilon}$ with the displacement field $\boldsymbol{u}$, using the definition of Equation (1.7). Let us consider the first term in the right-hand side of the equation, $(\nabla \cdot (\lambda Tr(\underline{\epsilon}) I_d))$, we note that by definition

$$Tr(\underline{\epsilon}) = \epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz} = \frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z} = \nabla \cdot (\boldsymbol{u}),$$

which gives

$$\begin{aligned} \nabla \cdot (\lambda Tr(\epsilon) I_d) &= \nabla \cdot \begin{pmatrix} \lambda\nabla \cdot (\boldsymbol{u}) & 0 & 0 \\ 0 & \lambda\nabla \cdot (\boldsymbol{u}) & 0 \\ 0 & 0 & \lambda\nabla \cdot (\boldsymbol{u}) \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x}(\lambda\nabla \cdot (\boldsymbol{u})) \\ \frac{\partial}{\partial y}(\lambda\nabla \cdot (\boldsymbol{u})) \\ \frac{\partial}{\partial z}(\lambda\nabla \cdot (\boldsymbol{u})) \end{pmatrix} \\ &= \nabla(\lambda\nabla \cdot \boldsymbol{u}), \end{aligned} \tag{1.17}$$

introducing the gradient ∇. For the second term in the right-hand side of Equation (1.16), $(\nabla \cdot (2\mu\underline{\epsilon}))$, we develop

$$\begin{aligned} 2\mu\epsilon &= \mu \begin{pmatrix} 2\frac{\partial u_x}{\partial x} & \frac{\partial u_y}{\partial x} + \frac{\partial u_x}{\partial y} & \frac{\partial u_z}{\partial x} + \frac{\partial u_x}{\partial z} \\ \frac{\partial u_y}{\partial x} + \frac{\partial u_x}{\partial y} & 2\frac{\partial u_y}{\partial y} & \frac{\partial u_z}{\partial y} + \frac{\partial u_y}{\partial z} \\ \frac{\partial u_z}{\partial x} + \frac{\partial u_x}{\partial z} & \frac{\partial u_z}{\partial y} + \frac{\partial u_y}{\partial z} & 2\frac{\partial u_z}{\partial z} \end{pmatrix} \\ &= \mu \begin{pmatrix} \frac{\partial u_x}{\partial x} & \frac{\partial u_x}{\partial y} & \frac{\partial u_x}{\partial z} \\ \frac{\partial u_y}{\partial x} & \frac{\partial u_y}{\partial y} & \frac{\partial u_y}{\partial z} \\ \frac{\partial u_z}{\partial x} & \frac{\partial u_z}{\partial y} & \frac{\partial u_z}{\partial z} \end{pmatrix} + \mu \begin{pmatrix} \frac{\partial u_x}{\partial x} & \frac{\partial u_y}{\partial x} & \frac{\partial u_z}{\partial x} \\ \frac{\partial u_x}{\partial y} & \frac{\partial u_y}{\partial y} & \frac{\partial u_z}{\partial y} \\ \frac{\partial u_x}{\partial z} & \frac{\partial u_y}{\partial z} & \frac{\partial u_z}{\partial z} \end{pmatrix} \\ &= \mu\big[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\big], \end{aligned} \tag{1.18}$$

where ^{T} lies for the transposed of the matrix, so that a matrix A^T has coefficients $A^T_{ij} = A_{ji}$. We now inject Equations (1.17) and (1.18) in Equation (1.16). This defines the isotropic elastic wave equation:

$$\rho\frac{\partial^2 \boldsymbol{u}}{\partial t^2} - \nabla\big(\lambda\nabla\cdot\boldsymbol{u}\big) - \nabla\cdot\big(\mu\big[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\big]\big) = 0. \tag{1.19}$$

Vectorial identities can be applied to rewrite this equation in a slightly different fashion, we briefly explicit another popular version of the elastic isotropic equation, starting with the term $(\nabla(\lambda\nabla\cdot\boldsymbol{u}))$, we develop the gradient operator,

$$\nabla(\lambda\nabla\cdot\boldsymbol{u}) = \nabla\lambda\nabla\cdot\boldsymbol{u} + \lambda\nabla\nabla\cdot\boldsymbol{u}.$$

The second term, $\nabla\cdot\big(\mu\big[\nabla\boldsymbol{u}+(\nabla\boldsymbol{u})^T\big]\big)$, is rewritten after developing the divergence operator such that

$$\begin{aligned}\nabla\cdot\big(\mu\big[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\big]\big) =& \nabla\cdot\mu\big[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\big] + \mu\nabla\cdot\big[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\big]\\ =& \nabla\cdot\mu\big[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\big] + \mu\nabla\cdot(\nabla\boldsymbol{u}) + \mu\nabla\cdot\big([\nabla\boldsymbol{u}]^T\big).\end{aligned}$$

Vectorial identity provides $\nabla\cdot(\nabla\boldsymbol{u}) = \nabla\nabla\cdot\boldsymbol{u} - \nabla\times\nabla\times\boldsymbol{u}$, where $\nabla\times$ denotes the rotational. We also notice that $\nabla\cdot([\nabla\boldsymbol{u}]^T) = \nabla\nabla\cdot\boldsymbol{u}$, thus we can write,

$$\nabla\cdot\big(\mu\big[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\big]\big) = \nabla\cdot\mu\big[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\big] + 2\mu\nabla\nabla\cdot\boldsymbol{u} - \mu\nabla\times\nabla\times\boldsymbol{u}.$$

We finally replace in Equation (1.19) to obtain

$$\rho\frac{\partial^2 \boldsymbol{u}}{\partial t^2} - \nabla\lambda\nabla\cdot\boldsymbol{u} - \nabla\cdot\mu\big[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\big] - (\lambda+2\mu)\nabla\nabla\cdot\boldsymbol{u} + \mu\nabla\times\nabla\times\boldsymbol{u} = 0. \tag{1.20}$$

The propagation of waves in isotropic elastic medium is defined via an hyperbolic partial differential equation for which the two equivalent formulations (1.19) and (1.20) are the most commons. The general wave equation with stress, strain and stiffness tensors has been replaced by the formulation involving the physical parameters (the density and the Lamé coefficients λ and μ) and the displacement field $\boldsymbol{u}$.

Remark 1.4 (P- and S-waves decomposition)**.** *We have seen in Figures 1.1 and 1.2 that body waves propagating in solid can be separated into the P- and S-waves. In particular for elastic isotropy we have defined the associated P- and S-wave speeds, see Equation* (1.15)*. The differentiation appears also from the mathematical wave equation by employing the Helmholtz decomposition (named after Hermann Ludwig Ferdinand von Helmholtz,* 1821–1894 *and later extended by the work of William Vallance Douglas Hodge,* 1903–1975*). It is applied for the displacement field* $\boldsymbol{u}$*, which is expressed as a composition of two functions,*

$$\boldsymbol{u} = \nabla\phi + \nabla\times\boldsymbol{\psi}.$$

Here ϕ is a scalar function and $\boldsymbol{\psi}$ a vector field. We remind the general vectorial identities $\nabla \times (\nabla \phi) = 0$ and $\nabla \cdot (\nabla \times \boldsymbol{\psi}) = 0$. Therefore, the displacement is divided into two: $\nabla\phi$ is rotational free and actually corresponds to the P-waves (compressional waves) and $\nabla \times \boldsymbol{\psi}$, which is divergence free, is related to the shear, S-waves. Let us denote $\boldsymbol{u} = \boldsymbol{u}^{(p)} + \boldsymbol{u}^{(s)}$ where $\boldsymbol{u}^{(p)}$ and $\boldsymbol{u}^{(s)}$ are identified with $\nabla\phi$ and $\nabla \times \boldsymbol{\psi}$ respectively. From (Kupradze et al., 1976, Theorem 2.5 p.123), we have that, in the context of homogeneous media, $\boldsymbol{u}^{(p)}$ and $\boldsymbol{u}^{(s)}$ are each solution of

$$\begin{cases} (\Delta + k_p^2)\boldsymbol{u}^{(p)} = 0, \\ (\Delta + k_s^2)\boldsymbol{u}^{(s)} = 0, \end{cases}$$

where the wavenumbers k_p and k_s are given by

$$\begin{cases} k_p = \omega\sqrt{\dfrac{\rho}{\lambda + 2\mu}} = \dfrac{\omega}{c_p}, \\ k_s = \omega\sqrt{\dfrac{\rho}{\mu}} = \dfrac{\omega}{c_s}. \end{cases}$$

These new equations interestingly relate to the acoustic wave equation that is detailed in Section 1.5.

1.4.4 2D wave equations

We have presented the wave equations for a three dimensional domain of interest Ω. In this context the displacement field $\boldsymbol{u}$ has one component per directions, i.e. three: u_x, u_y and u_z. The two-dimensional formulation of the equations reduces this number to two, by suppressing one of the direction. It does not change the isotropic elastic wave Equation (1.19) (nor the acoustic wave propagation that is to follow in Section 1.5). However, a major modification of the two-dimensional approach lies in the tensors. The strain, stress and stiffness tensors loose one dimension and the Voigt notation must be written accordingly. In 2D, the Equation (1.9) becomes, where only x and z directions are kept,

$$\begin{pmatrix} \sigma_{xx}(\mathbf{x},t) \\ \sigma_{zz}(\mathbf{x},t) \\ \sigma_{zx}(\mathbf{x},t) \end{pmatrix} = \begin{pmatrix} C_{11}(\mathbf{x}) & C_{12}(\mathbf{x}) & C_{13}(\mathbf{x}) \\ C_{12}(\mathbf{x}) & C_{22}(\mathbf{x}) & C_{23}(\mathbf{x}) \\ C_{13}(\mathbf{x}) & C_{23}(\mathbf{x}) & C_{33}(\mathbf{x}) \end{pmatrix} \begin{pmatrix} \epsilon_{xx}(\mathbf{x},t) \\ \epsilon_{zz}(\mathbf{x},t) \\ 2\epsilon_{xz}(\mathbf{x},t) \end{pmatrix}.$$

Using the symmetry of the tensors, six independent coefficients remain in the 2D case.

In the context of orthotropic medium, the stiffness tensor is reduced to four coefficients and becomes

$$\underline{\underline{C}}_{\text{orthotropic 2D}} = \begin{pmatrix} C_{11} & C_{12} & 0 \\ C_{12} & C_{22} & 0 \\ 0 & 0 & C_{33} \end{pmatrix}.$$

In order to consider titled transverse isotropy in the model, we follow the same idea as with Equation (1.12) and the rotation matrix, with only one angle θ, is defined by

$$R_{2D}^{TTI} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}.$$

1.5 Acoustic propagation

The elastic case presented above corresponds to waves propagating through solid materials. When the medium is a fluid, we refer to the phenomenon as *acoustic* propagation. In such media, there is no shear waves and thus the shear modulus is null ($\mu = 0$). It is easily understandable that no shear appears in a fluid because it cannot have any aspect of rigidity. Then the Equation (1.19) is simplified and the acoustic isotropic wave equation becomes

$$\rho \frac{\partial^2 \boldsymbol{u}}{\partial t^2} - \nabla(\lambda \nabla \cdot \boldsymbol{u}) = 0.$$

It is most common to employ the bulk modulus, with given expression Equation (1.14), which in this case can directly substitute the first Lamé parameter λ, as μ is zero. Therefore, we obtain

$$\rho \frac{\partial^2 \boldsymbol{u}}{\partial t^2} - \nabla(\kappa \nabla \cdot \boldsymbol{u}) = 0.$$

This expression can further be developed,

$$\rho \frac{\partial^2 \boldsymbol{u}}{\partial t^2} - \big(\nabla \kappa\big)\big(\nabla \cdot \boldsymbol{u}\big) - \kappa \nabla\big(\nabla \cdot \boldsymbol{u}\big) = 0.$$

An important modification of the acoustic case appears as we can account for the (scalar) pressure field $p = p(\mathbf{x}, t)$ instead of the (vector) displacement field $\boldsymbol{u}$. The pressure and displacement can be linked under the assumption of incompressible fluid, with the Euler equations (reminded for example in Colton and Kress (1998); Kirsch (1996); Demanet (2014)). It defines

$$\begin{cases} \dfrac{\partial^2 \boldsymbol{u}(\mathbf{x}, t)}{\partial t^2} = -\dfrac{1}{\rho(\mathbf{x})} \nabla p(\mathbf{x}, t), \\ \dfrac{\partial p(\mathbf{x}, t)}{\partial t} = -\kappa(\mathbf{x}) \nabla \cdot \dfrac{\partial \boldsymbol{u}(\mathbf{x}, t)}{\partial t}, \end{cases}$$

where the gravity force is not considered according to our initial assumption. The second equation can be derived with respect to the time t and the first equation can be injected to the right-hand side, then the wave equation with respect to the pressure follows

$$\frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} = \kappa(\mathbf{x}) \nabla \cdot \left(\frac{1}{\rho(\mathbf{x})} \nabla p(\mathbf{x}, t) \right). \tag{1.21}$$

The rewriting in terms of pressure of the acoustic wave propagation produces a crucial modification of the equation. It is now a scalar equation and acknowledges one and only one equation, whatever the dimension of the domain. When considering the displacement, we have vectorial equations which are as many as the dimensions. Under the assumption that the density is constant, $\rho(\mathbf{x}) = \rho$, the vectorial identity $\nabla \cdot \nabla = \nabla^2 = \Delta$ reveals the Laplacian (named after Pierre-Simon de Laplace, 1749–1827), and the equation becomes

$$\frac{\partial^2 p(\mathbf{x},t)}{\partial t^2} - \frac{\kappa(\mathbf{x})}{\rho}\Delta p(\mathbf{x},t) = 0.$$

Replacing with the velocity $c_p(\mathbf{x}) = c(\mathbf{x}) = \sqrt{\dfrac{\kappa(\mathbf{x})}{\rho(\mathbf{x})}}$ gives

$$\frac{1}{c(\mathbf{x})^2}\frac{\partial^2 p(\mathbf{x},t)}{\partial t^2} - \Delta p(\mathbf{x},t) = 0. \tag{1.22}$$

Remark 1.5 (Acoustic anisotropy)**.** *Anisotropy can also be incorporated in acoustic media via some special characterization but it is physically difficult to apprehend and we have not used it in our work. We mention the work of Duveneck et al. (2008); Plessix and Cao (2011); Alkhalifah and Plessix (2014).*

1.6 Formulation of the time-harmonic waves propagation

When we consider the wave motion and intend to retrieve the wavefield, p or $\boldsymbol{u}$, the formulation in the time domain is natural because we think of the physical phenomenon involved. The time dependency of the field (expressed with the time variable t) is equally intuitive. Wave propagations are primarily defined by their starting and ending, which correspond intuitively to the temporal notion: when it starts (the perturbation generated by the source) and when it ends (when the waves have vanished completely or when they have left the area of interest). Similarly, the spatial location of such phenomena is equally intuitive and justifies the general spatio-temporal dependency of the quantity of interest, $(\mathbf{x}, t)$. Nonetheless, in our project, we only deal with time-harmonic formulation of the wave equations. The time-harmonic formulation is less natural to picture, because it does not account for the time variable but the frequency instead. This formulation can also be referred to as the *frequency domain*. Our motivation is mainly related to the numerical resolution of the wave equations and will be made explicit later.

1.6.1 Time-harmonic waves

The time-harmonic formulation, e.g. Colton and Kress (1998), of the wave equations is based on solutions of the form

$$\boldsymbol{u}(\mathbf{x},t) = \widehat{\boldsymbol{u}}(\mathbf{x})e^{-\mathrm{i}\omega t}, \tag{1.23}$$

where the time and space variable are separated. i represents the complex imaginary unit such that $\mathrm{i}^2 = -1$ and ω is the angular frequency with

$$\omega = 2\pi f, \tag{1.24}$$

where $f \in \mathbb{R}$ is the frequency in Hertz (Hz, named after Heinrich Rudolf Hertz, 1857–1894), $1\ \mathrm{Hz} = 1\ \mathrm{s}^{-1}$).

Remark 1.6. *In the sequel of this manuscript we often refer to the angular frequency ω simply with the term* frequency*. It is only motivated to simplify the writing, we believe ω is a common symbol which speaks for itself with an intrinsic 2π multiplication.*

This provides a very powerful rewriting of the equations, because the separation of variable directly impacts the derivatives, in particular

$$\begin{aligned}\frac{\partial \boldsymbol{u}(\mathbf{x},t)}{\partial t} &= -\,\mathrm{i}\omega\widehat{\boldsymbol{u}}(\mathbf{x})e^{-\mathrm{i}\omega t},\\ \frac{\partial^2 \boldsymbol{u}(\mathbf{x},t)}{\partial t^2} &= -\,\omega^2\widehat{\boldsymbol{u}}(\mathbf{x})e^{-\mathrm{i}\omega t}.\end{aligned}$$

Concerning the space derivative, the exponential part can simply be taken out of the gradient and the divergence operator. we illustrate the modifications for the isotropic elastic wave equation that we obtained Equation (1.19), injecting the Equation (1.23) gives

$$-\rho\omega^2\widehat{\boldsymbol{u}}e^{-\mathrm{i}\omega t} - \nabla\big(\lambda\nabla\cdot\widehat{\boldsymbol{u}}\big)e^{-\mathrm{i}\omega t} - \nabla\cdot\big(\mu\big[\nabla\widehat{\boldsymbol{u}} + (\nabla\widehat{\boldsymbol{u}})^T\big]\big)e^{-\mathrm{i}\omega t} = 0,$$

where the spatial dependency $\mathbf{x}$ is omitted. One can simplify the $e^{-\mathrm{i}\omega t}$ to obtain the time-harmonic elastic wave equation:

$$-\rho\omega^2\widehat{\boldsymbol{u}} - \nabla\big(\lambda\nabla\cdot\widehat{\boldsymbol{u}}\big) - \nabla\cdot\big(\mu\big[\nabla\widehat{\boldsymbol{u}} + (\nabla\widehat{\boldsymbol{u}})^T\big]\big) = 0.$$

There is no more dependency in time in this equation, it is instead replaced by the appearance of the frequency variable. We can reproduce the same steps for the other wave equations. The time-harmonic acoustic wave equation with respect to pressure is established from Equation (1.21),

$$-\omega^2\widehat{p}(\mathbf{x}) - \kappa(\mathbf{x})\nabla\cdot\left(\frac{1}{\rho(\mathbf{x})}\nabla\widehat{p}(\mathbf{x})\right) = 0.$$

When we consider the acoustic wave equation with constant density ρ, Equation (1.22),

the time-harmonic waves verify

$$\Big(-\frac{\omega^2}{c(\mathbf{x})^2} - \Delta \Big) \widehat{p}(\mathbf{x}) = 0.$$

This is the well known Helmholtz equation, named after Hermann Ludwig Ferdinand von Helmholtz (1821–1894). The wavenumber k and potential q are defined subsequently. The wavenumber is the ratio of the angular frequency and wave speed,

$$k(\mathbf{x}, \omega) = \left(\frac{\omega}{c(\mathbf{x})} \right),$$

the potential is the wavenumber squared, $q(\mathbf{x}, \omega) = k^2(\mathbf{x}, \omega)$.

Remark 1.7 (Complex frequency)**.** *We have defined ω with Equation (1.24), where f is naturally though as a scalar of $\mathbb{R}$. This can be extended with the consideration of* complex frequency *ω. Because only ω squared appears in the wave equation, we now define it such that*

$$-\omega^2 = (\sigma + 2\mathrm{i}\pi f)^2, \tag{1.25}$$

where $\sigma \in \mathbb{R}$ is a damping coefficient. In this configuration,

$$-\omega^2 = \sigma^2 + 4\mathrm{i}\pi f\sigma - 4\pi^2 f^2,$$

so that when the damping is set to zero, ω^2 is identified with $4\pi^2 f^2$, which corresponds with the standard real frequency. This is revealed to be particularly useful when we examine the inverse problem.

Remark 1.8. *We note that the partial differential equations (PDE) obtained for the time domain and their time-harmonic formulation counterpart, differ in types. The time domain PDE is hyperbolic while the time-harmonic PDE is weakly elliptic.*

1.6.2 Note on the Fourier transform

The relation between time-harmonic wave equation and the frequency domain is clear as the frequency parameter supplants the time. Then it is natural to think that the Fourier transform is closely related. It is indeed the case even if some precautions must be taken regarding the mathematical aspect.

The Fourier transform is the operator that maps from the time to the frequency domain. Because the frequency dependency of a signal is harder to picture than the common time dependency, we will first give some physical intuitions in a very simple setup. Let us consider a single oscillating particle of space in one dimension. The displacement of this particle, say $m(t)$, is pictured in Figure 1.4 (left). We have considered a quite naive oscillating motion

defined by a sum of sine and cosine functions. For the frequency domain representation, the first step is to formulate the displacement to retrieve the sine and cosine functions involved, such that

$$m(t) = \sum_n X_n \cos(2\pi f_n t) + Y_n \sin(2\pi f_n t), \tag{1.26}$$

where X_n and Y_n are scaling coefficients and f_n is a real parameter. This is the representation of the function in terms of Fourier series where f_n represents the frequency components of the signal, see Figure 1.4 (middle). Representing the particle motion in the frequency domain gives picks for the frequency components, see Figure 1.4 (right).

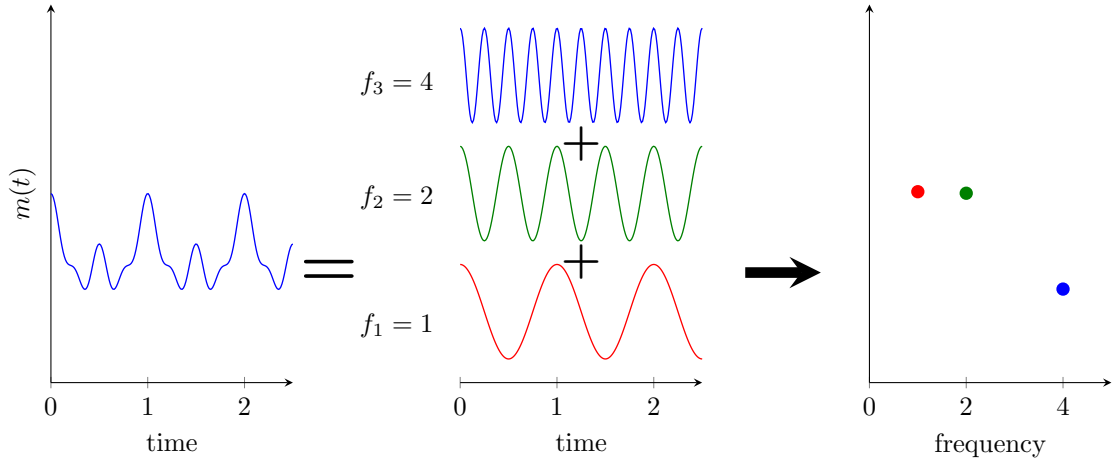


Figure 1.4: From time to frequency domain, illustration of the decomposition: the left plot represents a motion of a particle in one dimension with time. It is given by $m(t) = 10\cos(2\pi x) + 10\cos(2\pi 2x) + 5\cos(2\pi 4x)$. The middle plot shows how the initial motion is composed of a sum of cosine functions, with appropriate content indicated with f_j, $j = \{1, 2, 3\}$. The right plot translates the time-dependent motion into its frequency domain counterpart, where the amplitude of the frequency f_j content is connected to the weight of the cosines, i.e. the X_j in the Equation (1.26).

This simple example illustrates the relation between time and frequency domain, from the decomposition given in the Equation (1.26). In general, the signals do not have a discrete set of frequency component and generate a continuous representation in the frequency domain. Alternatively, the sum of sine and cosine functions can be substituted with an equivalent sum of exponential, such that,

$$m(t) = \sum_n \widehat{X}_n e^{2i\pi f_n t} + \widehat{Y}_n e^{-2i\pi f_n t}.$$

The mathematical way to step from time to frequency is the *Fourier transform*, which, as the name indicates, comes from the work of Joseph Fourier (1768–1830).

1.6.2.1 Definition

We consider an integrable function of time $h(t)$ such that

$$\begin{aligned} h \quad &: \quad \mathbb{R} \to \mathbb{C}, \\ &\quad\; t \to h(t). \end{aligned}$$

The Fourier transform is defined by $\mathfrak{F}(h) = \widehat{h}$ and introduce a new variable, ξ, following

$$\mathfrak{F}(h) \quad : \xi \to \widehat{h}(\xi),$$

with definition

$$\big(\mathfrak{F}(h)\big)(\xi) = \widehat{h}(\xi) = \int_{-\infty}^{+\infty} h(t) e^{-2\mathrm{i}\pi\xi t} dt.$$

Assuming $\widehat{h}$ is integrable, the inverse Fourier transform, $\mathfrak{F}^{-1}$ is given by

$$\big(\mathfrak{F}^{-1}(\widehat{h})\big)(t) = \int_{-\infty}^{+\infty} \widehat{h}(\xi) e^{2\mathrm{i}\pi\xi t} d\xi,$$

so that $\big(\mathfrak{F}^{-1}(\widehat{h})\big)(t) = h(t)$. In the above expressions, the frequency ξ is given in Hertz. The angular frequency ω can easily be incorporated with $\omega = 2\pi\xi$.

1.6.2.2 Derivation

The wave propagation equations in the time domain make appear the second derivative of the field with respect to the time t: $\partial^2 \boldsymbol{u}/\partial t^2$. An important property of the Fourier transform concerns the derivative, which is simplified via the use of integration by part. Let us consider h to be differentiable and assume $\partial h/\partial t$ to be differentiable as well. We have,

$$\begin{aligned} \left(\mathfrak{F}\Big(\frac{\partial h(t)}{\partial t}\Big)\right)(\xi) &= \int_{-\infty}^{+\infty} \frac{\partial h(t)}{\partial t} e^{-2\mathrm{i}\pi\xi t} dt \\ &= \big[h(t) e^{-2\mathrm{i}\pi\xi t}\big]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} h(t)(-2\mathrm{i}\pi\xi) e^{-2\mathrm{i}\pi\xi t} dt \\ &= (2\mathrm{i}\pi\xi) \int_{-\infty}^{+\infty} h(t) e^{-2\mathrm{i}\pi\xi t} dt \\ &= (2\mathrm{i}\pi\xi)\mathfrak{F}(h) = \mathrm{i}\omega\widehat{h}(\xi), \end{aligned}$$

where we have use the fact that h is integrable by definition. More generally, we can extend this formulation for higher order derivatives, in particular, for the second order derivative,

we have

$$\Big(\mathfrak{F}\Big(\frac{\partial^2 h(t)}{\partial t^2}\Big)\Big)(\xi) = -(2\pi\xi)^2\widehat{h}(\xi)$$
$$= -\omega^2\widehat{h}(\xi).$$

Remark 1.9. *We have used the special solutions to obtain the time-harmonic equations because it would not be correct to apply the Fourier transform directly onto the wave equations. In particular, because the solution of the time domain wave equations we have proposed is not in L^2, due to the absence of attenuation in the formulation. We have illustrated the intrinsic relation between the two approaches, but applying the Fourier transform requires assumptions on the solution first. We will indistinctly refer to* frequency domain *and* time-harmonic *in the rest of the manuscript. Regarding the complex frequency introduced in the Remark 1.7, it is actually related to the Laplace transform instead of the Fourier transform when $f = 0$.*

1.6.3 Summary of the time-harmonic equations

Because we will not use the time domain formulation in the sequel of the manuscript we use the notation $\boldsymbol{u}$ and p by abuse of notation and omit the $\widehat{\ }$ to clarify the equations. We review the main equations that we will use throughout the manuscript. We give the acoustic wave equation involving the pressure field, the elastic wave equation involving the displacement field and the general anisotropic wave equation:

$$\text{Helmholtz} \quad \big(-\omega^2 c(\mathbf{x})^{-2} - \Delta\big)p(\mathbf{x}) = 0, \tag{1.27}$$

$$\text{acoustic} \quad -\omega^2 p(\mathbf{x}) - \kappa(\mathbf{x})\nabla\cdot\left(\frac{1}{\rho(\mathbf{x})}\nabla p(\mathbf{x})\right) = 0, \tag{1.28}$$

$$\text{elastic} \quad -\rho(\mathbf{x})\omega^2\boldsymbol{u}(\mathbf{x}) - \nabla\big(\lambda(\mathbf{x})\nabla\cdot\boldsymbol{u}(\mathbf{x})\big) - \nabla\cdot\big(\mu(\mathbf{x})\big[\nabla\boldsymbol{u}(\mathbf{x}) + (\nabla\boldsymbol{u}(\mathbf{x}))^T\big]\big) = 0, \tag{1.29}$$

$$\text{anisotropic} \quad -\rho(\mathbf{x})\omega^2\boldsymbol{u}(\mathbf{x}) - \nabla\cdot\underline{\sigma}(\mathbf{x}) = 0. \tag{1.30}$$

At this moment, we have assumed no source at the right-hand side and have not yet specified the boundary conditions, we have simply equate the formulation to zero. The boundary terms are addressed in the following Section 1.8.

1.7 The consideration of attenuation in the media

The frequency domain formulation is advantageous to easily incorporate visco-acoustic behavior in materials, in particular by the consideration of *complex valued wave speeds*

leading to attenuation, Muller (1983). Attenuation refers to the progressive decrease in the waves energy with the distance to the source. This aspect appears natural when we think of any physical phenomenon, yet it is not clearly considered in the wave equations we have derived in the previous sections (note that attenuation doe not exist only in vacuum). Also, because the attenuation usually depends on the frequency, it is less convenient to account for it in the time domain formulation. For a comprehensive review of attenuating media and models, we refer to Aki and Richards (2002); Ursin and Toverud (2002). An important aspect is the causality of the models of attenuation, which is not always guaranteed. Here we only indicate the steps to define attenuation in the wave equations and the use of the Kolsky–Futterman model to represent it. Other models have been defined and are used in the literature; we mention, for example, the Kjartansson's model (Kjartansson, 1979) used by Liao and McMechan (1996), which also makes use of a complex velocity. We consider the acoustic wave propagation following the Helmholtz equation, (1.27), so the propagation only depends on the wavefield c.

The incorporation of attenuation in the propagation of waves is standardly realized by the requirement of a complex wavenumber k (e.g. Aki and Richards (2002); Ursin and Toverud (2002)) defined with

$$k(\mathbf{x},\omega) = \frac{\omega}{c_a(\mathbf{x},\omega)} = \frac{\omega}{c(\mathbf{x},\omega)} + \mathrm{i}\alpha(\mathbf{x},\omega),$$

where c_a is the complex wave speed, c the phase velocity, which can here depend on the frequency, and α is the imaginary part of the wavenumber. A rewriting of this equation directly gives

$$c_a(\mathbf{x},\omega) = \frac{\omega c(\mathbf{x},\omega)}{\omega + \mathrm{i}\alpha(\mathbf{x},\omega)c(\mathbf{x},\omega)}.$$

We also introduce the quality factor Q as the inverse of the dissipation. It is defined when $\alpha c \ll 2|\omega|$ (Ursin and Toverud, 2002) by

$$Q(\mathbf{x},\omega) = \frac{|\omega|}{2\alpha(\mathbf{x},\omega)c(\mathbf{x},\omega)}.$$

We represent the attenuation using the Kolsky-Futterman model (Kolsky, 1956; Futterman, 1962) which has been used in applications and defines the following relation (Ursin and Toverud, 2002, Eq. 17),

$$\begin{cases} \dfrac{1}{c(\mathbf{x},\omega)} = \dfrac{1}{c_{\omega_r}(\mathbf{x})} + \dfrac{\log\left|\dfrac{\omega_r}{\omega}\right|}{\pi c_{\omega_r}(\mathbf{x}) Q_{\omega_r}(\mathbf{x})}, \\ \alpha(\mathbf{x},\omega) = \dfrac{|\omega|}{2c_{\omega_r}(\mathbf{x}) Q_{\omega_r}(\mathbf{x})}, \end{cases}$$

where c_{ω_r} and Q_{ω_r} represent the values of the phase velocity c and the quality factor Q at

a reference angular frequency, ω_r.

Replacing the expression of c and α given by the Kolsky–Futterman model produces the following definition of the quality factor,

$$Q(\mathbf{x},\omega) = Q_{\omega_r}(\mathbf{x}) + \frac{1}{\pi}\log\left|\frac{\omega_r}{\omega}\right|.$$

It is straightforward to notice that if Q and c are both frequency independent, then $c = c_{\omega_r}$ and $Q_{\omega_r} = Q$. Following this simplification, the complex wave speed c_a is given by

$$c_a(\mathbf{x}) = \frac{\omega c(\mathbf{x})}{\omega + \dfrac{\mathrm{i}|\omega|}{2Q(\mathbf{x})}} = \frac{2\omega c(\mathbf{x})Q(\mathbf{x})}{2\omega Q(\mathbf{x}) + \mathrm{i}|\omega|} = \frac{4c(\mathbf{x})Q(\mathbf{x})^2 - 2\mathrm{i}\,\mathrm{sgn}(\omega)c(\mathbf{x})Q(\mathbf{x})}{4Q(\mathbf{x})^2 + 1},$$

where $\mathrm{sgn}(\omega)$ is the sign of ω. In the case of weak attenuation, $4Q^2 \gg 1$ the expression can be approximated with

$$c_a(\mathbf{x},\omega) = c(\mathbf{x}) - \frac{\mathrm{i}\,\mathrm{sgn}(\omega)c(\mathbf{x})}{2Q(\mathbf{x})} = c(\mathbf{x})\left(1 - \frac{\mathrm{i}\,\mathrm{sgn}(\omega)}{2Q(\mathbf{x})}\right). \tag{1.31}$$

Eventually, we observe that in this case, the complex wave speed is composed of a real part being the phase wave speed and an imaginary part depending on a quality factor Q:

$$\mathrm{Im}\left(c_a(\mathbf{x})\right) = \frac{\mathrm{sgn}(\omega)}{2}\frac{c(\mathbf{x})}{Q(\mathbf{x})}.$$

Clearly if the quality factor tends to be very high, there is no attenuation in the material and only the real wave speed remains. This notion of complex wave speed and attenuation factor depending on the frequency has motivated the use of the time-harmonic form of the wave equations, e.g. Muller (1983); Körnig and Müller (1989). The given formulation of the Kolsky–Futterman representation of visco-acoustic behavior was also used in the context of subsurface parameter reconstruction in Ribodetti and Virieux (1998); Malinowski et al. (2011).

The extension to elastic isotropic wave, Equation (1.29), naturally introduces two separate quality factor, say Q_p and Q_s which are related to the P- and S-waves respectively. As an alternative, Štekl and Pratt (1998) defined complex valued Lamé parameters λ and μ, yet one should carefully consider the way from one formulation to the other.

1.8 Some aspects regarding wave propagation PDEs

The amount of studies for the PDEs associated with the wave equations is monumental, let us point out some of the aspects that we have overlooked so far, in particular the boundary

conditions and their mathematical study.

1.8.1 Source and boundary conditions

1.8.1.1 Problem statement

In order to solve the wave equations we have defined, additional information must be incorporated: the source and the boundary conditions. By boundary we refer to the limit of the domain of interest (numerical or physical). These two facets are actually not always independent, namely when the source is located on the boundary. As for any PDE, they are required for the characterization of the solution and in order to prove well-posedness of the problem.

The source in particular, as the initial perturbation, is indispensable to even start the physical phenomenon of wave propagation. The location of this source, as well as its general characteristics (amplitude, phase), impact directly the solution. In seismic, the source can usually be represented by a Gaussian-shaped function, that we illustrate in Figure 1.5. If the source is inside the domain of interest, then it appears in the right-hand side of the wave equation following the decomposition of the forces of Subsection 1.3.3, it is represented by $\boldsymbol{f}$ in the Equation (1.4).

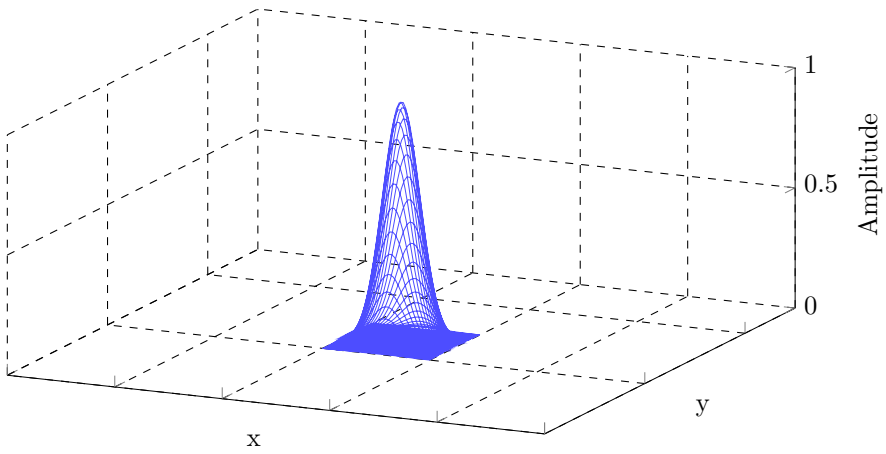


Figure 1.5: Illustration of the source amplitude as a Gaussian shape located at the center of a two-dimensional surface.

The boundary conditions give information on how the field behaves when reaching the limit of the domain of interest. In particular, what happens to the wave propagating in the Earth interior (acoustic or elastic) when it reaches the surface. Different types of boundary conditions exist: Dirichlet, Neuman, Robin, etc. In seismic we usually focus on the two first

only. The Dirichlet boundary condition imposes the value of the field and the Neumann boundary condition the value of the normal derivative. Let us consider a domain $\Omega \subset \mathbb{R}^n$ with boundary denoted by Γ, in dimension n (usually 2 or 3). The general formulation of the PDE becomes, by incorporating the different boundary conditions,

$$\begin{cases} \mathcal{P}(u(\mathbf{x},\omega)) = f(\mathbf{x}), & \mathbf{x} \in \Omega, \\ u(\mathbf{x},\omega) = g(\mathbf{x}), & \mathbf{x} \in \Gamma_1, \\ \partial_\nu u(\mathbf{x},\omega) = h(\mathbf{x}), & \mathbf{x} \in \Gamma_2, \end{cases}$$

where $\mathcal{P}$ is an operator, which can be assimilated to a wave operator of choice (acoustic, elastic, anisotropic), u is the wavefield, and ∂_ν represents the normal derivative. We have taken $\Gamma_1 \cap \Gamma_2 = \Gamma$.

1.8.1.2 Free-surface boundary conditions

The boundary conditions can reflect physical situation in addition to the source characteristic. In particular for a geophysical experiment where the wave propagates in the Earth's interior, the condition when the wave attains the surface (interface between the Earth and the air) is defined to be a *free surface*. It basically means that the surface is free to move. This refers to the interface between the air and solid material as well as between the liquid and the air. However, the condition naturally differs depending on the context.

Some intuitive characterization can be developed for the free surface conditions. In particular, the displacement should be free (the ground surface can moved) while there would not be any pressure. It means that for acoustic wave equations, involving the pressure p, the free surface can be characterized by imposing

$$p = 0, \qquad \text{acoustic free surface.}$$

The interface is a Dirichlet boundary condition. The free surface for elastic wave propagation (involving the displacement $\boldsymbol{u}$) is characterized by a Neumann boundary condition (the seismic acquisition follows a Neumann-to-Dirichlet map, see Baeten (1989)),

$$\partial_\nu \boldsymbol{u} = 0, \qquad \text{elasticstic free surface.}$$

We will later detail those conditions when introducing the seismic data used for the inverse problem, in Section 4.2.

Remark 1.10. *When we think about numerical simulation for Earth domain in geophysical exploration, it is not realistic (nor useful) to simulate over the whole Earth. Instead it is truncated, usually with a domain of interest representing a box of several kilometers cubed. If the Earth surface is still defined as a free surface, the lateral and bottom boundaries of the*

numerical domain (arising from the truncation) need special treatment. They must be dealt taking into account that the medium possibly continues on large distance and without any reflection. It usually results in the definition of absorbing boundary conditions or perfectly matched layers. We will naturally introduce those aspects during the numerical resolution of the wave equation, more precisely in Section 2.3.

1.8.2 Notes on the existence and uniqueness

Concerning the wave equations, the analytical study of the problem has been widely investigated and remains an active field of research, we mention for example Martin (2006) and the references therein. In particular, anisotropy and heterogeneity remain complicate to handle. For such study, the concept of *well-posedness* is crucial, and follows the notions introduced by Hadamard (1902, 1923). Let us consider a PDE, say $\mathcal{P}u = 0$ in Ω with appropriate boundary condition ($\mathcal{P}$ can be assimilated to a wave operator). The problem is said to be well-posed if

- the solution u exists (existence),
- the solution u is unique (uniqueness),
- the solution u depends continuously on the initial conditions.

In order to obtain such properties, $\mathcal{P}$ and the boundary conditions must be acutely defined. We give some simple considerations regarding homogeneous media in the following.

1.8.2.1 Sommerfeld radiation condition

For the Helmholtz equation (1.27), the Sommerfeld radiation condition assures that waves are not coming from infinity, it is named after Arnold Sommerfeld (1868–1951). Assuming the propagation in a sphere, it states that (Colton and Kress, 1998, Definittion 2.3),

$$\lim_{r\to+\infty} r\big(\frac{\partial p}{\partial r} - \mathrm{i}kp\big) = 0,$$

where k is the wavenumber and r the distance to the center of the sphere. It ensures uniqueness, see, for example, Colton and Kress (1998); Kirsch (1996), where existence is also proven.

1.8.2.2 Elasticity

For the study of elasticity , we refer to the pioneering work of Kupradze (1963); Kupradze et al. (1976), in particular for existence and uniqueness. Following the decomposition in P- and S-waves, see Remark 1.4, two conditions are imposed to assure that the waves do not propagate from infinity,

$$\begin{cases} \lim\limits_{r\to+\infty} r\big(\dfrac{\partial \boldsymbol{u}^p}{\partial r} - \mathrm{i}k\boldsymbol{u}^p\big) = 0, \\ \lim\limits_{r\to+\infty} r\big(\dfrac{\partial \boldsymbol{u}^s}{\partial r} - \mathrm{i}k\boldsymbol{u}^s\big) = 0, \end{cases}$$

These are often referred to as the Summerfeld–Kupradze conditions.

1.8.3 Analytical solution of the Helmholtz equation: Green's function

The Green's function are the solutions of the Helmholtz equations. Imposing the Sommerfeld radiation condition and an homogeneous velocity $c(\mathbf{x}) = c$, they are given by

$$\begin{cases} p^{(1D)}(\mathbf{x}) = \dfrac{e^{\mathrm{i}\omega c^{-1}\mathbf{x}}}{2\mathrm{i}\omega c^{-1}}, & \text{in one dimension,} \\ p^{(2D)}(\mathbf{x}) = \dfrac{i}{4}H_0(\omega c^{-1}\mathbf{x}), & \text{in two dimensions,} \\ p^{(3D)}(\mathbf{x}) = \dfrac{e^{\mathrm{i}\omega c^{-1}\mathbf{x}}}{4\pi\mathbf{x}}, & \text{in three dimensions,} \end{cases} \tag{1.32}$$

where H_0 is the Hankel function of the first kind. We illustrate the Green's function in two dimensions in Figure 1.6, on a 5 km square example with velocity 2000 $\mathrm{m\,s^{-1}}$ and frequency 4 Hz.

We observe the pattern of the wave propagating from the centrally located source. However, the Green's functions given in Equation (1.32) are restrictive because they assume an homogeneous medium. That is why we introduce the numerical discretization in Chapter 2, to accordingly account for Earth complexity.

1.9 Conclusion

In this chapter, we have defined the mathematical equations for the propagation of waves, starting from Newton's second law. The general formulation, with a full stiffness tensor, is further simplified depending on the type of medium. The distinction between acoustic

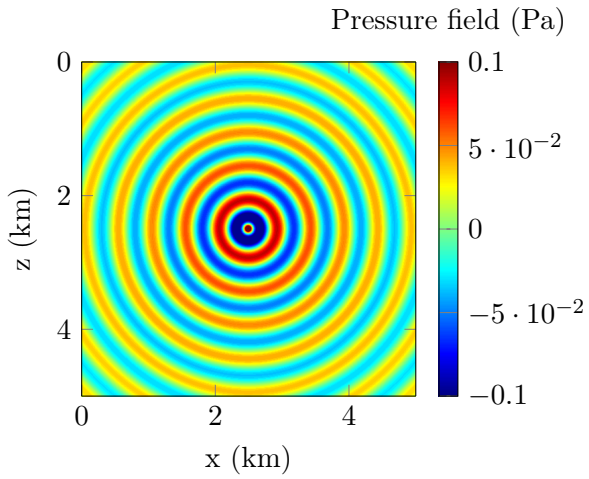


Figure 1.6: Illustration of the two-dimensional Green's function, see Equation (1.32). We have taken a 5 km square with constant velocity $c = 2000\ \mathrm{m\,s^{-1}}$ and frequency 4 Hz.

(fluid) and elastic (solid) medium imposes the modification of the quantity of interest, which is the scalar pressure field for acoustic media and the displacement vector field for elastic media. We have noted the underlying notion of Eulerian and Lagrangian approaches in this differentiation. The time-harmonic formulation of the wave equations substitutes the time variable with the angular frequency. It is the formulation elected for the rest of the project. We have addressed the case of attenuation in the medium with one of the existing models of representation, to define a complex valued wave speed. Following the introduction of boundary conditions, we have proposed initial material regarding the problem, regarding the existence and uniqueness of the solution. Homogeneous media provide advantageous behavior and allow the definition of analytical solutions, via the Green's functions. However, heterogeneous media are revealed to be much more complex to deal with. That is why we now study the standard numerical tools employed for the resolution of wave propagation in heterogeneous media.

Chapter 2

Numerical modeling of waves propagation

Abstract

We introduce the methods employed to solve the wave equations numerically. It consists in the discretization of the domain of interest and consequently the wave equations. In this project, we have used Finite Differences (FD) and Galerkin methods (continuous and discontinuous). The discretization of the harmonic equations generates a linear system, which is large depending on the domain of interest. We employ direct solvers, motivated by the multi right-hand sides possibilities and the re-use of the factorization. We carry out numerical experiments to illustrate the modeling of waves, in two and three dimensions and considering a TTI medium.

Contents

In Chapter 1, we have established the equations for the propagation of waves in the frequency domain. The Equations (1.27), (1.28), (1.29) and (1.30) summarize the possibilities, depending on acoustic, elastic, and anisotropic media. Analytical solutions exist, such as the Green's functions that we have illustrated in Section 1.8, but they require a homogeneous medium, which is not acceptable to model the Earth. The multi-scattering of obstacles can also be resolved analytically from boundary element methods, and we refer to the work of Martin (2006); Barucq et al. (2016) and the references therein. In particular, Barucq et al. (2016) provides a comprehensive numerical implementation of the method and compare the efficiency depending on a finite element method and the solver (direct or iterative) used.

In order to find the solution of a PDE with heterogeneous coefficients, it is common to rely on a discretization of the equation. First, the domain of interest is decomposed into a discrete set of partitions. Then, one can select the method to discretize the equation which is accordingly rewritten. Different methods exist, and we simply present the ones that have been used in this project. These are the Finite Differences (FD), the Finite Elements (FE, or Continuous Galerkin, CG), the Discontinuous Galerkin (DG) and the Hybridizable Discontinuous Galerkin (HDG). The respective formulations are prescribed in Sections 2.1 and 2.2.

The FD approach presents the advantage of being relatively simple to implement but is usually restricted to the consideration of a structured partition of the domain. It is uneasy to handle the topography of subsurface structures of complex geometries, or it may require a very small step size for the discretization, which induces the increase in the computational time. The alternative of Galerkin method (named after Boris Grigoryevich Galerkin, 1871–1945) is interesting because it allows an unstructured mesh which can easily account for the geometry. The high flexibility is clearly an advantage for the method, but they are certainly less straightforward than FD. Historically the finite differences method was initially promoted in the article of Richard Courant, Kurt Friedrichs and Hans Lewy (1928), for the time domain wave equation. Finite elements were developed a few decades later; we refer to Thomée (2001) for a comprehensive review of the historical steps from FD to FE. DG appears from the work of Reed and Hill (1973) considering hyperbolic equations; we refer to Arnold et al. (2002) for a historical review. Those methods are now widely employed to provide numerical solution of PDEs for time and frequency domains formulations; for example, we mention Monk (2003) (for the Maxwell's equations), Ihlenburg (2006); Cohen (2003); Ainsworth et al. (2006); Hesthaven and Warburton (2007); Riviere (2008); Brossier et al. (2010a); Allaire (2012); Cohen and Pernet (2017), among (many) others.

We discuss the boundary conditions to apply for the numerical discretization. In particular, in the case of seismic, the domain of interest is a section of Earth, and the waves

should be allowed to pursue the propagation outside the considered domain. Hence, artificial boundaries are required to ensure that no waves are reflected from the numerical boundaries, we review two common methods for this purpose: the Perfectly Matched Layers (PML) and Absorbing Boundary Conditions (ABC), see Section 2.3.

The harmonic approach of the problem, by suppressing the time variable, naturally avoid the discretization in time and the underlying choice of implicit or explicit scheme. Instead, it eventually leads to a (possibly large) linear system, which requires efficient techniques for the solver, due to computational limitation in memory. In this project, we have employed direct solvers only, motivated by the possibility to handle multiple right-hand sides. It is a crucial aspect in seismic where a campaign usually consists of several hundreds of shots. It also allows the re-use of the factors obtained from the factorization, which permits faster computation of the gradient for the inverse problem we depict in the sequel.

The resolution of the forward problem is actually not the purpose of this project, which concentrates on the seismic inverse problem instead. However, the inverse problem relies on the resolution of the forward problem and it is natural to introduce the main formulations for a self-sustained manuscript. Hence we will try to remain concise on the subject, pointing out that the discretized methods implemented in the software are due to the work of Wang et al. (2011, 2012) for FD, Chaumont-Frelet (2015) for FE and DG, and Bonnasse-Gahot (2015) for HDG. Numerical experiments are presented to illustrate the propagation of harmonic waves in Section 2.5, for two and three-dimensional domains, including anisotropy.

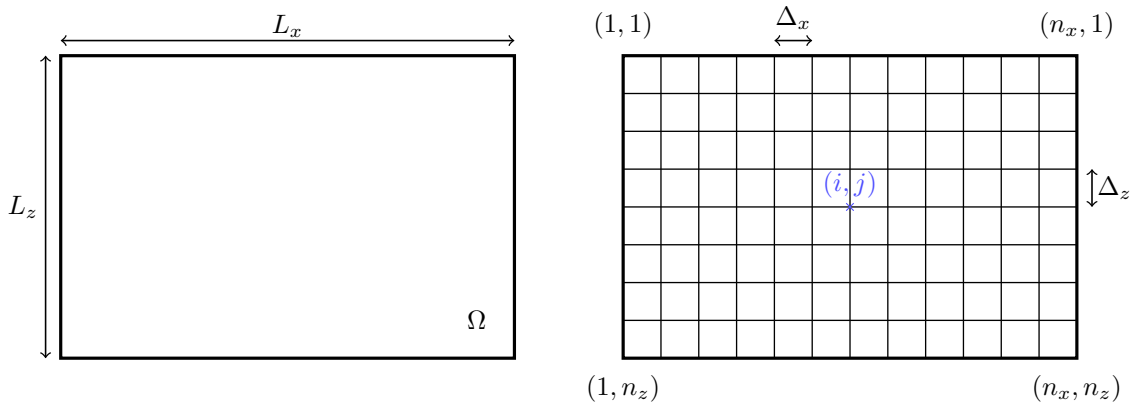
2.1 Introduction to the Finite Differences method

The Finite Differences (FD) method standardly relies on a structured decomposition of the domain of interest and a Taylor approximation of the derivative. We will review the method and illustrate with a simple configuration: a two-dimensional domain $\Omega \subset \mathbb{R}^2$ of rectangular shape. The method is divided in two steps, first the domain (structured) discretization and then the discretization of the equation, with the derivative operator.

2.1.1 Spatial discretization

We assume a two-dimensional domain of interest, Ω to have length L_x (x-axis) and depth L_z (or height, z-axis). The finite differences method commonly takes a structured decomposi-

tion of the domain, which is partitioned in a regular grid of nodes. This grid is determined by the step of discretization, denoted Δ_x for the x-axis and Δ_z for the z-axis. This process naturally creates $n_x = (L_x/d_x + 1)$ partitions in length and $n_z = (L_z/d_z + 1)$ partitions in depth. The total number of grid points follows to be $n = n_x \times n_z$. We illustrate the domain discretization in Figure (2.1). The nodal points are labeled as (i, j) where i is the z index, from 1 to n_z and j stands for the x index, from 1 to n_x.



(a) Two-dimensional domain Ω of size $L_x \times L_z$.

(b) Structured discretization of the two-dimensional domain Ω: the horizontal axis is partitioned into n_x nodes and the vertical axis into n_z nodes. The discretization generates a total of $n_x \times n_z$ nodes, equidistant per direction.

Figure 2.1: Illustration of a structured discretization applied on a two-dimensional rectangular domain for the use of the FD method.

Following the spatial discretization, the FD method seeks for the discretized solution of the PDE defined at every node.

2.1.2 Operators discretization

Let us consider the PDE to be the acoustic harmonic wave equation, (see Equation (1.28)),

$$-\omega^2 \kappa^{-1}(\mathbf{x}) p(\mathbf{x}) - \nabla \cdot \left(\rho(\mathbf{x})^{-1} \nabla p(\mathbf{x})\right) = 0, \qquad \text{in } \Omega, \tag{2.1}$$

where we postpone the consideration of boundary conditions. The FD method aims the recovery of a discretized solution, at the nodal points only. Here it is materialized by the vector P defined by

$$\begin{aligned} P &= \{p(\mathbf{x}_{1,1}); p(\mathbf{x}_{1,2}); \ldots; p(\mathbf{x}_{i,j}); \ldots; p(\mathbf{x}_{n_z,n_x})\} \\ &= \{P_{1,1}; P_{1,2}; \ldots; P_{i,j}; \ldots; P_{n_z,n_x}\}, \end{aligned}$$

where $\mathbf{x}_{(i,j)}$ indicates the position at the node (i,j), accordingly with the nodal discretization given in Figure 2.1. We introduce $P_{i,j}$ for the value of the vector P corresponding with $p(\mathbf{x}_{i,j})$. Similarly the model parameters must be discretized, and we refer to $\kappa_{i,j}$ and $\rho_{i,j}$ for the bulk modulus and density respectively.

The discretization of the derivative is based upon the Taylor formula, which indicate at first order that for an integrable function f,

$$f(x+h) = f(x) + h\frac{\partial f(x)}{\partial x} + o(h^2) \qquad \lim_{h\to 0} o(h^2) = 0.$$

Neglecting the term in $o(h^2)$ gives an approximation of the derivative,

$$\frac{\partial f(x)}{\partial x} \simeq \frac{f(x+h) - f(x)}{h}.$$

This formulation can be substituted if we consider $f(x-h)$ instead of $f(x+h)$, giving the approximation

$$\frac{\partial f(x)}{\partial x} \simeq \frac{f(x) - f(x-h)}{h},$$

we can eventually sum the two to obtain

$$\frac{\partial f(x)}{\partial x} \simeq \frac{f(x+h) - f(x-h)}{2h}.$$

In order to discretize the derivative operator in the FD framework, we apply this approximation onto the structured grid. For instance the gradient of the pressure field at the nodal position $\mathbf{x}_{i,j}$ is approximated by

$$\nabla p(\mathbf{x}_{i,j}) = \begin{pmatrix} \dfrac{\partial p(\mathbf{x}_{i,j})}{\partial x} \\ \dfrac{\partial p(\mathbf{x}_{i,j})}{\partial z} \end{pmatrix} \simeq \begin{pmatrix} \dfrac{P_{i,j+1} - P_{i,j-1}}{2\Delta_x} \\ \dfrac{P_{i+1,j} - P_{i-1,j}}{2\Delta_z} \end{pmatrix}.$$

For second order derivatives, the Taylor's theorem gives for $\pm h$,

$$f(x+h) = f(x) + h\frac{\partial f(x)}{\partial x} + \frac{h^2\partial^2 f(x)}{2\partial x^2} + o(h^3) \qquad \lim_{h^2\to 0} o(h^3) = 0,$$
$$f(x-h) = f(x) - h\frac{\partial f(x)}{\partial x} + \frac{h^2\partial^2 f(x)}{2\partial x^2} + o(h^3) \qquad \lim_{h^2\to 0} o(h^3) = 0.$$

The sum of those two expressions, neglecting the term $o(h^3)$, gives the approximation of the second order derivative:

$$\frac{\partial^2 f(x)}{\partial x^2} \simeq \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}.$$

If we apply onto the discretized pressure field, we have

$$\begin{cases} \dfrac{\partial p}{\partial x^2}(\mathbf{x}_{i,j}) \simeq \dfrac{P_{i,j+1} - 2P_{i,j} + P_{i,j-1}}{\Delta_x^2}, \\ \dfrac{\partial p}{\partial z^2}(\mathbf{x}_{i,j}) \simeq \dfrac{P_{i+1,j} - 2P_{i,j} + P_{i-1,j}}{\Delta_z^2}. \end{cases}$$

In particular, the Laplacian is obtained by summing the two second order derivatives.

The discretization of the wave operator follows from the injection of the Taylor approximations for the derivatives at every nodal points. Let us illustrate with Equation (2.1), in the position $\mathbf{x}_{i,j}$ we have

$$\begin{aligned} &- \omega^2 \kappa^{-1}(\mathbf{x}_{i,j}) p(\mathbf{x}_{i,j}) - \nabla \cdot \big(\rho^{-1}(\mathbf{x}_{i,j}) \nabla p(\mathbf{x}_{i,j})\big) \\ &= - \omega^2 \kappa^{-1}(\mathbf{x}_{i,j}) p(\mathbf{x}_{i,j}) - \bigg(\frac{\partial \rho^{-1}}{\partial x}(\mathbf{x}_{i,j}) \frac{\partial p}{\partial x}(\mathbf{x}_{i,j}) + \frac{\partial \rho^{-1}}{\partial z}(\mathbf{x}_{i,j}) \frac{\partial p}{\partial z}(\mathbf{x}_{i,j}) \\ &\qquad\qquad + \rho^{-1}(\mathbf{x}_{i,j}) (\Delta p)(\mathbf{x}_{i,j}) \bigg) \\ &\simeq - \omega^2 \kappa_{i,j}^{-1} P_{i,j} - \bigg(\frac{\big(\rho_{i,j+1}^{-1} - \rho_{i,j-1}^{-1}\big)\big(P_{i,j+1} - P_{i,j-1}\big)}{4\Delta_x^2} + \frac{\big(\rho_{i+1,j}^{-1} - \rho_{i-1,j}^{-1}\big)\big(P_{i+1,j} - P_{i-1,j}\big)}{4\Delta_z^2} \\ &\qquad\qquad + \rho_{i,j}^{-1} \bigg(\frac{P_{i,j+1} - 2P_{i,j} + P_{i,j-1}}{\Delta_x^2} + \frac{P_{i+1,j} - 2P_{i,j} + P_{i-1,j}}{\Delta_z^2} \bigg) \bigg). \end{aligned}$$

Then we organize to factor out the nodal pressure such that,

$$\begin{aligned} &- \omega^2 \kappa^{-1}(\mathbf{x}_{i,j}) p(\mathbf{x}_{i,j}) - \nabla \cdot \big(\rho^{-1}(\mathbf{x}_{i,j}) \nabla p(\mathbf{x}_{i,j})\big) \\ \simeq\ & P_{i,j} \bigg(- \omega^2 \kappa_{i,j}^{-1} + \rho_{i,j}^{-1} \bigg(\frac{2}{\Delta_x^2} + \frac{2}{\Delta_z^2} \bigg) \bigg) \\ &- P_{i,j+1} \bigg(\frac{\big(\rho_{i,j+1}^{-1} - \rho_{i,j-1}^{-1}\big)}{4\Delta_x^2} + \frac{\rho_{i,j}^{-1}}{\Delta_x^2} \bigg) + P_{i,j-1} \bigg(\frac{\big(\rho_{i,j+1}^{-1} - \rho_{i,j-1}^{-1}\big)}{4\Delta_x^2} - \frac{\rho_{i,j}^{-1}}{\Delta_x^2} \bigg) \\ &- P_{i+1,j} \bigg(\frac{\big(\rho_{i+1,j}^{-1} - \rho_{i-1,j}^{-1}\big)}{4\Delta_z^2} + \frac{\rho_{i,j}^{-1}}{\Delta_z^2} \bigg) + P_{i-1,j} \bigg(\frac{\big(\rho_{i+1,j}^{-1} - \rho_{i-1,j}^{-1}\big)}{4\Delta_z^2} - \frac{\rho_{i,j}^{-1}}{\Delta_z^2} \bigg). \end{aligned} \tag{2.2}$$

The discretization generates one equation per nodal points and each of them involves five grid points: the node itself and its four neighbors (at the exception of boundary nodes where the neighboring node outside the domain can be substituted with zeros). It produces a linear system of the form $AP = 0$ where the matrix A is of size $n \times n$ and encompasses the values resulting from the Taylor approximation. In this case only five coefficients are

non-zeros per line of the matrix, according to the formulation (2.2).

2.1.3 Extension

In our approach, we have simply used the neighboring nodes but it is possible to extend the Taylor formula and increase the number of nodes involved. This leads to the higher order FD, and other extensions are possible

Staggered grid. In the standard formulation of the FD, only the nodal values of the field have been considered, namely $P_{i,j}$. Alternatively the consideration of intermediate positions has been introduced by Madariaga (1976) for time domain wave problem, in particular for the time variable discretization. Instead of limiting the discretization at nodal values (i, j), we consider half position such that $(i + 1/2, j)$. This is referred to as staggered grid. This formulation has been successfully employed by Virieux (1984, 1986) for the time domain wave equation.

Rotated grid. Another alternative is to consider a rotation of the standard vertical and horizontal axes for the discretization of the derivatives. Instead of considering the vertical and horizontal neighbors only, the method considers rotated axes by $\pi/4$, as illustrated in Figure 2.2. It basically proceeds to a change of coordinates.

(a) Classical formulation. (b) Rotated formulation

Figure 2.2: The standard FD formulation (left) utilizes the neighbor nodes following orthonormal axis horizontal and vertical. In this case a derivative discretization for a node of interest (in blue) will need information from the nodes in red. The axes are rotated by $\pi/4$ on the right to generate a formulation where the information is needed on the 'diagonal' neighbors.

The use of rotated axis for FD scheme has been applied for the frequency domain wave equation in Jo et al. (1996) in acoustic and Štekl and Pratt (1998) for the elastic wave

equation. The major difference resides in the step size for the discretized derivative, the distance is now the same for the two axes. Introducing $\Delta_{xz} = \sqrt{\Delta_x^2 + \Delta_z^2}$, the derivative is approximated with

$$\nabla p(\mathbf{x}_{i,j}) = \begin{pmatrix} \dfrac{P_{i+1,j-1} - P_{i-1,j+1}}{2\Delta_{xz}} \\ \\ \dfrac{P_{i+1,j+1} - P_{i-1,j-1}}{2\Delta_{xz}} \end{pmatrix}.$$

Assuming the grid to be equally spaced in the two axis, $\Delta_x = \Delta_z$, then the new step can be replaced by $\Delta_{xz} = \sqrt{2}\Delta_x = \sqrt{2}\Delta_z$.

Mixed grid. The centered and rotated schemes can be considered jointly. It defines a mixed scheme where the eight adjacent nodal points are used to defined the discretization, i.e all the red points in Figure 2.2. This approach was used in Jo et al. (1996). This idea of considering all the closest points is also employed in Marfurt (1984) when treating the elastic wave equation, where he also addresses the FE method. Eventually, the scheme is developed as a weighted sum of the two formulations (rotated and staggered). Referring by $\alpha \in [0, 1]$ the weighting coefficient, the derivative is expressed as

$$\nabla p(\mathbf{x}_{i,j}) = \alpha \begin{pmatrix} \dfrac{P_{i+1,j} - P_{i-1,j}}{2\Delta_x} \\ \\ \dfrac{P_{i,j+1} - P_{i,j-1}}{2\Delta_z} \end{pmatrix} + (1-\alpha) \begin{pmatrix} \dfrac{P_{i+1,j-1} - P_{i-1,j+1}}{2\Delta_{xz}} \\ \\ \dfrac{P_{i+1,j+1} - P_{i-1,j-1}}{2\Delta_{xz}} \end{pmatrix}.$$

This approach is used in Hustedt et al. (2004) in the frequency domain with extension to the so-called parsimonious grid. Operto et al. (2007) develops the method for three-dimensional cases. Wave equations for TTI media are treated in Operto et al. (2009). This method has been proven to be of order fourth in Hustedt et al. (2004).

We see that the FD method has many variations but all of them lead to the formation of the linear system which size is the number of nodes in the representation. The major drawback is probably the difficulty of the FD method to consider topography or complex geometry in the subsurface. It inherits from its structured domain decomposition which the Galerkin approach is free of.

2.2 Introduction to Galerkin methods

The FD method seeks a discretized representation of the solution at nodal positions. The numerical discretization based upon Galerkin methods seeks the solution over the whole domain, by assuming it is (usually) represented with piecewise polynomial functions and considering the variational formulation.

We illustrate the application of the Galerkin method in a domain Ω with boundary $\Gamma = \Gamma_1 \cap \Gamma_2$ and the Helmholtz equation,

$$\begin{cases} \left(-\Delta - \dfrac{\omega^2}{c^2(\mathbf{x})} \right) p(\mathbf{x}) = f(\mathbf{x}) & \text{in } \Omega, \\ p(\mathbf{x}) = 0 & \text{on } \Gamma_1, \\ \partial_\nu p(\mathbf{x}) = 0 & \text{on } \Gamma_2. \end{cases} \tag{2.3}$$

We have considered a Dirichlet boundary condition on Γ_1 and a Neumann boundary condition for Γ_2, where ν is the normal direction. We further assume that the source term, f is in $L^2(\Omega)$.

We take the velocity c to be a piecewise continuous function defined on Ω. Obviously, c never vanishes and is bounded in practice, because it is the physical wave speed. By using the Fredholm theory it is possible to prove that the Problem (2.3) is well-posed in the following sense:

$$\forall f \in L^2(\Omega), \text{ the problem admits a unique solution } u \in \{w \in H^1(\Omega) \ ; \ \Delta w \in L^2(\Omega)\},$$

where H^1 is a Hilbert space and L^2 the space of square integrable functions.

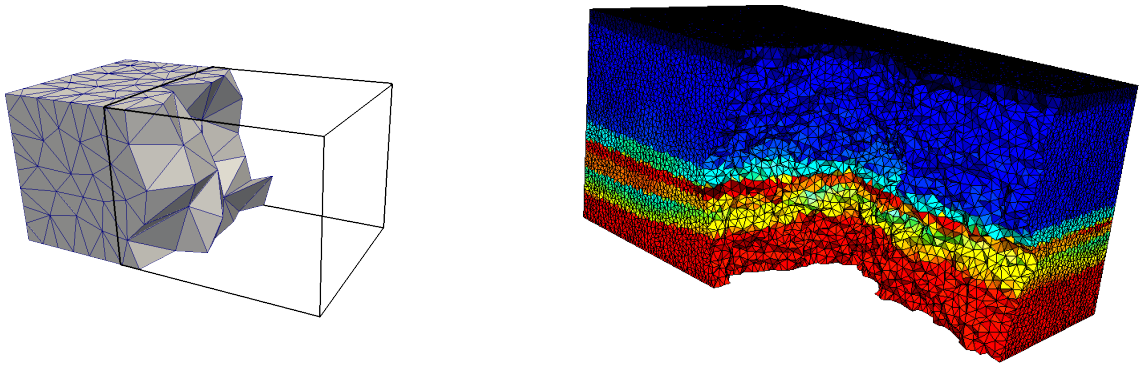
2.2.1 Domain discretization, reference element

The discretization of the domain for Galerkin methods is standardly realized with triangles in two dimensions and tetrahedra in three dimensions. This is the decomposition selected in this project, yet one can also employ other kind of elements, such as quadrangles for example. The domain of interest Ω is partitioned in N_K non-overlapping cells K_j such that

$$\Omega = \left\{ \cup_{k=1}^{N_K} K_k, \quad K_i \cap K_j = \emptyset \ \ \forall i \neq j \right\},$$

In Figure 2.3(a) we illustrate the decomposition of a simple domain in 3D, with few tetrahedra. In Figure 2.3(b) we show the discretization of a wave speed into a piecewise constant representation defined over the cells of a mesh. In our case, we mainly work

with paved domain but the implementation is totally suitable for other geometries, such as geophysical topography or even balls (via curved element for example).



(a) Illustration of tetrahedral mesh, the tetrahedral cells are indicated in blue, the black lines indicate the rest of the domain (transparent).

(b) Representation of a velocity onto a three-dimensional mesh. A quarter of the domain has been ignored to allow the visualization of inner cells. The blue area is water and several structures are present in the subsurface, see Chapter 7 for more details on the model.

Figure 2.3: Illustration of three-dimensional tetrahedral meshes for rectangle domains and parameter decomposition into a piecewise constant.

We introduce the reference element $\widehat{K}$ and the linear transformation T_e from $\widehat{K}$ to any cell K_e, as illustrated Figure 2.4. It allows any cell of the mesh (assuming triangles or tetrahedra) to be mapped to this reference element.

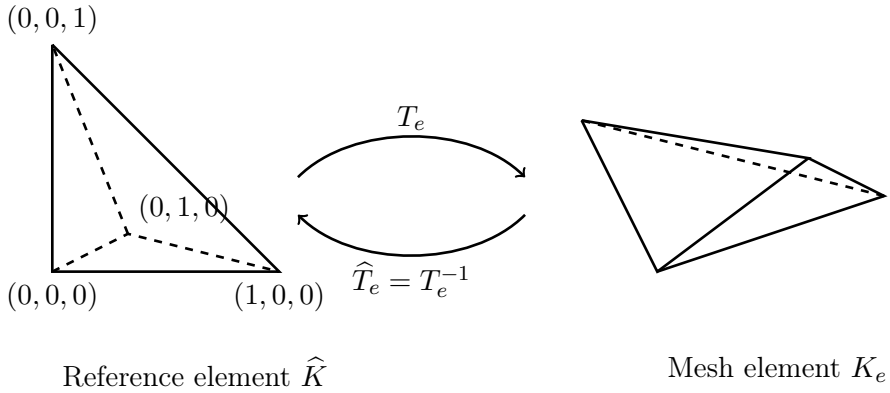
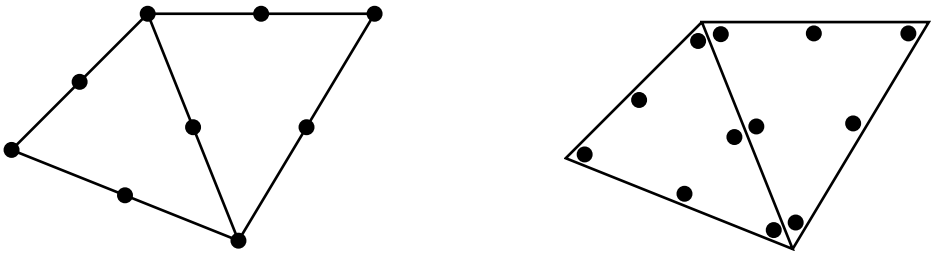


Figure 2.4: Illustration of the reference element. The tetrahedron K_e is mapped with the reference element $\widehat{K}$ with the transformation T_e.

The difference between continuous and discontinuous Galerkin methods lies in particular in the representation of the space where the approximate solution is defined. As the name indicates, the DG method considers discontinuous shape functions and they are continuous for CG/FE. In Figure 2.5, we illustrate the degrees of freedom of order two for two triangle cells.



(a) Continuous Galerkin (b) Discontinuous Galerkin

Figure 2.5: Illustration of the degrees of freedom between CG/FE and DG for a two-dimensional domain decomposed with triangles and order 2 polynomial. Shared degrees of freedom are present in the CG method, on the contrary every degree of freedom belongs to a single element in DG.

2.2.2 Continuous Galerkin discretization

The Continuous Galerkin (CG) or Finite Elements (FE) method consider the approximate solution of the PDE to be a continuous piecewise polynomial functions. We recast the problem into a variational formulation. For this purpose we introduce the function space V such that

$$V = \big\{ v \in H^1(\Omega) \quad \text{such that} \quad v\,|_{\Gamma_1} = 0 \big\}.$$

Then if p is solution of the Problem (2.3) for a given $f \in L^2(\Omega)$ we have that $\forall v \in V$,

$$\int_\Omega \Big(-\omega^2 c(\mathbf{x})^{-2} p(\mathbf{x}) \overline{v}(\mathbf{x}) - \Delta p(\mathbf{x}) \overline{v}(\mathbf{x}) \Big)\, \mathrm{d}\mathbf{x} = \int_\Omega f(\mathbf{x}) \overline{v}(\mathbf{x})\, \mathrm{d}\mathbf{x}, \tag{2.4}$$

where $\overline{v}$ stands for the conjugate of the test function v.

Let us apply integration by parts to the variational formulation (2.4), we obtain

$$\int_\Omega \Big(-\omega^2 c(\mathbf{x})^{-2} p(\mathbf{x}) \overline{v}(\mathbf{x}) + \nabla p(\mathbf{x}) \nabla \overline{v}(\mathbf{x}) \Big)\, \mathrm{d}\mathbf{x} - \big\langle \partial_\nu p(\mathbf{x}), v \big\rangle_\Gamma = \int_\Omega f(\mathbf{x}) \overline{v}(\mathbf{x})\, \mathrm{d}\mathbf{x}.$$

Here $<\ ,\ >_\Gamma$ stands for the dual pairing $H^{-\frac{1}{2}}(\Gamma)$, $H^{\frac{1}{2}}(\Gamma)$. We remind the Neumann condition on Γ_2 where $\partial_\nu p = 0$ and note that $v = 0$ on Γ_1. The variational formulation can be written accordingly,

$$\int_\Omega \Big(-\frac{\omega^2}{c(\mathbf{x})^2} u(\mathbf{x}) \overline{v}(\mathbf{x}) + \nabla u(\mathbf{x}) \nabla \overline{v}(\mathbf{x}) \Big)\, \mathrm{d}\mathbf{x} = \int_\Omega f(\mathbf{x}) \overline{v}(\mathbf{x})\, \mathrm{d}\mathbf{x}. \tag{2.5}$$

We now introduce the space of discretization V_h, defined with piecewise polynomial functions of order less than or equal to r over the tetrahedra,

$$V_h = \big\{ w \in V \quad \text{such that} \quad \forall i, \quad w\,|_{K_i} \in \mathbb{P}_r(K_i) \big\}.$$

Here $\mathbb{P}_r$ denotes the space of polynomials of order less than or equal to r given in three

dimensions by

$$\mathbb{P}_r = \Big\{ h(x,y,z) = \sum_{i,j,k=0}^{r} h_{ijk} x^i y^j z^k \Big\}. \tag{2.6}$$

Hence p is approximated by a sum of piecewise polynomial functions. We write,

$$p(\mathbf{x}) \mid_{K_e} = \sum_k p_k \phi_k(\mathbf{x}),$$

where the $\phi_k \in \mathbb{P}_r$ are the basis functions. In our case we choose the Lagrangian basis functions. We define P_h the vector encompassing the functions coefficients p_j with

$$p(\mathbf{x}) = \sum_j p_j \phi_j(\mathbf{x}).$$

We further choose $v = \{\phi_i\}$. Then the Equation (2.5) can be written as a linear system

$$AP_h = F,$$

with $A = M + L$, where M is the mass matrix and L the stiffness matrix. The different coefficients are given by

$$\begin{cases} [M]_{i,j} = \displaystyle\int_\Omega -\omega^2 c(\mathbf{x})^{-2} \phi_i(\mathbf{x}) \phi_j(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \\ [L]_{i,j} = \displaystyle\int_\Omega \nabla\phi_i(\mathbf{x}) \nabla\phi_j(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \\ [F]_i = \displaystyle\int_\Omega f(\mathbf{x}) \phi_i(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \end{cases}$$

where we have assumed that $\phi \in \mathbb{R}$ to neglect the conjugation. Because the basis functions are defined on the tetrahedra, the integrals can further be specified by the cells contained in the support of ϕ, for instance

$$[L]_{i,j} = \sum_{e \in (\mathrm{supp}(\phi_i) \cap \mathrm{supp}(\phi_j))} \int_{K_e} \nabla\phi_i(\mathbf{x}) \nabla\phi_j(\mathbf{x}) \, \mathrm{d}\mathbf{x},$$

and similarly for $[M]_{i,j}$ and $[F]_i$. Let us now reformulate p with the reference element basis functions $\widehat{\phi}$,

$$p(\mathbf{x}) \mid_{K_e} = \sum_k p_k \big(\widehat{\phi_k} \circ \widehat{T}_e(\widehat{\mathbf{x}}) \big),$$

where $\widehat{\mathbf{x}}$ denotes the coordinates in the reference element. The matrix coefficients can be

rewritten accordingly using DT to denote the Jacobian of T:

$$\begin{cases} [M]_{i,j} = \displaystyle\sum_{e\in(\mathrm{supp}(\phi_i)\cap\mathrm{supp}(\phi_j))} |\det(DT_e)| \int_{\widehat{K}_e} -\omega^2 c^{-2}(\mathbf{x})\widehat{\phi}_i(\widehat{\mathbf{x}})\widehat{\phi}_j(\widehat{\mathbf{x}})\,\mathrm{d}\widehat{\mathbf{x}}, \\ [L]_{i,j} = \displaystyle\sum_{e\in(\mathrm{supp}(\phi_i)\cap\mathrm{supp}(\phi_j))} |\det(DT_e)| \int_{\widehat{K}_e} (DT_e^{-1})(D\widehat{T}_e^{-T})\nabla\widehat{\phi}_i(\widehat{\mathbf{x}})\nabla\widehat{\phi}_j(\widehat{\mathbf{x}})\,\mathrm{d}\widehat{\mathbf{x}}, \\ [F]_i = \displaystyle\sum_{e\in\mathrm{supp}(\phi_i)} |\det(DT_e)| \int_{\widehat{K}_e} f(\mathbf{x})\widehat{\phi}_i(\widehat{\mathbf{x}})\,\mathrm{d}\widehat{\mathbf{x}}. \end{cases}$$

Eventually, we obtain a linear system, where the integrals are computed using a Gauss quadrature.

2.2.3 Discontinuous Galerkin discretization

The DG method, by allowing a discontinuous function between element, requires a different variational formulation, where the jumps of the function are incorporated. For this purpose we introduce inward and outward normals, as illustrated in Figure 2.6 with two adjacent cells, say K^+ and K^-.

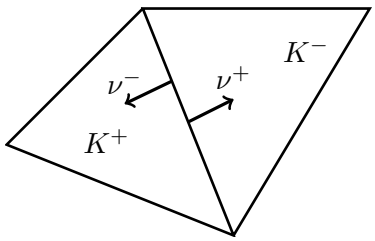


Figure 2.6: Illustration of the notation for DG method at the interface between two cells.

We define the *jump* of a scalar w by

$$[w] = w^{K^+}\cdot\nu^+ + w^{K^-}\cdot\nu^- = w^{K^+}\cdot\nu^+ - w^{K^-}\cdot\nu^+,$$

where w^{K^+} and w^{K^-} are the quantity on the cell K^+ and K^- respectively. For a vector $\boldsymbol{u}$ we define

$$[\boldsymbol{u}] = \boldsymbol{u}^{K^+}\cdot\nu^+ + \boldsymbol{u}^{K^-}\cdot\nu^- = \boldsymbol{u}^{K^+}\cdot\nu^+ - \boldsymbol{u}^{K^-}\cdot\nu^+.$$

We now introduce the *average*,

$$\{w\} = \frac{1}{2}\big(w^{K^+}|_{\Upsilon_1^{(e)}} + w^{K^-}|_{\Upsilon_1^{(e)}}\big),$$

where $\Upsilon_1^{(e)}$ is the shared boundary. We note that

$$[\boldsymbol{w_1}]\{w_2\} + [w_2]\{\boldsymbol{w_1}\} = [\boldsymbol{w_1} w_2] \tag{2.7}$$

Assuming p is solution of the problem (2.3), we impose $[p] = 0$ to ensure the continuity, and $[\nabla p] = 0$. We now introduce the notation $\Upsilon^{(e)} = \{\Upsilon_i^{(e)}, \quad i = 1, \ldots, d+1\}$ which corresponds to the boundary of the inner element K_e and where d is the dimension. For example in two dimensions we have three edges per triangle, see Figure 2.6.

For the variational formulation, we now let the function space V to be

$$V = \big\{ v \in L^2(\Omega) \quad \text{such that} \quad v\,|_{K_e} \in H^1(K_e),\ \forall e; \quad v\,|_{\partial K_e \cap \Gamma_1} = 0 \ \text{ if } \ \partial K_e \cap \Gamma_1 \neq \emptyset \big\}.$$

The variational formulation is given $\forall v \in V$,

$$\int_\Omega \Big(-\omega^2 c(\mathbf{x})^{-2} p(\mathbf{x}) \overline{v}(\mathbf{x}) - \Delta p(\mathbf{x}) \overline{v}(\mathbf{x}) \Big)\, \mathrm{d}\mathbf{x} = \int_\Omega f(\mathbf{x}) \overline{v}(\mathbf{x})\, \mathrm{d}\mathbf{x},$$

similarly to the previous situation but where v is allowed to have discontinuities. The integration over a cell K_e is given by

$$\begin{aligned}
&\int_{K_e} \Big(-\omega^2 c(\mathbf{x})^{-2} p(\mathbf{x}) \overline{v}(\mathbf{x}) - \Delta p(\mathbf{x}) \overline{v}(\mathbf{x}) \Big)\, \mathrm{d}\mathbf{x} \\
&= \int_{K_e} \Big(-\omega^2 c(\mathbf{x})^{-2} p(\mathbf{x}) \overline{v}(\mathbf{x}) + \nabla p(\mathbf{x}) \nabla \overline{v}(\mathbf{x}) \Big)\, \mathrm{d}\mathbf{x} - \int_{\Upsilon^{(e)}} \partial_{\nu_e^+} p(\mathbf{x}) \overline{v}(\mathbf{x})\, \mathrm{d}\mathbf{x},
\end{aligned}$$

where the test function v is possibly not continuous.

In order to sum over the cells, we introduce the set of cell boundaries $\Upsilon = \{\cup_{e=1}^{N_K} \Upsilon^{(e)}\}$, where only the boundaries shared between two cells are considered. From the relation between the outward and inward normal, ν^+ and ν^-, we obtain

$$\begin{aligned}
&\sum_{K_e} \int_{K_e} \Big(-\omega^2 c(\mathbf{x})^{-2} p(\mathbf{x}) \overline{v}(\mathbf{x}) - \Delta p(\mathbf{x}) \overline{v}(\mathbf{x}) \Big)\, \mathrm{d}\mathbf{x} \\
&= \sum_{K_e} \int_{K_e} \Big(-\omega^2 c(\mathbf{x})^{-2} p(\mathbf{x}) \overline{v}(\mathbf{x}) + \nabla p(\mathbf{x}) \nabla \overline{v}(\mathbf{x}) \Big)\, \mathrm{d}\mathbf{x} \\
&\qquad - \int_\Upsilon \partial_{\nu^+} p(\mathbf{x}) \overline{v}(\mathbf{x}) + \partial_{\nu^-} p(\mathbf{x}) \overline{v}(\mathbf{x})\, \mathrm{d}\mathbf{x} - \int_\Gamma \partial_\nu p(\mathbf{x}) \overline{v}(\mathbf{x})\, \mathrm{d}\mathbf{x} \\
&= \sum_{K_e} \int_{K_e} \Big(-\omega^2 c(\mathbf{x})^{-2} p(\mathbf{x}) \overline{v}(\mathbf{x}) + \nabla p(\mathbf{x}) \nabla \overline{v}(\mathbf{x}) \Big)\, \mathrm{d}\mathbf{x} \\
&\qquad - \int_\Upsilon [\nabla p(\mathbf{x}) \overline{v}(\mathbf{x})]\, \mathrm{d}\mathbf{x} - \int_\Gamma \partial_\nu p(\mathbf{x}) \overline{v}(\mathbf{x})\, \mathrm{d}\mathbf{x}.
\end{aligned}$$

Using the identity (2.7) and the fact that $[\nabla p] = 0$ we eventually get

$$\begin{aligned}
&\sum_{K_e} \int_{K_e} \Big(-\omega^2 c(\mathbf{x})^{-2} p(\mathbf{x}) \overline{v}(\mathbf{x}) - \Delta p(\mathbf{x}) \overline{v}(\mathbf{x}) \Big) \, \mathrm{d}\mathbf{x} \\
&= \sum_{K_e} \int_{K_e} \Big(-\omega^2 c(\mathbf{x})^{-2} p(\mathbf{x}) \overline{v}(\mathbf{x}) + \nabla p(\mathbf{x}) \nabla \overline{v}(\mathbf{x}) \Big) \, \mathrm{d}\mathbf{x} - \int_{\Upsilon} \{\partial_\nu p(\mathbf{x})\} [\overline{v}(\mathbf{x})] \, \mathrm{d}\mathbf{x} \\
&\qquad - \int_{\Gamma} \partial_\nu p(\mathbf{x}) \overline{v}(\mathbf{x}) \, \mathrm{d}\mathbf{x}.
\end{aligned}$$

This formulation is not coercive and it is common to introduce a penalty term (Riviere, 2008). Therefore, we add the term $\alpha_e [p][v]$, where the index e indicates that α depends on the cell. We get

$$\begin{aligned}
&\sum_{K_e} \int_{K_e} \Big(-\omega^2 c(\mathbf{x})^{-2} p(\mathbf{x}) \overline{v}(\mathbf{x}) + \nabla p(\mathbf{x}) \nabla \overline{v}(\mathbf{x}) \Big) \, \mathrm{d}\mathbf{x} - \int_{\Upsilon} \{\partial_\nu p(\mathbf{x})\} [\overline{v}(\mathbf{x})] + \alpha_e [p][v] \, \mathrm{d}\mathbf{x} \\
&\qquad - \int_{\Gamma} \partial_\nu p(\mathbf{x}) \overline{v}(\mathbf{x}) \, \mathrm{d}\mathbf{x}.
\end{aligned}$$

Remark 2.1. *In order to handle the Dirichlet boundary condition on Γ_1, we have introduced the space function V where the function is zero on Γ_1. As an alternative, we can instead include a penalization term onto this boundary. Replacing with the Neumann boundary condition on Γ_2, the formulation becomes*

$$\begin{aligned}
&\sum_{K_e} \int_{K_e} \Big(-\omega^2 c(\boldsymbol{x})^{-2} p(\boldsymbol{x}) \overline{v}(\boldsymbol{x}) + \nabla p(\boldsymbol{x}) \nabla \overline{v}(\boldsymbol{x}) \Big) \, \mathrm{d}\boldsymbol{x} - \int_{\Upsilon} \{\partial_\nu p(\boldsymbol{x})\} [\overline{v}(\boldsymbol{x})] + \alpha_e [p][v] \, \mathrm{d}\boldsymbol{x} \\
&\qquad - \int_{\Gamma_1} \partial_\nu p(\boldsymbol{x}) \overline{v}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} + \tilde{\alpha}_e [p][v] \, \mathrm{d}\boldsymbol{x}.
\end{aligned}$$

The space of discretization V_h is defined with piecewise polynomial functions of order less than or equal to r over the tetrahedra, similarly to the FE method,

$$V_h = \big\{ w \in V \quad \text{such that} \quad \forall i, \quad w \,|_{K_i} \in \mathbb{P}_r(K_i) \big\},$$

where $\mathbb{P}_r$ denotes the space of polynomials of order less than or equal to r given in the equation (2.6). Following this discretization we can inject in the variational formulation and one can eventually obtain a linear system. It follows exactly the same techniques as for the CG/FE method that we have detailed above, and one similarly makes use of the reference element and the Gauss quadrature for the computation.

2.2.4 Hybridizable Discontinuous Galerkin

The HDG method introduces two levels of discretization. A global linear system is formed for Lagrange multipliers located on the interface of the cells only. From these multipliers

the field of interest is retrieved locally on the cell. The implementation of this method in our software has been realized by Bonnasse-Gahot (2015) where we refer for all details as well as the validation of the method. In Subsection 4.4.7, we briefly review the formulation and depict the specificities it generates regarding the computation of the gradient of the cost function (introduced for the resolution of the inverse problem).

2.3 Artificial boundary conditions

Boundary conditions are obviously essential for numerical approximations. We must clearly distinguish the actual boundary (interface between the air and the medium), and the boundary forced by the limitation of the domain of interest. For the actual interface boundary between the air and the ground, we have already mentioned the free surface boundary conditions, see Section 1.8. In Figure 2.7 we illustrate the different boundaries in the seismic context, in two dimensions.

Figure 2.7: Distinction between artificial boundary conditions and free surface. The interface between the air and the medium, Γ_1 is a free surface. In seismic numerical applications, Γ_2 must ensure no reflection from incident waves.

The numerical computation is performed on a spatially limited, finite, domain of interest, which can be of several kilometers but the selected seismic domain is only a part of the Earth subsurface. The waves should leave the numerical domain freely. Hence the numerical treatment must ensure that no wave are reflecting or appearing from these numerical boundaries. For this purpose, we review the two most common approaches: perfectly matched layer and absorbing boundary conditions, we refer to Diaz (2005); Nataf (2013) and the references therein for additional details.

Perfectly Matched Layers (PML). The perfectly matched layers (PML) have been initially promoted in electromagnetic for the Maxwell equations with the work of Bérenger (1994). It consists in the incorporation of an artificial layer where incident waves are not reflected. We illustrate the layers in Figure 2.8.

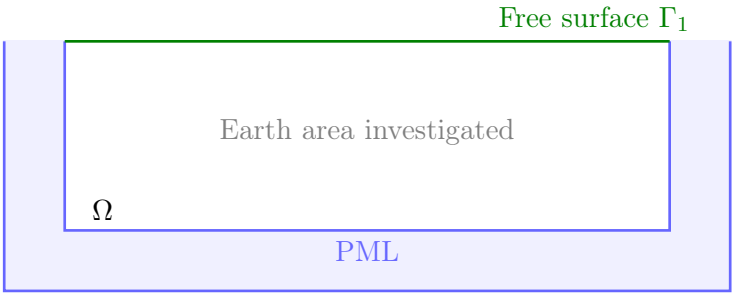


Figure 2.8: Illustration of the PML: a layer is incorporated on the boundary of the domain where no reflection is expected.

In the PML layer, the derivative is rewritten such that

$$\partial_x \longmapsto \left(1 + \mathrm{i}\frac{\sigma(x)}{\omega}\right)^{-1} \partial_x,$$

where σ is a selected function. It can have different shapes and in our case it is defined from the work of Turkel and Yefet (1998); Wang et al. (2011) which includes a cosine function.

Absorbing Boundary Conditions (ABC). As an alternative one can consider absorbing boundary conditions which are limited to Γ_2, they are defined by Engquist and Majda (1977),

$$\partial_\nu p(\mathbf{x}) - \mathrm{i}\omega c^{-1}(\mathbf{x}) = 0, \qquad \text{on } \Gamma_2.$$

Higher order ABC can be obtained, at order two it is defined by

$$-\mathrm{i}\omega c^{-1}(\mathbf{x})\partial_\nu p(\mathbf{x}) - \omega^2 c^{-2}(\mathbf{x})p(\mathbf{x}) - \frac{1}{2}\partial_\tau^2 p(\mathbf{x}) = 0, \qquad \text{on } \Gamma_2,$$

where ∂_τ represents the tangential derivative.

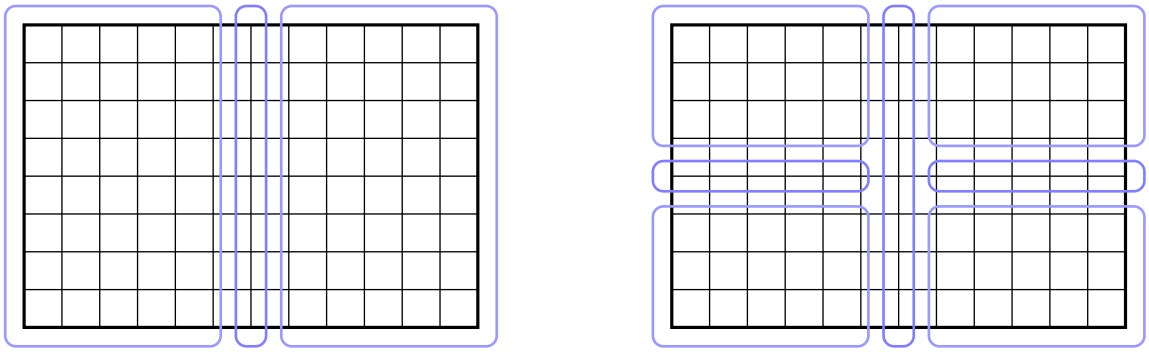
2.4 Direct solvers for linear system solution

The discretization of the wave equations leads to the formation of a linear system, whatever method is used, FD, FE or DG. The underlying system can represent a large scale matrix and we rely on external packages for the resolution of these linear systems. Two main families exist for the solver: direct and iterative. In our project, we only rely on direct solvers, which factorize the matrix. This step represents an important computational effort (time and memory) but is of importance as it allows multi-right-hand sides and the re-use of the factors. The multi-right hand sides technique refers to the possibility of solving several independent right-hand sides at low cost. Seismic acquisition usually consists in hundreds of independent sources so this option is extremely valuable. Similarly the reuse of the factor

is helpful for the inverse problem and the computation of the gradient (see Section 4.4). Despite current advances, iterative solvers are still restricted on those aspects, but they are usually less demanding for the computational memory, and we do not discard them for future applications.

MUMPS. MUltifrontal Massively Parallel sparse direct Solver (Amestoy et al., 2001, 2006) is a reference among direct solvers for sparse matrices. Born from the collaborations of six French institutes (CERFACS, CNRS, ENS Lyon, INP Toulouse, Inria and University of Bordeaux), it is in constant evolution and benefits from continuous support. Recently, block low rank techniques have been implemented to reduce the factorization time.

Hsolver. We have also been able to use the Hsolver, which has been developed by the work of Wang et al. (2011, 2012) as part of the Geo-Mathematical Imaging Group[1]. This solver is optimized for the FD method by intrinsically considering the structured decomposition. It is based upon a structured nested dissection of the domain to achieve high parallelization. In Figure 2.9 we illustrate the concept of nested dissection for a structured domain decomposition. The size of the interface between two subdomains depends on the order of the FD method.



(a) Level 1. (b) Level 2.

Figure 2.9: FD discretization (black) and nested dissection decomposition (blue). At level one (left) two sets are separated by an interface so that the two main groups are independent and interact only with the interface set (assuming the FD discretization involves only the closest neighbors). Higher levels of nested dissection follow the same concept.

Then the factorization of the matrix is conducted using a multi-frontal method with the

[1]The GMIG, previously at Purdue University, is now at Rice University, see http://maartendehoop.rice.edu/

possibility of promoting Hierarchically SemiSeparable (HSS) matrix, see Xia et al. (2009, 2010). Therefore, this solver is particularly efficient but limited to the FD method.

A follow-up of this solver is currently realized, to allow the same features with unstructured mesh, see Xin et al. (2017). This is part of an ongoing collaboration and we expect to be able to evaluate the performance (compared to MUMPS) in a near future.

2.5 Numerical experiments in the frequency domain

We briefly illustrate the numerical resolution of the wave equation with a three-dimensional acoustic experiment and a two-dimensional anisotropic one. In the Appendix B, we provide the detail of the computational toolbox implemented during this project. Regarding the HDG method, thorough experiments have been conducted in Bonnasse-Gahot (2015) where the reader is referred for additional information.

In Figure 2.10, we show the resolution of the acoustic isotropic wave equation where we have used the velocity of Figure 2.3(b). It is a model of size $2.54 \times 1.44 \times 1.22$ km, which is further detailed in Chapter 7. For this experiment we have taken the frequency to be 10 Hz and used a FE method.

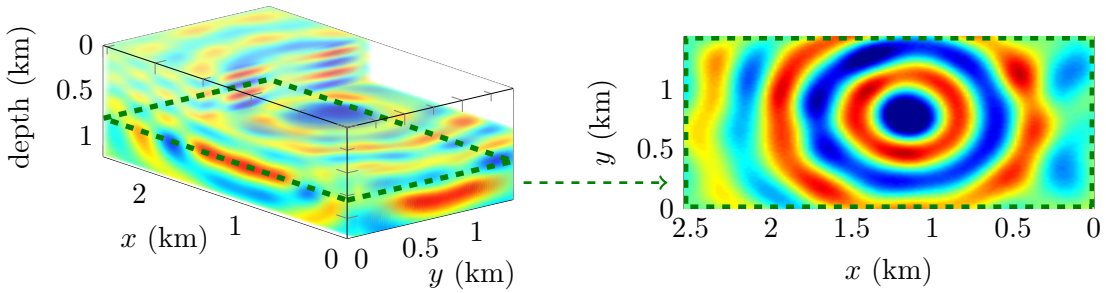
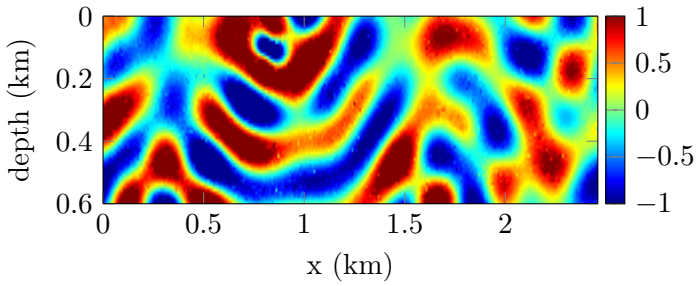


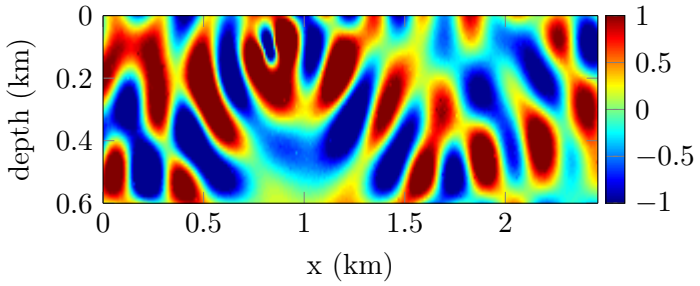
Figure 2.10: Resolution of the acoustic three-dimensional wave equation for the wave speed given in Figure 2.3(b) for which the density is constant: $\rho = 1000\ \mathrm{kg\,m^{-3}}$.

In Figure 2.11, we show the resolution of the TTI wave equation (1.30) in two dimensions for homogeneous parameters. The domain is of size 2.46 by 0.6 km and the parameters have the following constant values: $c_p = 5000\ \mathrm{m\,s^{-1}}$, $c_s = 2400\ \mathrm{m\,s^{-1}}$, $\rho = 1000\ \mathrm{kg\,m^{-2}}$, $\theta = 10°$, the Thomsen parameters are given by $\epsilon = 0.20$, $\delta = 0.10$. For the simulation the frequency is set to 10 Hz.

In Figure 2.12, we show the resolution of the TTI wave equation in two dimensions for variable parameters. This is the same domain as for the previous experiment, 2.46 by 0.6 km. The medium parameters are given in Figure 2.13.

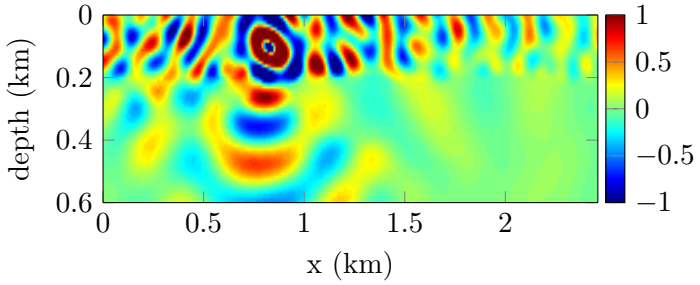


(a) U_x

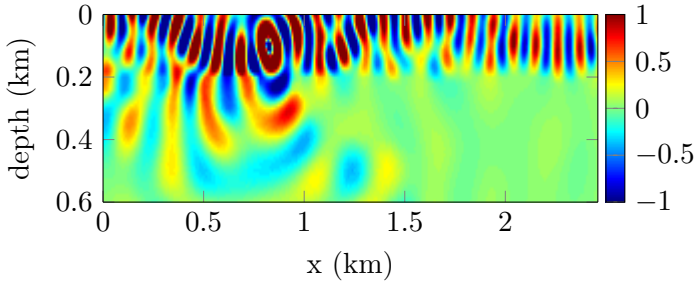


(b) U_z

Figure 2.11: Simulation of TTI wave propagation using a homogeneous two-dimensional medium of size 2.46 by 0.6 km. The constant parameters are $c_p = 5000\ \mathrm{m\,s^{-1}}$, $c_s = 2400\ \mathrm{m\,s^{-1}}$, $\rho = 1000\ \mathrm{kg\,m^{-2}}$, $\theta = 10°$, and the Thomsen parameters $\epsilon = 0.20$, $\delta = 0.10$.

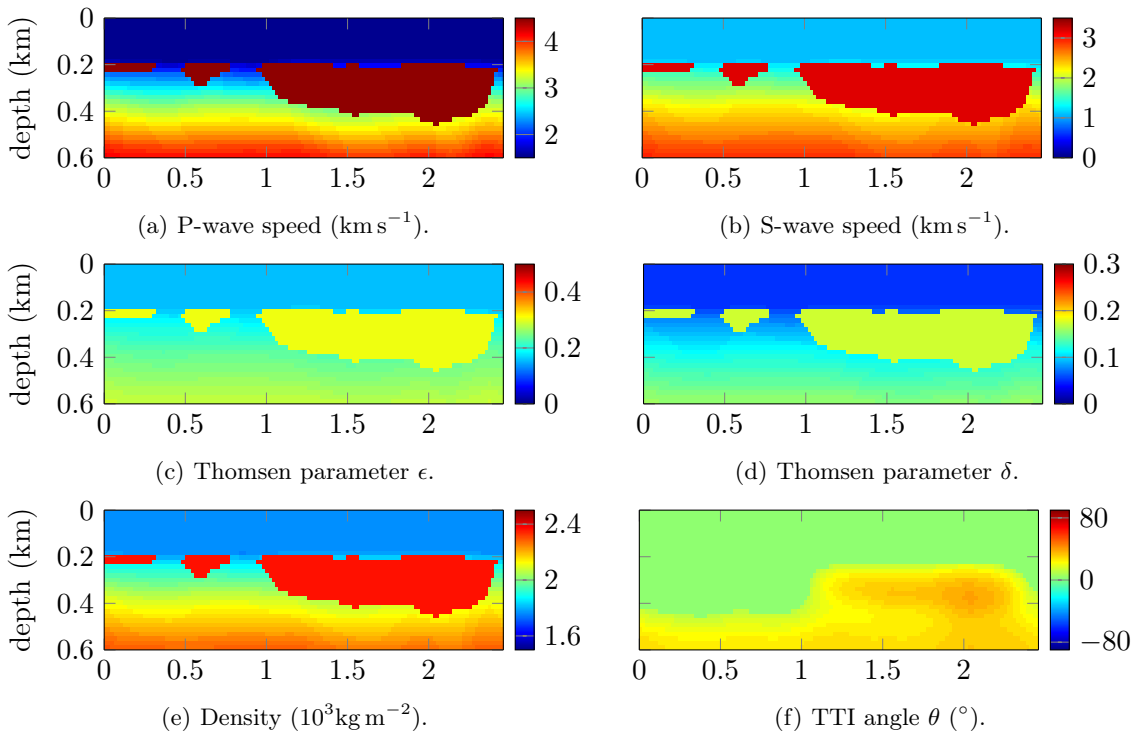


(a) U_x



(b) U_z

Figure 2.12: Simulation of TTI wave propagation using a two-dimensional medium of size 2.46 by 0.6 km, the model parameters are given Figure 2.13.



(a) P-wave speed ($\mathrm{km\,s^{-1}}$).

(b) S-wave speed ($\mathrm{km\,s^{-1}}$).

(c) Thomsen parameter ϵ.

(d) Thomsen parameter δ.

(e) Density ($10^3\mathrm{kg\,m^{-2}}$).

(f) TTI angle θ (°).

Figure 2.13: Experimental TTI models of size 2.46 by 0.6 km for the numerical resolution of the wave equation, the x axis is indicated in km.

2.6 Conclusion

In this chapter, we have briefly introduced the methods for the numerical resolution of the wave equations. We have given the ones that have been used during this project: FD, FE/CG, DG and HDG methods. The FD method relies on a structured grid and is usually the fastest option. However, to handle complex geometry and topography, the other methods are certainly more appropriate. we have given some illustrations for the resolution of the wave equation. The inverse problem requires the resolution of several forward problems, and it is consequently fundamental to have a robust, efficient and fast discretization method.

Chapter 3

Stability results for the Helmholtz inverse problem

Abstract

In this chapter, we introduce the principal concepts relative to inverse problems, in particular, the one we face in seismic, associated with harmonic wave equations. We give the general framework and notations that will be employed throughout the manuscript. This inverse problem is nonlinear and ill-posed. We show that the conditional Lipschitz stability of the inverse problem for the Helmholtz equation can be obtained and we give the result that defines the lower and upper bound of the underlying stability constant. The analytical formulations are then confronted with the numerical situation where we demonstrate the sharpness of the results in the case of seismic reconstruction. Numerically, we focus on one of the main *conditions* for the conditional stability to hold: the use of piecewise constant parameters and we propose different methods for such model representations. Following the stability analysis, we establish a relation between frequency and scale.

Contents

Chapters 1 and 2 were devoted to the forward problem for wave propagation. The mathematical formulation of the equations and their numerical resolution have been processed. In this context, there is the need to know all parameters of the system: the type of medium, the associated material properties (which define the stiffness tensor) and the source term. In seismic, the subsurface Earth parameters must be known, and then one can anticipate the propagation of waves. The inverse problem results from the opposite vision and aspires, instead, in the recovery of the unknown subsurface parameters from the observation of wave propagation. In addition, the observation is usually limited at the surface. This notion of forward or inverse appears quite clear in geophysics: the forward problem predicts the waves, and the inverse tends to recover the medium parameters. It can however be a matter of perspective depending on the context, and we refer to Keller (1976) for an interesting discussion. The method to conduct the reconstruction of the parameters depends on the situation. One would be tempted to say that there are as many methods of reconstruction as there are of inverse problems. It depends fundamentally on the kind of forward operator and the data available. In the context of scattering, we refer to Colton and Kress (1998). Bal (2012) gives the details on some techniques utilized to solve inverse problems. For seismic reconstruction, we note the early work of Pekeris (1934), which studies the inverse boundary problem in seismology for the recovery of the density and the Lamé parameters in the half-space.

The success of the reconstruction procedure is correlated with the method and the quantity of information at our disposal: the quantity of observations, prior knowledge on the unknowns, etc. In our seismic applications, the acquisition setup is usually located at the surface only, i.e., we disregard boreholes. This configuration provides a physical framework where the body is not altered by the experiment (depending on the power of the source): it is called a *nonintrusive* or *non-invasive* process. The first consequence is that the procedure suffers from this limited amount of information. Mathematically we formulate the reconstruction as an inverse boundary value problem (IBVP) for obvious reasons: the data are only acquired onto the boundary of the domain of interest, and similarly for the initial source term. In our case as in many, the inverse problems is nonlinear and ill-posed, Kirsch (1996); Colton and Kress (1998); Isakov (2006); Bal (2012); Tarantola (2005).

The notions of existence and uniqueness are crucial in the inverse problem, always following the conception of Hadamard (1923; 1902). The uniqueness in particular guarantees that one and only one solution exists for the inverse problem: is there only one medium that can produce the observed data? The answer is not trivial in general, depends on the problem (homogeneous, anisotropy, etc) and incorporating noise and partial data further complicates the process. Here again, a priori information and the amount of data are revealed to be capital, see Tarantola and Valette (1982) for discussions related to the seismic problem. In order to mitigate the ill-posedness of the inverse problem, one can incorporate

regularization, as advocated, for example, in Kirsch (1996); Isakov (2006).

The growth in the comprehension of inverse problems is marked with medical imaging, in particular, the Electrical Impedance Tomography (EIT), where the recovery is based upon electrical measurements (our seismic problem relates to mechanical measurements). The formulation of this problem by Calderón (1980) gave birth to the eponymous Calderón problem which is a milestone. This problem is also referred to as the *Gel'fand* problem, see Novikov (1994). The considered type of data inherits from this EIT approach with the Dirichlet-to-Neumann map. In Section 3.1, we review the formulation of the problem. In this chapter, we focus on the inverse problem associated with this Helmholtz equation (1.27). It corresponds with an acoustic medium where the objective is the reconstruction of the velocity c.

This problem has been increasingly studied since then, and here we only point at the appropriate references on the topic. We do not aim to cover the techniques and methods in details, as it is not the purpose of this manuscript. In particular, existence and uniqueness for this problem have been obtained by Sylvester and Uhlmann (1987) in three dimensions with the use of complex geometrical optics (CGO) solutions. These are exponential solutions of the equation (e.g. Uhlmann and Wang (2008); Feldman et al. (2015)). It clearly represents a landmark of the Calderón problem and inverse problem in general. The two-dimensional case is also investigated by Novikov (1986); Nachman (1988) and Nachman (1996) gives a global uniqueness result. For the Calderón (or Gel'fand) problem, we mention the results of Henkin and Novikov (1988); Novikov (1988, 2004, 2010, 2011). For the state of the art review, we refer to Sylvester and Uhlmann (1990); Novikov (1994); Uhlmann (2009); Feldman et al. (2015) and the references therein. The case of partial data is studied in Kenig et al. (2007); Imanuvilov et al. (2010); Feldman et al. (2015). Nakamura and Uhlmann (1993, 1994) study reconstruction in the case of elasticity, where uniqueness holds if the Lamé parameter μ is close to constant, see Nakamura and Uhlmann (2003).

Here we focus on inquiring the stability of the inverse problem. It can be apprehended as the response of the solution to small perturbations in the acquired data. It provides a comprehensive relation between two sets of data, and their respective parameters. In Section 3.2, we introduce the notion of stability and the fundamental work associated. This project resulted in the production of Beretta et al. (2016), which is reproduced in the Appendix A. It establishes the lower and upper bound of the stability constant in the case of piecewise constant model representation. Supplementary materials are given in Section 3.3, and we develop different methods to decompose the model in Section 3.4. Sometimes the mathematical study requires relatively strong assumptions, such that full boundary information or the absence of noise, which are unrealistic in practical situations. We will see that despite those restrictions it can give precise guidelines on how to conduct

the reconstruction of unknown parameters. The stability in the context of elasticity, for the recovery of the density and the Lamé parameters, is considered by Beretta et al. (2013b, 2014). For the time domain, we mention the work of Stefanov and Uhlmann (1998, 2005, 2009); Datchev and de Hoop (2015) show the convergence of the Landweber iteration scheme when the frequency content of the data is considered bounded. It also motivates our approach in the frequency domain directly.

3.1 Minimal foundations

3.1.1 Notation

Let us first introduce the general inverse problem framework and associated quantities. We consider the domain $\Omega \subset \mathbb{R}^r$ with boundary Γ for which we aim the recovery (reconstruction) of selected *parameters* of interest. Typically r is equal to 2 or 3 as we consider two and three-dimensional cases. The parameter of interest will also be referred to as *model* due to the seismic context, it is denoted

$$m \in \mathcal{M} \qquad \text{represents the model (or parameter) to recover.} \tag{3.1}$$

m actually encompasses one or several physical parameters for the seismic subsurface (e.g., the density, the velocity, etc). $\mathcal{M}$ is the *model space* and represents the subspace where those parameters are. In our case it usually consists in a subset of $\mathbb{R}$ as we research physical coefficients. The model space takes any a priori knowledge on the subsurface model.

The information that are at the disposal for the reconstruction are referred to as the *data* or *measurements*, and denoted d. Due to physical feasibility, the data are acquired on some discrete part of the space which we denote Σ. There is no restriction for Σ, which can be away from Ω (think of helioseismology where information is not acquired in or at the surface of the sun but at relative distance). We write

$$d \in \mathcal{D} \qquad \text{represents the data (or measurements),} \tag{3.2}$$

where $\mathcal{D}$ is the data space. One can be tempted to assume that $\mathcal{D} \subset \mathbb{R}$, because it comes from physical apparels. However, because we work in the frequency domain we need to perform the Fourier transform of the data, so that eventually $\mathcal{D} \subset \mathbb{C}$ for our problem. Typically, $\mathcal{D}$ and $\mathcal{M}$ are assumed to be Banach or Hilbert spaces (Bal, 2012).

The *forward operator* prescribes the correspondence between the model and the quan-

tity observed (the data):

$$\mathcal{F} \ : \ \mathcal{M} \to \mathcal{D}, \quad \mathcal{F}(x) = z. \tag{3.3}$$

This notation is further extended for the numerical resolution of the inverse problem in Section 4.1. For example, in seismic, x is identified with the subsurface characterization (velocity, density, etc) and z corresponds to the measurements of waves at the receivers locations.

Remark 3.1. *It is usually not conceivable to assume that the observations d verifies $\mathcal{F}(m^\dagger) = d$ for some parameter $m^\dagger$. Several reasons are involved: the noise in the acquisition and limited accuracy of the devices. Another aspect is the model error: the approximation used to represent $\mathcal{F}$ numerically (any discretization scheme) generates imperfection. In the case of seismic, one straightforward implication is the use of special medium characteristics (isotropy, elasticity, titled isotropy) to represent the anisotropic Earth. This aspect is postponed for now but will be investigated in Chapter 5, where the maximal distance between the observations and the attainable set is investigated.*

3.1.2 The Calderón problem

Following the Helmholtz equation provided in the Equation (1.27), the parameter of interest, that describes the medium, is the velocity c. We establish that this problem can be derived from the commonly denominated *Calderón problem*.

The Calderón problem is named in reference of the work of Calderón (1980) who defines the inverse problem for Electrical Impedance Tomography (EIT) (see Uhlmann (2009) for historical background). It is formulated as the inverse boundary value problem associated with the conductivity equation:

$$\begin{cases} \nabla \cdot \big(\gamma(\mathbf{x}) \nabla u(\mathbf{x})\big) = 0 & \text{in } \Omega, \\ u(\mathbf{x}) = f(\mathbf{x}) & \text{on } \Gamma. \end{cases} \tag{3.4}$$

We assume the Dirichlet boundary condition represented by $f(\mathbf{x})$, γ stands for the electrical conductivity and is positive. Here f and u are voltage. Furthermore, we require $f \in H^{1/2}(\Gamma)$, the conductivity $\gamma \in C^\infty(\Omega)$ and the solution $u \in H^1(\Omega)$. The inverse problem consists of the reconstruction of the conductivity γ from observation at the boundary. In EIT the measurements are current (while the conductivity problem gives the voltage) so that those boundary data are defined as the Dirichlet-to-Neumann (DtoN) map Λ_γ,

$$\Lambda_\gamma : f(\mathbf{x}) \quad \to \quad \gamma(\mathbf{x}) \frac{\partial u(\mathbf{x})}{\partial \nu}\bigg|_\Gamma,$$

where ∂_ν is the normal derivative. In the case of EIT, the DtoN map is also referred to as

the *current-to-voltage* map for obvious physical reason. This problem has been extensively studied in the mathematical framework, as mentioned in the introduction.

3.1.3 Relation with the Helmholtz equation

At first it seems that the problem stated with Equation (3.4) differs from the wave equation we have established. Actually the conductivity equation can be generalized to the Schrödinger operator, hence showing similarities with the Helmholtz equation. A few simple steps have to be taken to see the relation, as reminded in Nachman (1988, 1996); Mandache (2001). Let us make the substitution

$$u(\mathbf{x}) = \gamma(\mathbf{x})^{-\frac{1}{2}} w(\mathbf{x}).$$

We can replace in the conductivity equation to have (omitting the $\mathbf{x}$ dependency for clarity)

$$\nabla \cdot \Big(\gamma \nabla \big(\gamma^{-\frac{1}{2}} w \big) \Big) = 0.$$

We precise the formula

$$\begin{cases} \nabla(\gamma^{-\frac{1}{2}}) = -\dfrac{1}{2}(\gamma^{-\frac{3}{2}})\nabla(\gamma), \\ \nabla(\gamma^{\frac{1}{2}}) = \dfrac{1}{2}(\gamma^{-\frac{1}{2}})\nabla(\gamma), \end{cases}$$

so that

$$\nabla(\gamma^{-\frac{1}{2}}) = -\frac{1}{2}\Big(\gamma^{-\frac{3}{2}}\Big)\Big(2\gamma^{\frac{1}{2}}\nabla(\gamma^{\frac{1}{2}})\Big) = -\gamma^{-1}\nabla(\gamma^{\frac{1}{2}}).$$

The conductivity equation can eventually be written as

$$\begin{aligned} &\nabla \cdot \Big(- w\nabla\big(\gamma^{\frac{1}{2}}\big) + \gamma^{\frac{1}{2}}\nabla\big(w\big) \Big) \\ =& - \nabla \cdot \Big(w\nabla(\gamma^{\frac{1}{2}}) \Big) + \nabla \cdot \Big(\gamma^{\frac{1}{2}}\nabla(w) \Big) \\ =& - w\Delta\gamma^{\frac{1}{2}} - \nabla(\gamma^{\frac{1}{2}})\nabla(w) + \gamma^{\frac{1}{2}}\Delta(w) + \nabla(\gamma^{\frac{1}{2}})\nabla(w) \\ =& - w\Delta\gamma^{\frac{1}{2}} + \gamma^{\frac{1}{2}}\Delta(w). \end{aligned}$$

The original equation is replaced by

$$-w\gamma^{-\frac{1}{2}}\Delta\gamma^{\frac{1}{2}} + \Delta(w) = 0.$$

We can further introduce $q = \gamma^{-\frac{1}{2}}\Delta\gamma^{\frac{1}{2}}$,

$$-wq + \Delta(w) = 0,$$

and we recognize the Schrödinger operator. Finally, assuming q is strictly negative we obtain the Helmholtz equation. Moreover, Nachman (1988) reveals that the Calderón problem can

be generalized to the general acoustic propagation encompassing a variable density,

$$-\nabla \cdot \big(\frac{1}{\rho}\nabla p\big) - \omega^2 \kappa p = 0.$$

3.2 Stability result for the Helmholtz inverse problem

The stability is a fundamental characteristic of the inverse problem. It provides a relation between measurements and models. Assuming two parameters m_1 and m_2, the stability is of the general form

$$\|m_1 - m_2\|_{\mathcal{M}} \leq C\big(\|\mathcal{F}(m_1) - \mathcal{F}(m_2)\|_{\mathcal{D}}\big),$$

where $\mathcal{F}$ is the forward operator, which maps the quantities observed. The function C indicates the stability of the problem. The objective of the reconstruction is often to minimize the quantity $\|\mathcal{F}(m_1) - \mathcal{F}(m_2)\|$ and the stability provides an error on the reconstruction ($\|m_1 - m_2\|$) in terms of the error in the forward maps. Hence we see the crucial aspect of this function C. We further refer to Bal (2012) where extended discussion and intuition is provided, in particular how the stability relates to the injectivity of the forward operator and defines the ill-posedness of the inverse problem.

We study the stability of the inverse boundary value problem associated with the Helmholtz equation, using the Dirichlet-to-Neumann (DtoN) map as the data. In this context, see Equation (1.27), the problem concerns the recovery of one single parameter, the velocity c. The stability associated with this inverse problem has been obtained by Alessandrini (1988), where the function C is proven to be the inverse of a logarithm. This has further been proven to be optimal by Mandache (2001), with reference to "exponential instability". The way to overcome the issue is to introduce assumptions and derive a conditional stability. In particular, Alessandrini and Vessella (2005) obtain a Lipschitz type stability result by the use of piecewise constant models. Lipschitz type stability (e.g. Bourgeois (2013)) indicates that the stability condition reduces to the formula

$$\|m_1 - m_2\| \leq \mathcal{C}\|\mathcal{F}(m_1) - \mathcal{F}(m_2)\|,$$

where $\mathcal{C}$ is the stability constant. The result is also developed in the case of EIT by Beretta and Francini (2011).

This is the point of departure of our work, which is given in the Appendix A, where we reproduce the article that has been published in collaboration with E. Beretta, M. V. de Hoop and O. Scherzer. The objective is to characterize the stability constant to understand under which conditions it varies. The smallness of the constant guarantees the accuracy

of the reconstruction in terms of the proximity between observed quantities. Let us only summarize the main steps established in this paper:

1. we consider the Helmholtz equation with Dirichlet boundary condition on the domain of interest Ω,

$$\begin{cases} (-\omega^2 c^{-2}(\mathbf{x}) - \Delta)u(\mathbf{x}) = 0 & \mathbf{x} \text{ in } \Omega, \\ u(\mathbf{x}) = g(\mathbf{x}) & \mathbf{x} \text{ on } \partial\Omega. \end{cases}$$

   The forward operator is defined by

$$\mathcal{F}_\omega(c^{-2}) = \Lambda_{\omega^2 c^{-2}}.$$

   Λ is the DtoN map,

$$\Lambda_{\omega^2 c^{-2}} \ : \ g \to \partial_\nu u \mid_{\partial\Omega},$$

   where ∂_ν stands for the normal derivative.

2. The domain Ω is partitioned in N non-overlapping subdomains $\mathcal{D}_k$,

$$\Omega = \Big\{ \cup_{k=1}^N \mathcal{D}_k \ ; \quad \mathcal{D}_i \cap \mathcal{D}_j = \emptyset \quad \forall i \neq j \Big\}. \tag{3.5}$$

   The velocity is defined via a piecewise constant representation over the partition,

$$c^{-2}(\mathbf{x}) = \sum_{k=1}^{N} c_k \chi_{\mathcal{D}_k}(x), \tag{3.6}$$

   where $\chi_{\mathcal{D}_k}$ is the characteristic function over the domain $\mathcal{D}_k$ and c_k are constants. Hence N stands for the number of coefficients used to represent the parameter. It indicates the scale of the model: low values of N imply a coarse representation and large values of N give a refined representation.

3. From the work of Alessandrini and Vessella (2005), this problem is known to have Lipschitz type stability such that for two velocities c_0 and c_1,

$$\|c_0^{-2} - c_1^{-2}\|^2 \leq \mathcal{C} \|\mathcal{F}(c_0^{-2}) - \mathcal{F}(c_1^{-2})\|^2.$$

   $\mathcal{C}$ is the stability constant.

4. The main result of this production is to establish the lower and upper bounds for the stability constant, $\mathcal{C}$, depending on the frequency ω and the scale N:

$$\frac{1}{4\omega^2} e^{K_1 N^{1/5}} \leq \mathcal{C} \leq \frac{1}{\omega^2} e^{(K(1+\omega^2 B_2) N^{4/7})}, \tag{3.7}$$

   where B_2, K_1 and K are constants. Hence we have the quantification of the stability evolution with frequency and scale.

5. We provide numerical estimates for the evolution of the stability constant with scale. It validates the sharpness of the bounds in the case when the full DtoN map is used as the data (i.e. where the data are acquired on all the boundary).

6. We estimate the stability constant for the geophysical context where data are only located at the surface, and when artificial boundaries are introduced on the other sides. We show that the estimates follows accurately the full DtoN description of the Equation (3.7) and conjecture that it should hold in the context of partial data.

Following this work, we see that frequency and scale are crucial for the stability of the inverse problem. It is natural to think that increasing the number of unknowns in problem deteriorates its stability. Here we have shown that the stability worsen (i.e. the stability constant increases) exponentially with the number of coefficients to represent the model (N). Despite the lack of analytical results, our numerical estimation intuits that the case of partial data and absorbing boundaries would follow the same path.

In Section 3.3 we provide additional materials regarding the stability associated with the Helmholtz inverse problem. We give the evolution of the stability constant estimation with frequency in Subsection 3.3.1. Consequently, we illustrate the possibility for pairing frequency and scale in Subsection 3.3.2.

3.3 Supplementary materials

This section is the continuation of the work realized in the article given in the Appendix A, on page 93. We follow the same framework and extend the analysis with the same techniques, thus the reader must have read the Appendix A prior to read this supplement.

3.3.1 Numerical stability estimates with frequency

In the Appendix A, we have shown the evolution of the stability constant with scale, and demonstrated that it follows the exponential behavior anticipated by the lower and upper bounds given in the Equation (3.7). We now show the evolution of the stability constant estimates with frequency. We follow the setup of Section A.3, where the models and the procedure employed are prescribed. Contrary to the estimation of Figure A.5, we now fix the partition and vary the frequency. We plot the resulting estimates in Figure 3.1, in the context of partial data (described Subsection A.3.2).

The estimates decrease with increasing frequencies, in the selected range which is motivated by seismic so that the frequency varies from 1 to 50 Hz. Regarding the lower and upper bounds given in the Equation (3.7), we see that the lower bound decreases with

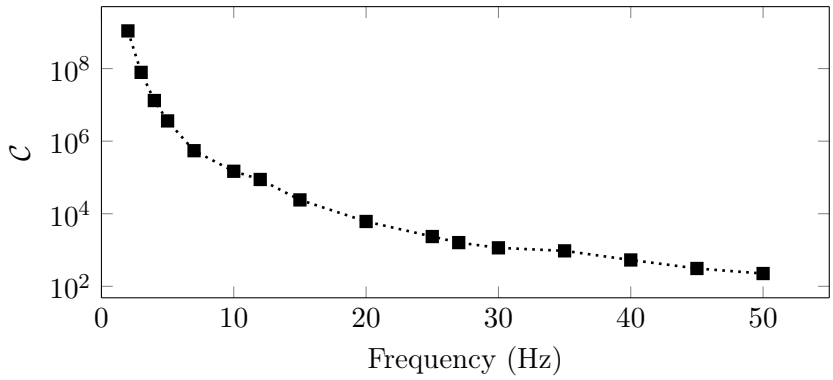


Figure 3.1: The black squares represent the computational estimates of the stability constant (··■··) depending on the frequency at selected partition $N = 4\ 597\ 248$. The model chosen for c_0 and c_1 are given Figures A.3 and A.2.

frequency (in ω^{-2}) but the upper bound makes appear the frequency in the exponential. Hence we can suspect that the frequency regime we experiment is relatively low and that the exponential increase with frequency appears at higher frequencies. However, as it has been mentioned in the introduction of the Appendix A, the exponential increase with frequency in the upper bound, Equation (3.7), may actually not be optimal. It is inherent from the use of CGO solutions: this may be an artifact of the method of proof and this is also remarked in Nagayasu et al. (2013). A more precise frequency characterization is subject of further research.

3.3.2 Pairing frequency and scale

Following the Appendix A and the previous Subsection 3.3.1, we conclude that the frequency and scale have the following impact on the stability of the inverse problem:

- the stability worsens (i.e. the stability constant increases) with increasing scale N (exponentially),
- the stability improves (i.e. the stability constant decreases) with increasing frequency.

Following this statement, a very natural idea is to pair frequency and scale so that the stability constant remains (relatively) stable. The low frequencies should consequently use a low scale and the higher frequencies can use a refined scale. To illustrate this idea we plot the evolution of the stability constant estimates with frequency for different partitions in Figure 3.2. We employ the setup described in Section A.3, Subsection A.3.2.

The decrease of the stability constant with frequency is sufficiently significant to allow a refinement in scale. If we consider a threshold for our estimate we can establish a progression

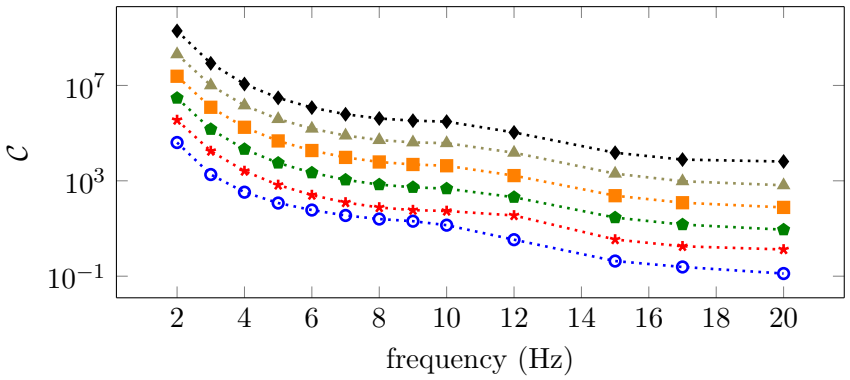


Figure 3.2: Evolution of the computational estimates of the stability constant depending on the frequency. The models chosen for c_0 and c_1 are given Figures A.3 and A.2. The blue circles (··o··) employ the partition $N = 336$. The red crosses (··*··) employ the partition $N = 2\,880$. The green pentagons (··⬟··) employ the partition $N = 23\,040$. The orange squares (··■··) employ the partition $N = 187\,492$. The grey triangles (··▲··) employ the partition $N = 1\,527\,168$. The black diamonds (··◆··) employ the partition $N = 4\,597\,248$.

in scale with frequency. We illustrate this selection in Figure 3.3 for an arbitrary threshold which allows a smooth progression in scale associated with the increase in frequency.

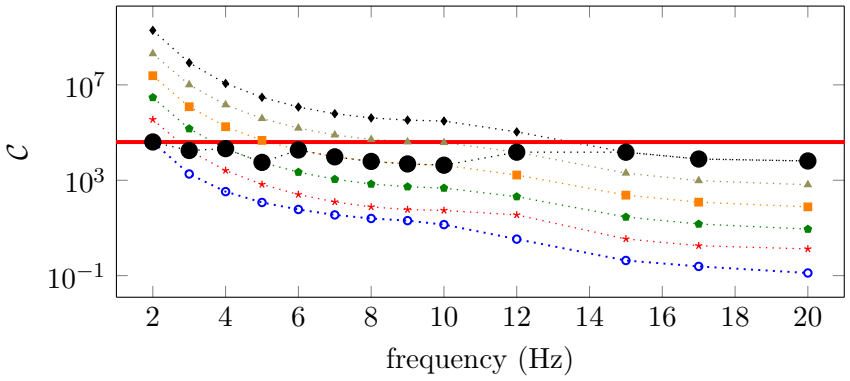


Figure 3.3: Illustration of the possible simultaneous progression in frequency and scale. The models chosen for c_0 and c_1 are given Figures A.3 and A.2. The blue circles (··o··) employ the partition $N = 336$. The red crosses (··*··) employ the partition $N = 2\,880$. The green pentagons (··⬟··) employ the partition $N = 23\,040$. The orange squares (··■··) employ the partition $N = 187\,492$. The grey triangles (··▲··) employ the partition $N = 1\,527\,168$. The black diamonds (··◆··) employ the partition $N = 4\,597\,248$. The plain black circles (··●··) select the largest scale which has an estimates lower than 4×10^4, value indicated by the red line (——).

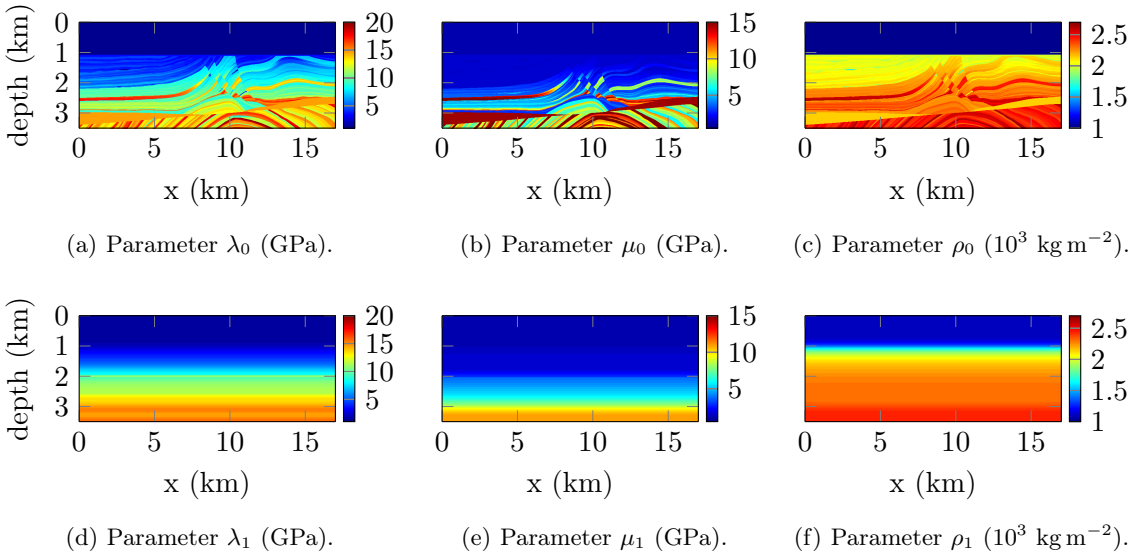
The stability analysis provides an understanding of frequency and scale matching. Yet, in terms of stability, nothing prevents us from selecting arbitrarily high frequency, to benefit from the improved stability (see Figure 3.1). There is no clear motivation for the frequency progression here. However, this progression, namely from low to high frequencies, is a well-known component for the numerical reconstruction. We shall explain the reason in Chapter 5, where the size of the radius of convergence is shown to evolve with frequency. The stability results associated with the convergence gives rise to our multi-level, multi-

frequency algorithm, where simultaneous progression of frequency and scale is employed, see Section 6.1.

3.3.3 Numerical stability estimates for elastic media

In the elastic case, Beretta et al. (2013b, 2014) show that the reconstruction can be obtained only if the Lamé parameters and the density are considered separately, still incorporating the piecewise constant assumption for the models representation. Here we want to probe numerically the stability constant in this elastic context. Hence we investigate the progression of the numerical estimates of the stability constant with frequency and scale. The exact progression is not known but we conjecture similar behavior as in the acoustic case.

We designed a two-dimensional elastic case in order to compute numerical estimation of the stability constant. It is based upon two sets of parameters, m_0 and m_1 defined by the Lamé parameters and the density such that $m_0 = \{\lambda_0, \mu_0, \rho_0\}$ and similarly for m_1. The models used are given in Figure 3.4, where m_0 is actually the Marmousi2 elastic model (see Chapter 5 and 6). Then the computational estimates follow the same method as prescribed in the Appendix A.



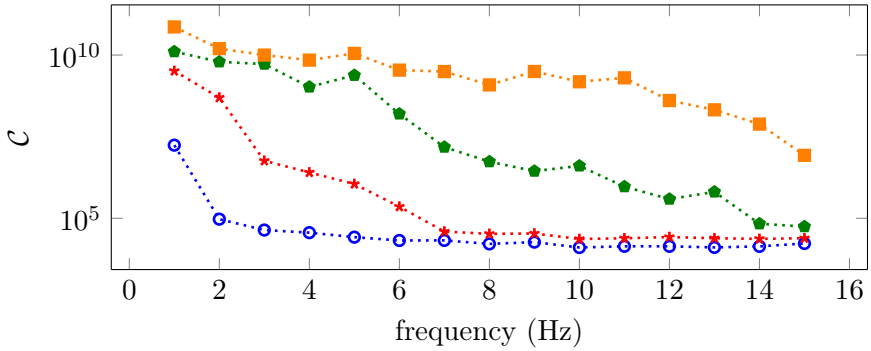
(a) Parameter λ_0 (GPa). (b) Parameter μ_0 (GPa). (c) Parameter ρ_0 (10^3 kg m^{-2}).

(d) Parameter λ_1 (GPa). (e) Parameter μ_1 (GPa). (f) Parameter ρ_1 (10^3 kg m^{-2}).

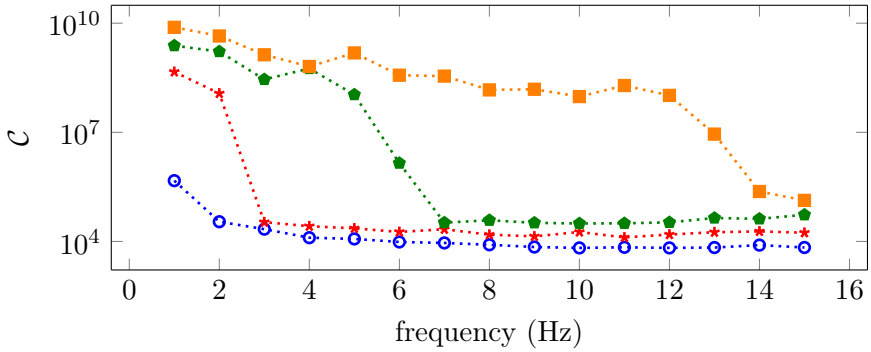
Figure 3.4: Elastic media of size 17×3.5 km used for the numerical estimates of the stability constant, m_0 is the Marmousi2 model.

In Figure 3.5 we show the estimates of the stability constant with frequency in the elastic case. We compare this evolution for different partitioning, in a similar fashion as in Figure 3.2 for the acoustic case. The method for the model decomposition is given in Section 3.4. Motivated by the work of Beretta et al. (2014), we only consider the Lamé

parameter λ and μ.



(a) Stability estimates for λ.



(b) Stability estimates for μ.

Figure 3.5: Evolution of the computational estimates of the stability constant depending on the frequency for elastic media. The models chosen for m_0 and m_1 are given Figure 3.4. We compare the evolution for the Lamé parameters λ and μ. The blue circles (··o··) employ the partition $N = 530$. The red crosses (··*··) employ the partition $N = 2\ 226$. The green pentagons (··⬟··) employ the partition $N = 9\ 116$. The orange squares (··■··) employ the partition $N = 36\ 975$.

We see that the estimates associated with the elastic problem behaves similarly to the acoustic one: they decrease with increasing frequencies, and increase with increasing scale. It demonstrates that the relation between frequency and scale should not be restricted to the acoustic situation. We later provide numerical experiments of reconstruction in the acoustic and elastic framework in Chapter 6, based on this simultaneous progression.

3.4 Numerical representation of the model

An important aspect to obtain the stability result is the consideration of a piecewise constant representation for the velocity, see Equations (3.5) and (3.6). In this section, we review some numerical possibilities that have been implemented for the construction.

Remark 3.2. *All of the representations that are subsequently presented have been imple-*

mented in the software developed to solve the seismic inverse problem, depicted in Chapter 4 and Appendix B. The illustrations we provide are two-dimensional for a better visualization but all representations have been implemented to treat three-dimensional domains as well (as it is used in the Appendix A).

3.4.1 Structured and unstructured representations

We illustrate the domain partition with a two-dimensional model, pictured in Figure 3.6. It is actually the Marmousi2 P-wave speed that is later employed in numerical experiments. It is of size 17×3.5 km with values between 1450 and 5500 $\mathrm{m\,s^{-1}}$. The model is originally represented via a structured architecture of 1701×351 nodes for a total of 597 051 coefficients.

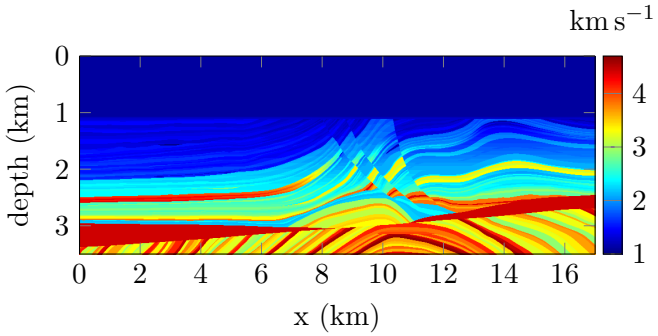
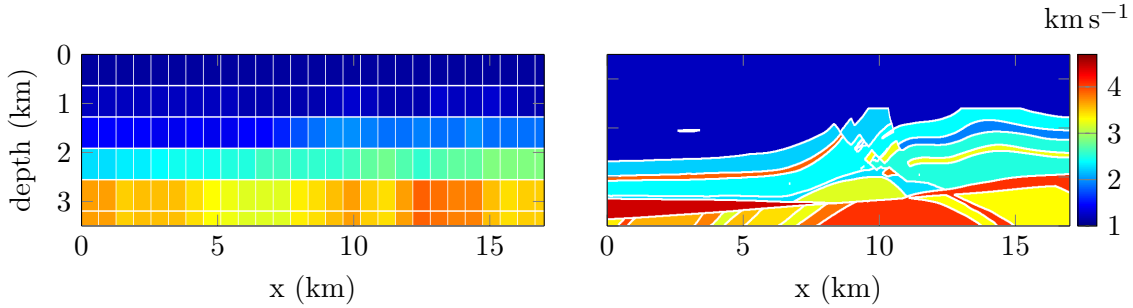


Figure 3.6: Initial model used for the illustration of model partition, it is the elastic Marmousi2 wave speed of size 17×3.5 km. the velocity varies between 1450 and 5500 $\mathrm{m\,s^{-1}}$ and is represented via 597 051 nodal coefficients.

The creation of the domain partition is said to be structured if the resulting subdomains are homogeneous. In 2D, the domain is divided into n_h horizontal sections and n_v consecutive vertical nodes. The new representation is the collection of the resulting subdomains, where the coefficients are given by the average value of the original image. When the number of consecutive nodes is given by a power of two, $n_h = n_v = 2^h$, with h an integer, it can actually be assimilated to Haar wavelets (Haar, 1911).

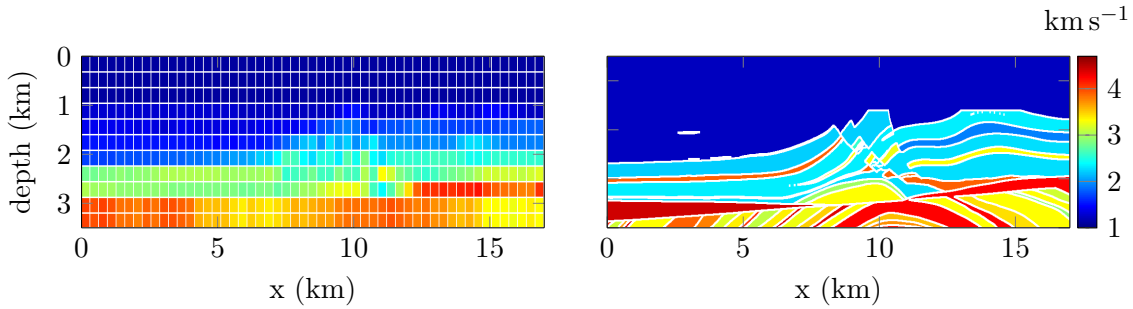
As an alternative we allow the possibility for unstructured decomposition. The procedure consists in creating subdomain as group of adjacent nodes which have values that are close. In our method, the notion of 'close' is made via an automatic criterion decided to reduce the amount of coefficients in the representation. This method allows much more flexibility but is based upon an initial image for the decision. In Figures 3.7 and 3.8 we compare the model decomposition with structured or unstructured decision, reducing the total number of coefficients to less than 0.1% of the original scale.



(a) Structured decomposition resulting in 130 subdomains with piecewise constant representation.

(b) Unstructured decomposition leading to 120 subdomains with piecewise constant representation.

Figure 3.7: Piecewise constant decomposition of the model pictured Figure 3.8 from structured and unstructured decision resulting in about 0.02% of the original number of coefficients. The subdomains are indicated with the white lines.



(a) Structured decomposition resulting in 530 subdomains with piecewise constant representation.

(b) Unstructured decomposition leading to 480 subdomains with piecewise constant representation.

Figure 3.8: Piecewise constant decomposition of the model pictured Figure 3.8 from structured and unstructured decision resulting in about 0.08% of the original number of coefficients. The subdomains are indicated with the white lines.

Due to the inherent lack of flexibility in the structured approach, it is not able to capture the variation accurately. However, the unstructured decomposition provides interesting structures in accordance with the original ones. We have drastically reduced the number of coefficients and yet the structures are interestingly designed (benefiting from the large upper layer of water where the velocity is constant).

The unstructured partition is naturally better suited for model representation, allowing profound flexibility. However, in the context of inverse problem, there is no indication on how to proceed to the decision. Indeed the initial subsurface guess does not usually have any information and hence it is impossible to decide on the appropriate subdomains. In the context of iterative minimization (see Chapter 4), we employ the gradient of the cost function to decide on the partition. Yet the structured decision remains relevant in practice.

3.4.2 Piecewise linear extension

The stability results we have presented necessitates a piecewise constant representation of the velocity. However, in the case where Cauchy data are acquired instead of the DtoN map, the results hold for piecewise linear decomposition, see Chapter 7. We simply illustrate the extension of the decomposition to this case. Following the domain partition, the velocity is represented with a linear function over each subdomain. The velocity is written as

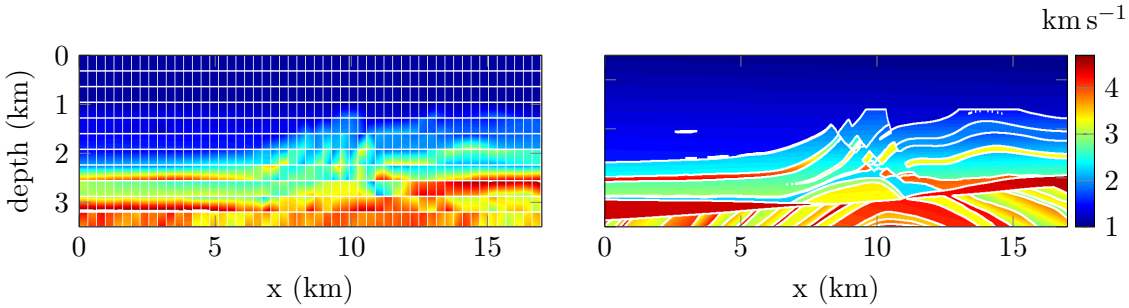
$$c^{-2}(\mathbf{x}) = \sum_{k=1}^{N} \mathcal{H}_k(\mathbf{x}) \chi_{\mathcal{D}_k}(x).$$

Taking the space variable $\mathbf{x} = \{x_1, \ldots, x_r\}$ in dimension r, $\mathcal{H}_k$ is given by

$$\mathcal{H}_k(\mathbf{x}) = \alpha_{k,0} + \sum_{j=1}^{r} \alpha_{k,j} x_j.$$

The total number of coefficients to represent the model is $(r+1)N$, i.e. $4N$ in three dimensions and $3N$ in two dimensions.

The structured and unstructured decisions are applied similarly, now imposing linear functions on the resulting subdomains. In Figure 3.9 we illustrate the structured and unstructured piecewise linear decompositions where the total number of coefficient is about 0.25% of the original model given in Figure 3.6. In Figure 3.10 we compare piecewise linear and piecewise constant unstructured decomposition resulting in the similar number of coefficients.

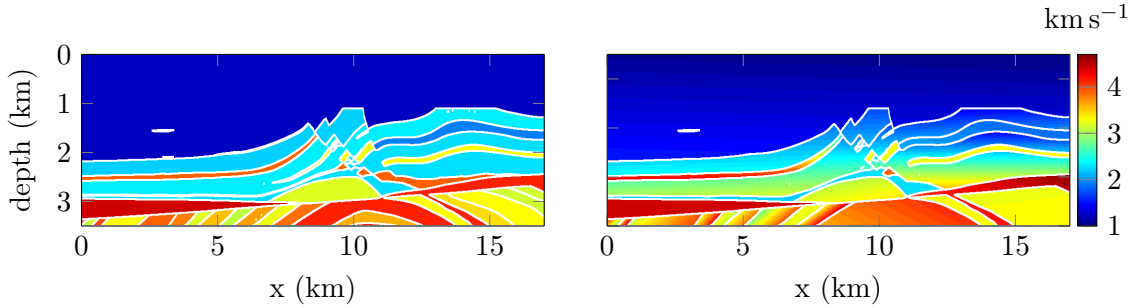


(a) Structured decomposition resulting in 530 subdomains with piecewise linear representation leading to $3 \times 530 = 1590$ coefficients.

(b) Unstructured decomposition leading to 470 subdomains with piecewise linear representation leading to $3 \times 470 = 1410$ coefficients.

Figure 3.9: Piecewise linear decomposition of the model pictured Figure 3.8 from structured and unstructured decision resulting in about 0.25% of the original number of coefficients. The subdomains are indicated with the white lines.

We have illustrated the procedure for our model decomposition, which relies on piecewise constant or piecewise linear representations. It can be seen as relatively simple but it is



(a) Unstructured decomposition leading to 359 subdomains with piecewise constant representation.

(b) Unstructured decomposition leading to 357 subdomains with piecewise linear representation leading to $3 \times 530 = 1590$ coefficients.

Figure 3.10: Comparison of piecewise linear and piecewise constant decompositions of the model pictured Figure 3.8 using unstructured decision resulting in about 0.06% of the original number of coefficients. The subdomains are indicated with the white lines.

entirely motivated by the stability results. Numerically speaking, many possibilities can be used to decompose a model in terms of wavelets (e.g., Mallat (2008)) and such techniques have been employed, for example, in seismic tomography for the purpose of regularization by Loris et al. (2007, 2010). For full waveform inversion, we mention the work of Yuan and Simons (2014); Yuan et al. (2015).

Remark 3.3. *In our applications, the relation between this model representation and the way the model is taken into account for the resolution of the forward problem (numerical discretization) is relatively simple. Indeed, the generated subdomains actually consist in a group of mesh cells, and the parameter value is constant in each mesh cell. Because our technique is employed to coarsen the representation, this approach is sufficient, see Remark 7.1. For higher flexibility, we can mention the techniques where the model is allowed to vary inside each mesh cell. It could certainly be used to obtain an higher correspondence between the model representation and its numerical discretization.*

3.5 Conclusion

In this chapter, we have studied the stability of the inverse problem associated with the Helmholtz equation. The conditional Lipschitz type stability is obtained when considering a piecewise constant representation of the model, and we have given analytical lower and upper bounds for the underlying stability constant. The main result is delivered in the Appendix A. It is obtained in the case where the full DtoN map is taken as the data, but we demonstrate its validity in the geophysical framework with partial data. Motivated by the stability analysis, we deploy a pairing between frequency and scale to apply for

numerical reconstruction. The low frequencies suffer from the increase of the stability constant, and low scale should be employed. On the other hand, higher frequencies, by reducing the stability constant, allow the scale refinement and consequently improve the resolution. The piecewise constant model partition is of importance, and for this purpose, we have established structured and unstructured decisions. Unstructured partitions, by their flexibility, provides much more contrast. However, the initial model can experience limited a priori information, and the structured partition is not to discard.

Appendix A

Article: Inverse boundary value problem for the Helmholtz equation: quantitative conditional Lipschitz stability estimates

E. Beretta, M. V. de Hoop, F. Faucher and O. Scherzer.

published in *SIAM Journal on Mathematical Analysis*, 48 (6): 3962–3983, 2016.

Key words. Inverse problems, Helmholtz equation, stability and convergence of numerical methods.
AMS subject classifications. 35R30, 86A22, 65N12, 35J25.

Abstract. We study the inverse boundary value problem for the Helmholtz equation using the Dirichlet-to-Neumann map at selected frequencies as the data. A conditional Lipschitz stability estimate for the inverse problem holds in the case of wavespeeds that are a linear combination of piecewise constant functions (following a domain partition) and gives a framework in which the scheme converges. The stability constant grows exponentially as the number of subdomains in the domain partition increases. We establish an order optimal upper bound for the stability constant. We eventually realize computational experiments to demonstrate the stability constant evolution for three-dimensional wavespeed reconstruction.

A.1 Introduction

In this paper we study the inverse boundary value problem for the Helmholtz equation using the Dirichlet-to-Neumann map at selected frequencies as the data. This inverse problem arises, for example, in reflection seismology and inverse obstacle scattering problems for electromagnetic waves Bao and Li (2005b); Symes (2009); Bao and Triki (2010). We consider wave speeds containing discontinuities.

Uniqueness of the mentioned inverse boundary value problem was established by Sylvester and Uhlmann (1987) assuming that the wave speed is a bounded measurable function. This inverse problem has been extensively studied from an optimization point of view. We mention, in particular, the work of Ben-Hadj-Ali et al. (2008).

It is well known that the logarithmic character of stability of the inverse boundary value problem for the Helmholtz equation Alessandrini (1988); Novikov (2011) cannot be avoided, see also Hähner and Hohage (2001); Hohage (1997). In fact, Mandache (2001) proved that despite of regularity a priori assumptions of any order on the unknown wavespeed, logarithmic stability is the best possible. However, conditional Lipschitz stability estimates can be obtained: accounting for discontinuities, such an estimate holds if the unknown wavespeed is a finite linear combination of piecewise constant functions with an underlying known domain partitioning Beretta et al. (2013a). It was obtained following an approach introduced by Alessandrini and Vessella (2005) and further developed by Beretta and Francini (2011) for Electrical Impedance Tomography (EIT) based on the use of singular solutions. If, on one hand, this method allows to use partial data, on the other hand it does not allow to find an optimal bound of the stability constant. Here, we revisit the Lipschitz stability estimate for the full Dirichlet-to-Neumann map using complex geometrical optics (CGO) solutions which give rise to a sharp upper bound of the Lipschitz constant in terms of the number of subdomains in the domain partitioning. We develop the estimate in $L^2(\Omega)$.

Unfortunately, the use of CGO's solutions leads naturally to a dependence of the stability constant on frequency of exponential type. This is clearly far from being optimal as it is also pointed out in the paper of Nagayasu et al. (2013). There the authors prove a stability estimate, in terms of Cauchy data instead of the Dirichlet-to-Neumann map using CGO solutions. They derive a stability estimate consisting of two parts: a Lipschitz stability estimate and a Logarithmic stability estimate. When the frequency increases the logarithmic part decreases while the Lipschitz part becomes dominant but with a stability constant which blows up exponentially in frequency.

We can exploit the quantitative stability estimate, via a Fourier transform, in the cor-

responding time domain inverse boundary value problem with bounded frequency data. Datchev and de Hoop (2015) showed how to choose classes of non-smooth coefficient functions, one of which is consistent with the class considered here, so that optimization formulations of inverse wave problems satisfy the prerequisites for application of steepest descent and Newton-type iterative reconstruction methods. The proof is based on resolvent estimates for the Helmholtz equation. Thus, one can allow approximate localization of the data in selected time windows, with size inversely proportional to the maximum allowed frequency. This is of importance to applications in the context of reducing the complexity of field data. We note that no information is lost by cutting out a (short) time window, since the boundary source functions (and wave solutions), being compactly supported in frequency, are analytic with respect to time. We cannot allow arbitrarily high frequencies in the data. This restriction is reflected, also, in the observation by Blazek et al. (2013) that the adjoint equation, which appears in the mentioned iterative methods, does not admit solutions.

As a part of the analysis, we study the Fréchet differentiability of the direct problem and obtain the frequency and domain partitioning dependencies of the relevant constants away from the Dirichlet spectrum. Our results hold for finite fixed frequency data including frequencies arbitrarily close to zero while avoiding Dirichlet eigenfrequencies; in view of the estimates, inherently, there is a finest scale which can be reached. Finally, we estimate the stability numerically and demonstrate the validity of the bounds, in particular in the context of reflection seismology.

A.2 Inverse boundary value problem with the Dirichlet-to-Neumann map as the data

A.2.1 Direct problem and forward operator

We describe the direct problem and some properties of the data, that is, the Dirichlet-to-Neumann map. We will formulate the direct problem as a nonlinear operator mapping F_ω from $L^\infty(\Omega)$ to $\mathcal{L}(H^{1/2}(\partial\Omega), H^{-1/2}(\partial\Omega))$ defined as

$$F_\omega(c^{-2}) = \Lambda_{\omega^2 c^{-2}},$$

where $\Lambda_{\omega^2 c^{-2}}$ indicates the Dirichlet to Neumann operator. Indeed, at fixed frequency ω^2, we consider the boundary value problem,

$$\begin{cases} (-\Delta - \omega^2 c^{-2}(x))u = & 0, \quad \text{in } \Omega, \\ u = & g \quad \text{on } \partial\Omega, \end{cases}$$

while $\Lambda_{\omega^2 c^{-2}} : g \to \frac{\partial u}{\partial \nu}|_{\partial\Omega}$, where ν denotes the outward unit normal vector to $\partial\Omega$. In this section, we will state some known results concerning the well-posedness of problem (A.2.1) (see, for example, Gilbarg and Trudinger (1983)) and regularity properties of the nonlinear map F_ω. We will sketch the proofs of these results because we will need to keep track of the dependencies of the constants involved on frequency. We invoke

Assumption A.1. *There exist two positive constants B_1, B_2 such that*

$$B_1 \leq c^{-2} \leq B_2 \quad \text{in } \Omega.$$

In the sequel of Section 2 $C = C(a, b, c, \dots)$ indicates that C depends only on the parameters $a, b, c, \dots$ and we will indicate different constants with the same letter C.

Proposition A.1. *Let Ω be a bounded Lipschitz domain in $\mathbb{R}^3$, $f \in L^2(\Omega)$, $g \in H^{1/2}(\partial\Omega)$ and $c^{-2} \in L^\infty(\Omega)$ satisfying Assumption A.1. Then, there exists a discrete set $\Sigma_{c^{-2}} := \{\tilde{\lambda}_n \quad |\tilde{\lambda}_n > 0, \ \forall n \in \mathbb{N}\}$ such that, for every $\omega^2 \in \mathbb{C}\backslash\Sigma_{c^{-2}}$, there exists a unique solution $u \in H^1(\Omega)$ of*

$$\begin{cases} (-\Delta - \omega^2 c^{-2}(x))u = & f \quad \text{in } \Omega, \\ u = & g \quad \text{on } \partial\Omega. \end{cases}$$

Furthermore, there exists a positive constant C such that

$$\|u\|_{H^1(\Omega)} \leq C \left(1 + \frac{\omega^2}{d(\omega^2, \Sigma_{c^{-2}})}\right) \left(\|g\|_{H^{1/2}(\partial\Omega)} + \|f\|_{L^2(\Omega)}\right), \tag{A.1}$$

where $C = C(\Omega, B_2)$ and $d(\omega^2, \Sigma_{c^{-2}})$ indicates the distance of ω^2 from $\Sigma_{c^{-2}}$.

Proof. We first prove the result for $g = 0$. Consider the linear operators $-\Delta : H_0^1(\Omega) \to H^{-1}(\Omega)$ and the multiplication operator

$$\begin{aligned} M_{c^{-2}} : L^2(\Omega) &\to L^2(\Omega), \\ u &\to c^{-2}u \end{aligned} \tag{A.2}$$

respectively. We can now consider the operator $K = \Delta^{-1} M_{c^{-2}} : H_0^1(\Omega) \to H_0^1(\Omega)$. The equation

$$(-\Delta - \omega^2 c^{-2}(x))u = f \ .$$

for $u \in H_0^1(\Omega)$ is equivalent to

$$(I - \omega^2 K)u = \Delta^{-1} f \tag{A.3}$$

Note that $K : H_0^1(\Omega) \to H_0^1(\Omega)$ is compact by Rellich–Kondrachov compactness theorem. Furthermore, by Assumption A.1 and the properties of Δ^{-1} it follows that K is self-adjoint and positive. Hence, K has a discrete set of positive eigenvalues $\{\alpha_n\}_{n\in\mathbb{N}}$ such that $\alpha_n \to 0$ as $n \to \infty$. Let $\tilde{\lambda}_n := \frac{1}{\alpha_n}, n \in \mathbb{N}$ and define $\Sigma_{c^{-2}} := \{\tilde{\lambda}_n \ : \ n \in \mathbb{N}\}$ and let $\omega^2 \in \mathbb{C}\backslash\Sigma_{c^{-2}}$, and show that it satisfies the assumptions of this proposition. Then, by the Fredholm alternative, there exists a unique solution $u \in H_0^1(\Omega)$ of (A.3).

To prove estimate (A.1) we observe that

$$u = \sum_{n=1}^{\infty} \langle u, e_n \rangle e_n, \quad Ku = \sum_{n=1}^{\infty} \alpha_n \langle u, e_n \rangle e_n$$

where $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal basis of $L^2(\Omega)$. Hence we can rewrite (A.3) in the form

$$\sum_{n=1}^{\infty} (1 - \omega^2 \alpha_n)\langle u, e_n \rangle e_n = \sum_{n=1}^{\infty} \langle h, e_n \rangle e_n \text{ where } h = \Delta^{-1} f$$

Hence,

$$\langle u, e_n \rangle = \frac{1}{1 - \frac{\omega^2}{\tilde{\lambda}_n}} \langle h, e_n \rangle, \quad \forall n \in \mathbb{N}$$

and

$$u = \sum_{n=1}^{\infty} \frac{1}{1 - \frac{\omega^2}{\tilde{\lambda}_n}} \langle h, e_n \rangle e_n$$

so that

$$\|u\|_{L^2(\Omega)} \leq \left(1 + \frac{\omega^2}{d(\omega^2, \Sigma_{c^{-2}})}\right) \|h\|_{L^2(\Omega)} \leq C \left(1 + \frac{\omega^2}{d(\omega^2, \Sigma_{c^{-2}})}\right) \|f\|_{L^2(\Omega)}$$

where $C = C(\Omega, B_2)$.

Now, by multiplying equation (A.1) with u, integrating by parts, using Schwartz' inequality, Assumptions (A.1) and (A.2.1) it follows in the case $g = 0$:

$$\|\nabla u\|_{L^2(\Omega)} \leq C \left(1 + \frac{\omega^2}{d(\omega^2, \Sigma_{c^{-2}})}\right) \|f\|_{L^2(\Omega)}$$

Hence, by (A.2.1) and (A.2.1) we finally get

$$\|u\|_{H^1(\Omega)} \leq C \left(1 + \frac{\omega^2}{d(\omega^2, \Sigma_{c^{-2}})}\right) \|f\|_{L^2(\Omega)}.$$

If g is not identically zero then we reduce the problem to the previous case by considering $v = u - \tilde{g}$ where $\tilde{g} \in H^1(\Omega)$ is such that $\tilde{g} = g$ on $\partial\Omega$ and $\|\tilde{g}\|_{H^1(\Omega)} \leq \|g\|_{H^{1/2}(\partial\Omega)}$ and we

derive easily the estimate

$$\|u\|_{H^1(\Omega)} \leq C \left(1 + \frac{\omega^2}{d(\omega^2, \Sigma_{c^{-2}})}\right) (\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)})$$

which concludes the proof. □

The constants appearing in the estimate of Proposition A.1 depends on c^{-2} and $\Sigma_{c^{-2}}$ which are unknown. To our purposes it would be convenient to have constants depending only on a priori parameters B_1, B_2 and other known parameters. Let us denote by Σ_0 the spectrum of $-\Delta$. Then, we have the following

Proposition A.2. *Suppose that the assumptions of Proposition A.1 are satisfied. Let* $\{\lambda_n\}_{n\in\mathbb{N}}$ *denote the Dirichlet eigenvalues of* $-\Delta$. *Then, for any* $n \in \mathbb{N}$,

$$\frac{\lambda_n}{B_2} \leq \tilde{\lambda}_n \leq \frac{\lambda_n}{B_1}.$$

If ω^2 *is such that,*

$$0 < \omega^2 < \frac{\lambda_1}{B_2}, \tag{A.4}$$

or, for some $n \geq 1$,

$$\frac{\lambda_n}{B_1} < \omega^2 < \frac{\lambda_{n+1}}{B_2}, \tag{A.5}$$

then there exists a unique solution $u \in H^1(\Omega)$ *of Problem A.2.1 and the following estimate holds*

$$\|u\|_{H^1(\Omega)} \leq C \left(\|g\|_{H^{1/2}(\partial\Omega)} + \|f\|_{L^2(\Omega)}\right),$$

where $C = C(B_1, B_2, \omega^2, \Sigma_0)$.

Proof. To derive estimate (A.2) we consider the Rayleigh quotient related to equation (A.2.1)

$$\frac{\int_\Omega |\nabla v|^2}{\int_\Omega c^{-2} v^2}.$$

By Assumption A.1, for any non trivial $v \in H_0^1(\Omega)$ we have

$$\frac{1}{B_2} \frac{\int_\Omega |\nabla v|^2}{\int_\Omega v^2} \leq \frac{\int_\Omega |\nabla v|^2}{\int_\Omega c^{-2} v^2} \leq \frac{1}{B_1} \frac{\int_\Omega |\nabla v|^2}{\int_\Omega v^2}.$$

Now, we apply Courant-Rayleigh minimax principle (see for instance (Davies, 1995, Theorem 4.5.1), where the infinite dimensional Courant-Rayleigh minimax principle has been considered): The following arguments are similar as in the simple one-dimensional Example

of Davies' book (Davies, 1995, Example 4.6.1). Due to Assumption A.1 the Hilbert space

$$L^2_c(\Omega) = \{v : \int_\Omega c^{-2}v^2 < \infty\},$$

with norm $\|v\|_{L^2_c} = \int_\Omega v^2 c^{-2}$ is equivalent to $L^2(\Omega)$.

$$\tilde{\lambda}_n := \inf_{\{\tilde{u}_1,\cdots,\tilde{u}_n \in H^1_0(\Omega)\}} \sup_{v\in \text{span}\{\tilde{u}_1,\cdots,\tilde{u}_n\}:\|v\|_{L^2_c}\leq 1} \frac{\int_\Omega |\nabla v|^2}{\int_\Omega c^{-2}v^2},$$

$$\lambda_n := \inf_{\{u_1,\cdots,u_n \in H^1_0(\Omega)\}} \sup_{v\in \text{span}\{u_1,\cdots,u_n\}:\|v\|_{L^2}\leq 1} \frac{\int_\Omega |\nabla v|^2}{\int_\Omega v^2}.$$

Note that $\|v\|_{L^2} \leq 1$ implies that $\|v\|^2_{L^2_c} \leq B_2$ and that $L^2_c(\Omega) = L^2(\Omega)$. Therefore

$$\lambda_n \leq \inf_{\{u_1,\cdots,u_n \in H^1_0(\Omega)\}} \sup_{v\in \text{span}\{u_1,\cdots,u_n\}:\|v\|^2_{L^2_c}\leq B_2} \frac{\int_\Omega |\nabla v|^2}{\int_\Omega v^2}.$$

Now, using the scale invariance of $\frac{\int_\Omega |\nabla v|^2}{\int_\Omega v^2}$ and that $c^{-2} \leq B_2$, we get

$$\lambda_n \leq B_2 \inf_{\{u_1,\cdots,u_n \in H^1_0(\Omega)\}} \sup_{v\in \text{span}\{u_1,\cdots,u_n\}:\|v\|_{L^2_c}\leq 1} \frac{\int_\Omega |\nabla v|^2}{\int_\Omega c^{-2}v^2} = B_2\tilde{\lambda}_n.$$

To get lower bound estimate for $\tilde{\lambda}_n$ observe that if $\|v\|_{L^2_c} \leq 1$ then $\|v\|^2_{L^2} \leq \frac{1}{B_1}$. Hence

$$\tilde{\lambda}_n \leq \inf_{\{\tilde{u}_1,\cdots,\tilde{u}_n \in H^1_0(\Omega)\}} \sup_{v\in \text{span}\{\tilde{u}_1,\cdots,\tilde{u}_n\}:\|v\|^2_{L^2}\leq \frac{1}{B_1}} \frac{\int_\Omega |\nabla v|^2}{\int_\Omega c^{-2}v^2}.$$

Now, using the scale invariance of $\frac{\int_\Omega |\nabla v|^2}{\int_\Omega v^2}$ and that $c^{-2} \geq B_1$, we get

$$\tilde{\lambda}_n \leq \inf_{\{\tilde{u}_1,\cdots,\tilde{u}_n \in H^1_0(\Omega)\}} \sup_{v\in \text{span}\{\tilde{u}_1,\cdots,\tilde{u}_n\}:\|v\|_{L^2}\leq 1} \frac{\int_\Omega |\nabla v|^2}{\int_\Omega c^{-2}v^2} = \frac{1}{B_1}\lambda_n.$$

Thus we have shown that

$$\frac{\lambda_n}{B_2} \leq \tilde{\lambda}_n \leq \frac{\lambda_n}{B_1}, \quad \forall n \in \mathbb{N}.$$

Hence, we have well-posedness of problem (A.2.1) if we select an ω^2 satisfying (A.4) or (A.5) and the claim follows. □

We observe that in order to derive the uniform estimates of Proposition A.2 we need to assume that either the frequency is small (A.4) or that the oscillation of c^{-2} is sufficiently small (A.5). This observation can also been found in Davies (1995).

In the seismic application we have in mind we might know the spectrum of some reference wavespeed c_0^{-2}. The following local result holds

Proposition A.3. *Let Ω and c_0^{-2} satisfy the assumptions of Proposition A.1 and let $\omega^2 \in \mathbb{C}\backslash\Sigma_{c_0^{-2}}$ where $\Sigma_{c_0^{-2}}$ is the Dirichlet spectrum of equation (A.2.1) corresponding to c_0^{-2}. Then, there exists $\delta = \delta(\Omega, \omega^2, B_2, \Sigma_{c_0^{-2}}) > 0$ such that, if*

$$\|c^{-2} - c_0^{-2}\|_{L^\infty(\Omega)} \leq \delta,$$

then $\omega^2 \in \mathbb{C}\backslash\Sigma_{c^{-2}}$ and the solution u of Problem A.1 corresponding to c^{-2} satisfies

$$\|u\|_{H^1(\Omega)} \leq C\left(1 + \frac{\omega^2}{d(\omega^2, \Sigma_{c_0^{-2}})}\right)\left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\right),$$

$C = C(\Omega, B_2)$.

Proof. Let $\delta_c := c^{-2} - c_0^{-2}$ and consider $u_0 \in H^1(\Omega)$ the unique solution of A.1 for c_0^{-2} and consider the problem

$$\begin{cases} -\Delta v - \omega^2 c_0^{-2} v - \omega^2 \delta_c v = & \omega^2 u_0 \delta_c \quad \text{in } \Omega, \\ v = & 0, \quad \text{on } \partial\Omega. \end{cases}$$

Let now

$$L := -\Delta - \omega^2 c_0^{-2}$$

then, by assumption, it is invertible from $H_0^1(\Omega)$ to $L^2(\Omega)$ and we can rewrite problem (A.2.1) in the form

$$(I - K)v = h, \tag{A.6}$$

where $K = \omega^2 L^{-1} M_{\delta_c}$ and M_{δ_c} is the multiplication operator defined in (A.2) and $h = L^{-1}(\omega^2 u_0 \delta_c)$. Observe now that from (A.1) $\|L^{-1}\| \leq C(1 + \frac{\omega^2}{d_0})$ with $C = C(\Omega, B_2)$ and where $d_0 = \text{dist}(\omega^2, \Sigma_{c_0^{-2}})$. Hence, we derive

$$\|K\| \leq \omega^2 \|L^{-1}\| \|M_{\delta_c}\| \leq \omega^2 \|L^{-1}\| \delta \leq C\omega^2 (1 + \frac{\omega^2}{d_0})\delta.$$

Hence, choosing $\delta = \frac{1}{2}(C\omega^2(1 + \frac{\omega^2}{d_0}))^{-1}$ the bounded operator K has norm smaller than one. Hence, $I - K$ is invertible and there exists a unique solution v of (A.6) in H_0^1 satisfying (A.1) with $C = C(B_2, \omega^2, \Omega, d_0)$ and since $u = u_0 + v$ the statement follows. □

Let ω^2 be such that either

$$0 < \omega^2 < \frac{\lambda_1}{B_2},$$

or for some $n \geq 1$

$$\frac{\lambda_n}{B_1} < \omega^2 < \frac{\lambda_{n+1}}{B_2},$$

and let

$$\mathcal{W} := \{c^{-2} \in L^\infty(\Omega) : B_1 \leq c^{-2} \leq B_2\} .$$

Then the direct operator

$$\begin{aligned} F_\omega : \mathcal{W} &\to \mathcal{L}(H^{1/2}(\partial\Omega), H^{-1/2}(\partial\Omega)), \\ c^{-2} &\mapsto \Lambda_{\omega^2 c^{-2}}, \end{aligned}$$

is well defined.

We will examine regularity properties of F_ω in the following lemmas. We will show the Fréchet differentiability of it.

Lemma A.1 (Fréchet differentiability)**.** *Let $c^{-2} \in L^\infty(\Omega)$ satisfy Assumption (A.1). Assume that $\omega^2 \in \mathbb{C}\backslash\Sigma_{c^{-2}}$. Then, the direct operator F_ω is Fréchet differentiable at c^{-2} and its Fréchet derivative $DF_\omega(c^{-2})$ satisfies*

$$\|DF_\omega[c^{-2}]\|_{\mathcal{L}(L^\infty(\Omega),\mathcal{L}(H^{1/2}(\partial\Omega),H^{-1/2}(\partial\Omega)))} \leq C\omega^2 \left(1 + \frac{\omega^2}{d(\omega^2, \Sigma_{c^{-2}})}\right)^2$$

where $C = C(\Omega, B_2)$.

Proof. Consider $c^{-2} + \delta c^{-2}$. Then, from Proposition A.3, if $\|\delta c^{-2}\|_{L^\infty}(\Omega)$ is small enough, $\omega^2 \notin \Sigma_{c^{-2}+\delta c^{-2}}$. An application of Alessandrini's identity then gives

$$\langle (\Lambda_{\omega^2(c^{-2}+\delta c^{-2})} - \Lambda_{\omega^2 c^{-2}})g \, , \, h\rangle = \omega^2 \int_\Omega \delta c^{-2} \, uv \, \mathrm{d}x,$$

where where $\langle \cdot, \cdot \rangle$ is the dual pairing with respect to $H^{-1/2}(\partial\Omega)$ and $H^{1/2}(\partial\Omega)$ and u and v solve the boundary value problems,

$$\begin{cases} (-\Delta - \omega^2(c^{-2} + \delta c^{-2}))u = & 0, \quad x \in \Omega, \\ u = & g, \quad x \in \partial\Omega, \end{cases}$$

and

$$\begin{cases} (-\Delta - \omega^2 c^{-2})v = & 0, \quad x \in \Omega, \\ v = & h, \quad x \in \partial\Omega, \end{cases}$$

respectively. We first show that the map F_ω is Fréchet differentiable and that the Fréchet derivative is given by

$$\langle DF_\omega[c^{-2}](\delta c^{-2})g \, , \, h\rangle = \omega^2 \int_\Omega \delta c^{-2} \, \tilde{u}v \, \mathrm{d}x,$$

where $\tilde{u}$ solves the equation

$$\begin{cases} (-\Delta - \omega^2 c^{-2})\tilde{u} = & 0, \quad x \in \Omega, \\ \tilde{u} = & g, \quad x \in \partial\Omega. \end{cases}$$

In fact, by (A.2.1), we have that

$$\langle (\Lambda_{\omega^2(c^{-2}+\delta c^{-2})} - \Lambda_{\omega^2 c^{-2}})g\,,\, h\rangle - \omega^2 \int_\Omega \delta c^{-2}\,\tilde{u}v\,\mathrm{d}x = \omega^2 \int_\Omega \delta c^{-2}\,(u-\tilde{u})v\,\mathrm{d}x.$$

We note that $u - \tilde{u}$ solves the equations

$$\begin{cases} (-\Delta - \omega^2 c^{-2})(u-\tilde{u}) = & -\omega^2 \delta c^{-2}\,u, & x \in \Omega, \\ u - \tilde{u} = & 0, & x \in \partial\Omega. \end{cases}$$

Using the fact that $u - \tilde{u}$ and v are in $H^1(\Omega)$ and that $\delta c^{-2} \in L^\infty(\Omega)$ and applying Cauchy-Schwarz inequality, we get

$$\left|\omega^2 \int_\Omega \delta c^{-2}\,(u-\tilde{u})v\,\mathrm{d}x\right| \leq \omega^2 \|\delta c^{-2}\|_{L^\infty(\Omega)} \|u-\tilde{u}\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)}.$$

Finally, using the stability estimates of Proposition A.1 applied to $u - \tilde{u}$ and to v and the stability estimates of Proposition A.3 applied to u we derive

$$\left|\omega^2 \int_\Omega \delta c^{-2}\,(u-\tilde{u})v\,\mathrm{d}x\right| \leq C\omega^4 \left(1 + \frac{\omega^2}{\mathrm{d}(\omega^2, \Sigma_{c^{-2}})}\right)^3 \|\delta c^{-2}\|^2_{L^\infty(\Omega)} \|g\|_{H^{1/2}(\partial\Omega)} \|h\|_{H^{1/2}(\partial\Omega)}.$$

Hence

$$\begin{aligned} &\left|\langle (\Lambda_{\omega^2(c^{-2}+\delta c^{-2})} - \Lambda_{\omega^2 c^{-2}})g\,,\, h\rangle - \omega^2 \int_\Omega \delta c^{-2}\,\tilde{u}v\,\mathrm{d}x\right| \\ \leq & C\omega^4 \left(1 + \frac{\omega^2}{\mathrm{d}(\omega^2, \Sigma_{c^{-2}})}\right)^3 \|\delta c^{-2}\|^2_{L^\infty(\Omega)} \|g\|_{H^{1/2}(\partial\Omega)} \|h\|_{H^{1/2}(\partial\Omega)}, \end{aligned}$$

which proves differentiability.
Finally, by

$$\langle DF_\omega[c^{-2}](\delta c^{-2})g\,,\, h\rangle = \omega^2 \int_\Omega \delta c^{-2}\,\tilde{u}v\,\mathrm{d}x,$$

and we get

$$\begin{aligned} \left|\langle DF_\omega[c^{-2}](\delta c^{-2})g\,,\, h\rangle\right| \leq & \omega^2 \|\delta c^{-2}\|_{L^\infty(\Omega)} \|\tilde{u}\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} \\ \leq & \omega^2 \left(1 + \frac{\omega^2}{\mathrm{d}(\omega^2, \Sigma_{c^{-2}})}\right)^2 \|\delta c^{-2}\|_{L^\infty(\Omega)} \|g\|_{H^{1/2}(\partial\Omega)} \|h\|_{H^{1/2}(\partial\Omega)}. \end{aligned}$$

from which (A.1) follows. □

A.2.2 Conditional quantitative Lipschitz stability estimate

Let B_2, r_0, r_1, A, L, N be positive with $N \in \mathbb{N}, N \geq 2$, $r_0 < 1$. In the sequel we will refer to these numbers as to the a priori data. To prove the results of this section we invoke the following common assumptions

Assumption A.2. *$\Omega \subset \mathbb{R}^3$ is a bounded domain such that*

$$|x| \leq Ar_1, \quad \forall x \in \Omega.$$

Moreover,

$$\partial\Omega \text{ of Lipschitz class with constants } r_1 \text{ and } L.$$

Let $\mathcal{D}_N$ be a partition of Ω given by

$$\mathcal{D}_N \triangleq \left\{ \{D_1, D_2, \ldots, D_N\} \mid \bigcup_{j=1}^{N} \overline{D}_j = \Omega \ , \ (D_j \cap D_{j'})^\circ = \emptyset, j \neq j' \right\}$$

such that

$$\{\partial D_j\}_{j=1}^N \text{ is of Lipschitz class with constants } r_0 \text{ and } L.$$

Assumption A.3. *The function $c^{-2} \in \mathcal{W}_N$, that is, it satisfies*

$$B_1 \leq c^{-2} \leq B_2, \quad \text{in } \Omega$$

and is of the form

$$c^{-2}(x) = \sum_{j=1}^{N} c_j \chi_{D_j}(x),$$

where $c_j, j = 1, \ldots, N$ are unknown numbers and $\{D_1, \ldots, D_N\} \in \mathcal{D}_N$.

Assumption A.4. *Assume*

$$0 < \omega^2 < \frac{\lambda_1}{B_2},$$

or, for some $n \geq 1$,

$$\frac{\lambda_n}{B_1} < \omega^2 < \frac{\lambda_{n+1}}{B_2}.$$

Under the above assumptions we can state the following preliminary result

Lemma A.2. *Let Ω and $\mathcal{D}_N$ satisfy Assumption (A.2) and let $c^{-2} \in \mathcal{W}_N$. Then, for every $s' \in (0, 1/2)$, there exists a positive constant C with $C = C(L, s')$ such that*

$$\|c^{-2}\|_{H^{s'}(\Omega)} \leq C(L, s') \frac{1}{r_0^{s'}} \|c^{-2}\|_{L^2(\Omega)}.$$

Proof. The proof is based on the extension of a result of Magnanini and Papi (1985) to the three-dimensional setting. In fact, following the argument in Magnanini and Papi (1985), one has that

$$\|\chi_{D_j}\|^2_{H^{s'}(\Omega)} \leq \frac{16\pi}{(1-2s')(2s')^{1+2s'}} |D_j|^{1-2s'} |\partial D_j|^{2s'}. \tag{A.7}$$

We now use the fact that $\{D_j\}_{j=1}^N$ is a partition of disjoint sets of Ω to show the following inequality

$$\|c^{-2}\|^2_{H^{s'}(\Omega)} \leq 2 \sum_{j=1}^N c_j^2 \|\chi_{D_j}\|^2_{H^{s'}(\Omega)} \tag{A.8}$$

In fact, in order to prove (A.8) recall that

$$\|c^{-2}\|^2_{H^{s'}(\Omega)} = \int_\Omega \int_\Omega \frac{|\sum_{j=1}^N c_j(\chi_{D_j}(x) - \chi_{D_j}(y))|^2}{|x-y|^{3+2s'}} \, dx \, dy$$

and observe that, since the $\{D_j\}_{j=1}^N$ is a partition of disjoint sets of Ω, we get

$$|\sum_{j=1}^N c_j(\chi_{D_j}(x) - \chi_{D_j}(y))|^2 = \sum_{j=1}^N c_j^2(\chi_{D_j}(x) - \chi_{D_j}(y))^2 - \sum_{i \neq j} c_i c_j \chi_{D_i}(x) \chi_{D_j}(y)$$

Again, by the fact that the $\{D_j\}_{j=1}^N$ are disjoint sets, we have

$$\begin{aligned}
& \sum_{i \neq j} |c_i c_j| \chi_{D_i}(x) \chi_{D_j}(y) \leq \sum_{i \neq j} \frac{c_i^2 + c_j^2}{2} \chi_{D_i}(x) \chi_{D_j}(y) \\
= & \sum_{i \neq j} \frac{c_i^2}{2} (\chi_{D_i}(x) - \chi_{D_i}(y))^2 \chi_{D_i}(x) \chi_{D_j}(y) + \sum_{i \neq j} \frac{c_j^2}{2} (\chi_{D_j}(x) - \chi_{D_j}(y))^2 \chi_{D_i}(x) \chi_{D_j}(y) \\
\leq & \sum_{i \neq j} \frac{c_i^2}{2} (\chi_{D_i}(x) - \chi_{D_i}(y))^2 \chi_{D_j}(y) + \sum_{i \neq j} \frac{c_j^2}{2} (\chi_{D_j}(x) - \chi_{D_j}(y))^2 \chi_{D_i}(x) \\
\leq & \sum_{i=1}^N \frac{c_i^2}{2} (\chi_{D_i}(x) - \chi_{D_i}(y))^2 \sum_{j=1}^N \chi_{D_j}(y) + \sum_{j=1}^N \frac{c_j^2}{2} (\chi_{D_j}(x) - \chi_{D_j}(y))^2 \sum_{i=1}^N \chi_{D_i}(y) \\
\leq & \sum_{i=1}^N \frac{c_i^2}{2} (\chi_{D_i}(x) - \chi_{D_i}(y))^2 + \sum_{j=1}^N \frac{c_j^2}{2} (\chi_{D_j}(x) - \chi_{D_j}(y))^2 \\
= & \sum_{i=1}^N c_i^2 (\chi_{D_i}(x) - \chi_{D_i}(y))^2
\end{aligned}$$

where we have used the fact that $\sum_{i=1}^N \chi_{D_i} \leq 1$. So, we have derived that

$$|\sum_{j=1}^N c_j(\chi_{D_j}(x) - \chi_{D_j}(y))|^2 \leq 2\sum_{j=1}^N c_j^2(\chi_{D_j}(x) - \chi_{D_j}(y))^2$$

from which it follows that

$$\begin{aligned}
\|c^{-2}\|^2_{H^{s'}(\Omega)} &= \int_\Omega\int_\Omega \frac{|\sum_{j=1}^N c_j(\chi_{D_j}(x) - \chi_{D_j}(y))|^2}{|x-y|^{3+2s'}}\,dx\,dy \\
&\leq 2\int_\Omega\int_\Omega \frac{\sum_{j=1}^N c_j^2(\chi_{D_j}(x) - \chi_{D_j}(y))^2}{|x-y|^{3+2s'}}\,dx\,dy \\
&\leq 2\sum_{j=1}^N c_j^2 \int_\Omega\int_\Omega \frac{(\chi_{D_j}(x) - \chi_{D_j}(y))^2}{|x-y|^{3+2s'}}\,dx\,dy = 2\sum_{j=1}^N c_j^2\|\chi_{D_j}\|^2_{H^{s'}(\Omega)}
\end{aligned}$$

which proves (A.8). so that finally from (A.8), (A.7) and from Assumption (A.2) we get

$$\|c^{-2}\|^2_{H^{s'}(\Omega)} \leq 2\sum_{j=1}^N c_j^2\|\chi_{D_j}\|^2_{H^{s'}(\Omega)} \leq C(s')\sum_{j=1}^N c_j^2|D_j|\left(\frac{|\partial D_j|}{|D_j|}\right)^{2s'} \leq \frac{C(L,s')}{r_0^{2s'}}\|c^{-2}\|^2_{L^2(\Omega)}.$$

□

We are now ready to state and prove our main stability result

Proposition A.4. *Assume (A.2) and let $c_1^{-1}, c_2^{-1} \in \mathcal{W}_N$ and let ω^2 satisfy Assumption A.4. Then, there exists a positive constant K, depending on A, r_1, L, such that,*

$$\|c_1^{-2} - c_2^{-2}\|_{L^2(\Omega)} \leq \frac{1}{\omega^2}\, e^{K(1+\omega^2 B_2)(|\Omega|/r_0^3)^{\frac{4}{7}}}\ \|\Lambda_{\omega^2 c_1^{-2}} - \Lambda_{\omega^2 c_2^{-2}}\|_{\mathcal{L}(H^{1/2}(\partial\Omega), H^{-1/2}(\partial\Omega))}.$$

Proof. To prove our stability estimate we follow the idea of Alessandrini of using CGO solutions but we use slightly different ones than those introduced in Sylvester and Uhlmann (1987) and in Alessandrini (1988) to obtain better constants in the stability estimates as proposed by Feldman et al. (2015). We also use the estimates proposed in Feldman et al. (2015) (see Theorem 4.4) and due to Hähner (1996) concerning the case of bounded potentials.

In fact, by Theorem 4.3 of Feldman et al. (2015), since $c^{-2} \in L^\infty(\Omega)$, $\|c^{-2}\|_{L^\infty(\Omega)} \leq B_2$, there exists a positive constant $C = C(\omega^2, B_2, A, r_1)$ such that for every $\zeta \in \mathbb{C}^3$ satisfying $\zeta \cdot \zeta = 0$ and $|\zeta| \geq C$ the equation

$$-\Delta u - \omega^2 c^{-2} u = 0$$

has a solution of the form

$$u(x) = e^{ix\cdot\zeta}(1 + R(x))$$

where $R \in H^1(\Omega)$ satisfies

$$\|R\|_{L^2(\Omega)} \leq \frac{C}{|\zeta|}, \quad \|\nabla R\|_{L^2(\Omega)} \leq C.$$

Let $\xi \in \mathbb{R}^3$ and let $\tilde{\omega}_1$ and $\tilde{\omega}_2$ be unit vectors of $\mathbb{R}^3$ such that $\{\tilde{\omega}_1, \tilde{\omega}_2, \xi\}$ is an orthogonal set of vectors of $\mathbb{R}^3$. Let s be a positive parameter to be chosen later and set for $k = 1, 2$,

$$\zeta_k = \begin{cases} (-1)^{k-1}\frac{s}{\sqrt{2}}(\sqrt{(1 - \frac{|\xi|^2}{2s^2})}\tilde{\omega}_1 + (-1)^{k-1}\frac{1}{\sqrt{2}s}\xi + \mathrm{i}\tilde{\omega}_2) & \text{for } \frac{|\xi|}{\sqrt{2}s} < 1, \\ (-1)^{k-1}\frac{s}{\sqrt{2}}((-1)^{k-1}\frac{1}{\sqrt{2}s}\xi + \mathrm{i}(\sqrt{(\frac{|\xi|^2}{2s^2} - 1)}\tilde{\omega}_1 + \tilde{\omega}_2)) & \text{for } \frac{|\xi|}{\sqrt{2}s} \geq 1. \end{cases}$$

Then an straightforward computation gives

$$\zeta_k \cdot \zeta_k = 0$$

for $k = 1, 2$ and

$$\zeta_1 + \zeta_2 = \xi.$$

Furthermore, for $k = 1, 2$,

$$|\zeta_k| = \begin{cases} s & \text{for } \frac{|\xi|}{\sqrt{2}s} < 1, \\ \frac{|\xi|}{\sqrt{2}} & \text{for } \frac{|\xi|}{\sqrt{2}s} \geq 1. \end{cases}$$

Hence,

$$|\zeta_k| = \max\{s, \frac{|\xi|}{\sqrt{2}}\}.$$

Then, by Theorem 4.3 of Feldman et al. (2015), for $|\zeta_1|, |\zeta_2| \geq C_1 = \max\{C_0\omega^2 B_2, 1\}$, with $C_0 = C_0(A, r_1)$, there exist u_1, u_2, solutions to $-\Delta u_k - \omega^2 c_k^{-2} u_k = 0$ for $k = 1, 2$ respectively, of the form

$$u_1(x) = e^{\mathrm{i}x\cdot\zeta_1}(1 + R_1(x)), \quad u_2(x) = e^{\mathrm{i}x\cdot\zeta_2}(1 + R_2(x))$$

with

$$\|R_k\|_{L^2(\Omega)} \leq \frac{C_0\sqrt{|\Omega|}}{s}\omega^2 B_2$$

and

$$\|\nabla R_k\|_{L^2(\Omega)} \leq C_0\sqrt{|\Omega|}\omega^2 B_2$$

for $k = 1, 2$. It is common in the literature to use estimates which contain $\sqrt{|\Omega|}$; Different estimates in terms of $|\Omega|$ are possible and just change the leading constant C_0.

Consider again Alessandrini's identity

$$\int_\Omega \omega^2(c_1^{-2}-c_2^{-2})u_1u_2\mathrm{d}x = \langle(\Lambda_1-\Lambda_2)u_1|_{\partial\Omega}, u_2|_{\partial\Omega}\rangle,$$

where $u_k \in H^1(\Omega)$ is any solution of $-\Delta u_k - \omega^2 c_k^{-2}u_k = 0$ and $\Lambda_k = \Lambda_{\omega^2 c_k^{-2}}$ for $k=1,2$. Inserting the solutions (A.2.2) in Alessandrini's identity we derive

$$\begin{aligned}
&\left|\int_\Omega \omega^2(c_1^{-2}-c_2^{-2})e^{\mathrm{i}\xi\cdot x}\mathrm{d}x\right| \\
\leq &\|\Lambda_1-\Lambda_2\|\|u_1\|_{H^{1/2}(\partial\Omega)}\|u_2\|_{H^{1/2}(\partial\Omega)} + \left|\int_\Omega \omega^2(c_1^{-2}-c_2^{-2})e^{\mathrm{i}\xi\cdot x}(R_1+R_2+R_1R_2)\mathrm{d}x\right| \\
\leq &\|\Lambda_1-\Lambda_2\|\|u_1\|_{H^1(\Omega)}\|u_2\|_{H^1(\Omega)} + E(\|R_1\|_{L^2(\Omega)}+\|R_2\|_{L^2(\Omega)}+\|R_1\|_{L^4(\Omega)}\|R_2\|_{L^4(\Omega)}).
\end{aligned}$$

where $E := \|\omega^2(c_1^{-2}-c_2^{-2})\|_{L^2(\Omega)}$. By (A.2.2), (A.2.2), (A.2.2) and since $\Omega \subset B_{2R}(0)$ we have

$$\|u_k\|_{H^1(\Omega)} \leq C\sqrt{|\Omega|}(s+|\xi|)e^{Ar_1(s+|\xi|)}, \quad k=1,2.$$

Let $s \geq C_2$ so that $s+|\xi| \leq e^{Ar_1(s+|\xi|)}$. Then, for $s \geq C_3 = \max(C_1,C_2)$, using (A.2.2), (A.2.2) and the standard interpolation inequality ($\|u\|_{L^4(\Omega)} \leq \|u\|^{3/4}_{L^6(\Omega)}\|u\|^{1/4}_{L^2(\Omega)}$) we get

$$|\omega^2(c_1^{-2}-c_2^{-2})\hat{}(\xi)| \leq C\sqrt{|\Omega|}\left(e^{4Ar_1(s+|\xi|)}\|\Lambda_1-\Lambda_2\| + \frac{\omega^2 B_2 E}{s}\right) \tag{A.9}$$

where the $\omega^2 c_k^{-2}$'s have been extended to all $\mathbb{R}^3$ by zero and $\hat{}$ denotes the Fourier transform. Hence, from (A.9), we get

$$\int_{|\xi|\leq\rho}|\omega^2(c_1^{-2}-c_2^{-2})\hat{}(\xi)|^2\mathrm{d}\xi \leq C|\Omega|\rho^3\left(e^{8Ar_1(s+\rho)}\|\Lambda_1-\Lambda_2\|^2 + \frac{\omega^4B_2^2E^2}{s^2}\right)$$

and hence

$$\begin{aligned}
\|\omega^2(c_1^{-2}-c_2^{-2})\hat{}\|^2_{L^2(\mathbb{R}^3)} \leq C|\Omega|\rho^3&\left(e^{8Ar_1(s+\rho)}\|\Lambda_1-\Lambda_2\|^2 + \frac{\omega^4B_2^2E^2}{s^2}\right) \\
&+ \int_{|\xi|\geq\rho}|\omega^2(c_1^{-2}-c_2^{-2})\hat{}(\xi)|^2\,\mathrm{d}\xi
\end{aligned}$$

where $C = C(A,r_1)$. By (A.2) and (A.8) we have that

$$\|\omega^2(c_1^{-2}-c_2^{-2})\|^2_{H^{s'}(\Omega)} \leq \frac{C}{r_0^{2s'}}E^2,$$

where C depends on L, s' and hence

$$\begin{aligned}
\rho^{2s'}\int_{|\xi|\geq\rho}|\omega^2(c_1^{-2}-c_2^{-2})\hat{}(\xi)|^2\,\mathrm{d}\xi &\leq \int_{|\xi|\geq\rho}|\xi|^{2s'}|\omega^2(c_1^{-2}-c_2^{-2})\hat{}(\xi)|^2\,\mathrm{d}\xi \\
&\leq \int_{\mathbb{R}^3}(1+|\xi|^2)^{s'}|\omega^2(c_1^{-2}-c_2^{-2})\hat{}(\xi)|^2\,\mathrm{d}\xi \leq \frac{C}{r_0^{2s'}}E^2.
\end{aligned}$$

Hence, we get

$$\int_{|\xi|\geq\rho} |\omega^2(c_1^{-2}-c_2^{-2})\hat{}(\xi)|^2\,\mathrm{d}\xi \leq \frac{CE^2}{r_0^{2s'}\rho^{2s'}}$$

for every $s'\in(0,1/2)$. Inserting last bound in (A.2.2) we derive

$$\|\omega^2(c_1^{-2}-c_2^{-2})\hat{}\|^2_{L^2(\mathbb{R}^3)} \leq C\left(\rho^3|\Omega|e^{8Ar_1(s+\rho)}\|\Lambda_1-\Lambda_2\|^2+\rho^3|\Omega|\frac{\omega^4B_2^2E^2}{s^2}+\frac{E^2}{r_0^{2s'}\rho^{2s'}}\right).$$

where $C=C(L,s')$. To make the last two terms in the right-hand side of the inequality of equal size we pick up

$$\sqrt[3]{|\Omega|}\rho=\left(\frac{|\Omega|}{r_0^3}\right)^{\frac{2s'}{3(3+2s')}}\left(\frac{1}{\alpha}\right)^{\frac{1}{3+2s'}}s^{\frac{2}{3+2s'}}$$

with $\alpha=\max\{1,\omega^4B_2^2\}$. Then, by Assumption A.2 and observing that we might assume without loss of generality that $\frac{|\Omega|}{r_0^3}>1$. In fact, if this is not the case we can choose a smaller value of r_0 so that the condition is satisfied.

$$\|\omega^2(c_1^{-2}-c_2^{-2})\|^2_{L^2(\Omega)} \leq CE^2\left(\frac{|\Omega|}{r_0^3}\right)^{\frac{2s'}{3+2s'}}\left(e^{C_4(\frac{|\Omega|}{r_0^3})^{\frac{2s'}{3(3+2s')}}s}\left(\frac{\|\Lambda_1-\Lambda_2\|}{E}\right)^2+\left(\frac{\alpha}{s^2}\right)^{\frac{2s'}{3+2s'}}\right)$$

for $s\geq C_3$ and where C depends on s',L,A,r_1 and C_4 depends on L,A,r_1. We now make the substitution

$$s=\frac{1}{C_4(\frac{|\Omega|}{r_0^3})^{\frac{2s'}{3(3+2s')}}}\left|\log\frac{\|\Lambda_1-\Lambda_2\|}{E}\right|$$

where we assume that

$$\frac{\|\Lambda_1-\Lambda_2\|}{E}<c:=e^{-\bar{C}\max\{1,\omega^2B_2\}(\frac{|\Omega|}{r_0^3})^{\frac{2s'}{3(3+2s')}}}$$

with $\bar{C}=\bar{C}(R)$ in order that the constraint $s\geq C_3$ is satisfied. Under this assumption,

$$\|\omega^2(c_1^{-2}-c_2^{-2})\|_{L^2(\Omega)} \leq C(\sqrt{\alpha})^{\frac{2s'}{3+2s'}}\left(\frac{|\Omega|}{r_0^3}\right)^{\frac{2s'}{3+2s'}\frac{9+10s'}{6(3+2s')}}E\left(\left|\log\frac{\|\Lambda_1-\Lambda_2\|}{E}\right|^{-\frac{2s'}{3+2s'}}\right)$$

where $C=C(L,s',A,r_1)$ and we can rewrite last inequality in the form

$$E\leq C(1+\omega^2B_2)^{\frac{2s'}{3+2s'}}\left(\frac{|\Omega|}{r_0^3}\right)^{\frac{2s'}{3+2s'}\frac{9+10s'}{6(3+2s')}}E\left(\left|\log\frac{\|\Lambda_1-\Lambda_2\|}{E}\right|^{-\frac{2s'}{3+2s'}}\right)$$

which gives

$$E\leq e^{C(1+\omega^2B_2)(\frac{|\Omega|}{r_0^3})^{\frac{9+10s'}{6(3+2s')}}}\|\Lambda_1-\Lambda_2\|$$

where $C = C(L, s', A, r_1)$. On the other hand if

$$\frac{\|\Lambda_1 - \Lambda_2\|}{E} \geq c,$$

then

$$\|\omega^2(c_1^{-2} - c_2^{-2})\|_{L^2(\Omega)} \leq c^{-1}\|\Lambda_1 - \Lambda_2\| \leq e^{\bar{C}(1+\omega^2 B_2)\left(\frac{|\Omega|}{r_0^3}\right)^{\frac{1}{3(3+2s')}}}\|\Lambda_1 - \Lambda_2\|$$

Hence, from (A.2.2) and (A.2.2) and recalling that $s' \in (0, \frac{1}{2})$, we have that

$$E \leq e^{C(1+\omega^2 B_2)(\frac{|\Omega|}{r_0^3})^{\frac{9+10s'}{6(3+2s')}}}\|\Lambda_1 - \Lambda_2\|$$

Choosing $s' = \frac{1}{4}$, we derive

$$\|c_1^{-2} - c_2^{-2}\|_{L^2(\Omega)} \leq \frac{1}{\omega^2} e^{K(1+\omega^2 B_2)(|\Omega|/r_0^3)^{\frac{4}{7}}}\|\Lambda_1 - \Lambda_2\|$$

where $K = K(L, A, r_1, s')$ and the claim follows. □

Remark A.1. *Here we state an L^∞-stability estimate, in contrast to the L^2-stability estimate in Proposition A.4.*

Observing that

$$\frac{1}{\sqrt{|\Omega|}}\|c_1^{-2} - c_2^{-2}\|_{L^2(\Omega)} \leq \|c_1^{-2} - c_2^{-2}\|_{L^\infty(\Omega)} \leq \frac{C}{r_0^{3/2}}\|c_1^{-2} - c_2^{-2}\|_{L^2(\Omega)},$$

where $C = C(L)$, and we immediately get the following stability estimate in the L^∞ norm

$$\|c_1^{-2} - c_2^{-2}\|_{L^\infty(\Omega)} \leq \frac{C}{\omega^2} e^{K(1+\omega^2 B_2)(|\Omega|/r_0^3)^{\frac{4}{7}}}\|\Lambda_1 - \Lambda_2\|$$

with $C = C(L)$.

Remark A.2. *In Beretta et al. (2013a) the following lower bound of the stability constant has been obtained in the case of a uniform polyhedral partition $\mathcal{D}_N$*

$$C_N \geq \frac{1}{4\omega^2} e^{K_1 N^{\frac{1}{5}}}$$

Choose a uniform cubical partition $\mathcal{D}_N$ of Ω of mesh size r_0. Then,

$$|\Omega| = N r_0^3 \tag{A.10}$$

and estimate (A.4) of Proposition A.4 gives

$$C_N = \frac{1}{\omega^2} e^{K(1+\omega^2 B_2)N^{\frac{4}{7}}},$$

which proves a sharp bound on the Lipschitz constant with respect to N when the global DtN map is known. In Beretta et al. (2013a) a Lipschitz stability estimate has been derived in terms of the local DtN map using singular solutions. This type of solutions allows to recover the unknown piecewise constant wavespeeds by determining it on the outer boundary of the domain and then, by propagating the singularity inside the domain, to recover step by step the wavespeed on the interface of all subdomains of the partition. This iterative procedure does not lead to sharp bounds of the Lipschitz constant appearing in the stability estimate. It would be interesting if one can get a better bound of the Lipschitz constant using oscillating solutions.

Remark A.3. *In Lemma A.1 we have seen that F_ω is Fréchet differentiable with Lipschitz derivative DF_ω for which we have derived an upper bound in terms of the apriori data. From the stability estimates we can easily derive the following lower bound*

$$\min_{c^{-2}\in\mathcal{W}_N;\ h\in\mathbb{R}^N,\ \|h\|_{L^\infty(\Omega)}=1} \|DF_\omega[c^{-2}]h\|_* \geq \omega^2 e^{-K(1+\omega^2B_2)(\frac{|\Omega|}{r_0^3})^{4/7}}.$$

where $K = K(L, A, r_1)$ and $\|\cdot\|_*$ indicates the norm in $\mathcal{L}(H^{1/2}(\partial\Omega), H^{-1/2}(\partial\Omega))$ i.e.

$$\|T\|_* = \sup\{|\langle Tg, f\rangle| : g, f \in H^{1/2}(\partial\Omega), \|g\|_{H^{1/2}(\partial\Omega)} = \|f\|_{H^{1/2}(\partial\Omega)} = 1\}$$

In fact, by the injectivity of DF_ω

$$\min_{c^{-2}\in\mathcal{W}_N;\ h\in\mathbb{R}^N,\ \|h\|_{L^\infty(\Omega)}=1} \|DF_\omega[c^{-2}]h\|_* = m_0/2 > 0$$

Then, there exists h_0 satisfying $\|h_0\|_{L^\infty(\Omega)} = 1$ and $c_0^{-2} \in \mathcal{W}_N$ such that

$$\|DF_\omega[c_0^{-2}]h_0\|_* \leq m_0\ .$$

Hence, by the definition of $\|\cdot\|_*$ it follows that

$$\left|\langle DF_\omega[c_0^{-2}](h_0)g, f\rangle\right| = \left|\int_\Omega h_0\tilde{u}_0v_0\right| \leq m_0\|\tilde{u}_0\|_{H^{1/2}(\partial\Omega)}\|v_0\|_{H^{1/2}(\partial\Omega)}$$

where $\tilde{u}_0$ and v_0 are solutions to the equation $(-\Delta - \omega^2c_0^{-2})u = 0$ in Ω with boundary data g and f, respectively. Proceeding like in the proof of the stability result (Proposition A.4) and Remark A.1 we derive that

$$1 = \|h_0^{-2}\|_{L^\infty(\Omega)} \leq \frac{1}{\omega^2}e^{K(1+\omega^2B_2)(\frac{|\Omega|}{r_0^3})^{4/7}}m_0$$

which gives the lower bound (A.3).

A.3 Computational experiments

In this section, we numerically compute the stability constant for the inverse problem associated with the Dirichlet-to-Neumann map. We illustrate the stability behavior and compare it with the analytical bounds derived in Section A.2. The estimates we provide here are obtained from the definition of the stability constant,

$$\|c_1^{-2} - c_2^{-2}\|^2 < \mathcal{C}\|F_\omega(c_1^{-2}) - F_\omega(c_2^{-2})\|^2, \tag{A.11}$$

where $\|c_1^{-2} - c_2^{-2}\|$ denotes the L^2-norm of the functions from the finite dimensional Ansatz space. In particular, we consider here a geophysical example of reconstruction where normal data are collected on the boundary. In this situation c_1 and c_2 are assimilated to two different wavespeeds. Hence the boundary value problem A.2.1 corresponds to the propagation of acoustic wave in the media for a boundary source g using the wavespeeds c_1 and c_2 respectively. In our experiments, Gaussian shaped (spatial) source functions (see Figure A.1) are applied. Then the normal data (measurements of the normal derivative of the field) are acquired on the boundary in order to generate the forward operator. The numerical stability estimates are finally obtained by the knowledge of all quantities of equation (A.11).

In Remark A.2, we have formulated the stability constant depending on the number of cubical partitions N in the model representation (equation (A.10)). This situation is well adapted for numerical applications where the domain is commonly discretized. Hence we want to verify the (exponential) dependence of the stability constant with N.

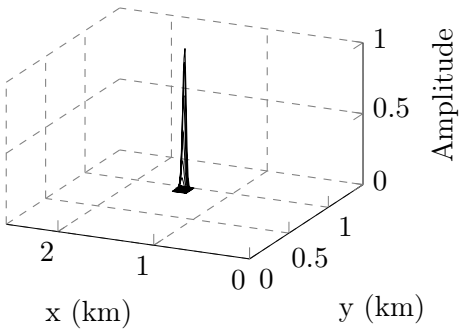


Figure A.1: Illustration of the source shape for a localized boundary source.

The model (assimilated to a wave speed here) is defined on a cubical (structured) domain partition of a rectangular block. With increasing N, the size of the cubes decreases, possibly non uniformly. We use piecewise constant functions on the cubes to define the wave speeds following the main assumption for the Lipschitz stability to hold. Such a partition can be related to Haar wavelets, where N determines the scale. These naturally introduce approximate representations, that is, when the scale of the approximation is coarser than the finest scale contained in the model.

In order to solve the forward problem, the numerical discretization of the operator is realized using discontinuous Galerkin method, where Dirichlet boundary conditions are invoked. The Dirichlet sources at the top boundary introduce Identity block in the discretized Helmholtz operator and give the following linear problem

$$\begin{pmatrix} A_{ii} & A_{i\partial} \\ A_{\partial i} & A_{\partial\partial} \end{pmatrix} \begin{pmatrix} u_i \\ u_\partial \end{pmatrix} = \begin{pmatrix} A_{ii} & A_{i\partial} \\ 0 & Id \end{pmatrix} \begin{pmatrix} u_i \\ u_\partial \end{pmatrix} = \begin{pmatrix} 0 \\ g \end{pmatrix},$$

where A represents the discretized operator, i labels interior points and ∂ labels boundary points, g has values at the source location and is zero elsewhere. This system verifies $u_\partial = u_{|\partial\Omega} = g$ (i.e. Dirichlet boundary condition) and $A_{ii}u_i + A_{i\partial}u_{|\partial\Omega} = 0$. The normal derivative data are generated by taking the normal derivative of the solution wave field u on the surface.

Our experiments use a three-dimensional model of size $2.55 \times 1.45 \times 1.22$km. The wave sepeed c_1 is viewed as a reference model (which is known in this test case) and is represented Figure A.2 (courtesy of Statoil). We also illustrate the different partitions of a model and the notion of approximation. Obviously the larger the number of subdomains is, the more precise will the representation be.

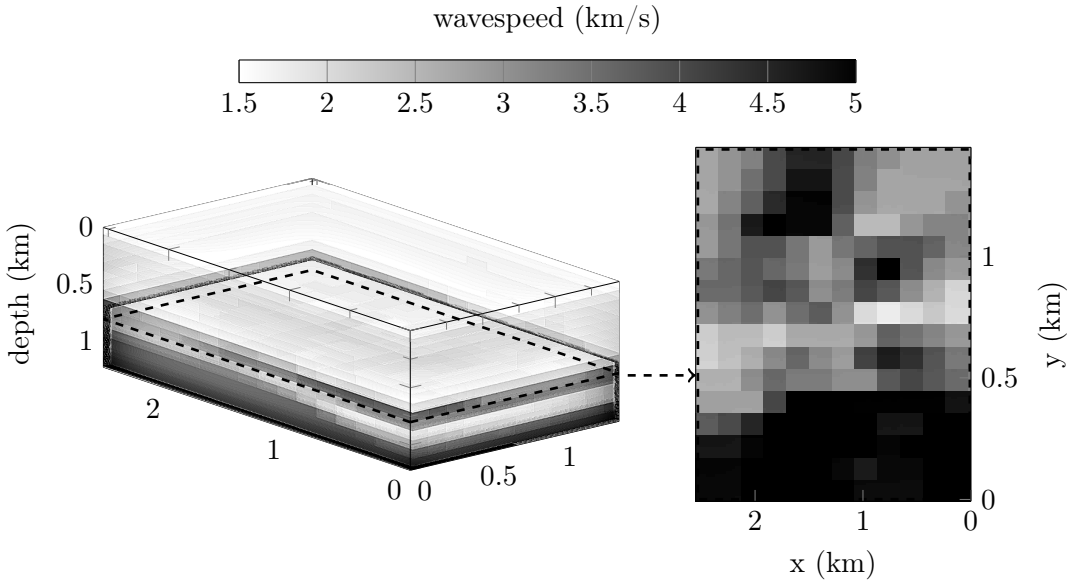
For the computation of the stability estimates we consider c_2 as the model shown in Figure A.3. This setup can be associated with the 'true' subsurface Figure A.2 and starting model Figure A.3. In this context we have chosen the initial guess with no knowledge of any structures by simply considering a one-dimensional variation in depth.

A.3.1 Estimates using the full Dirichlet-to-Neumann map

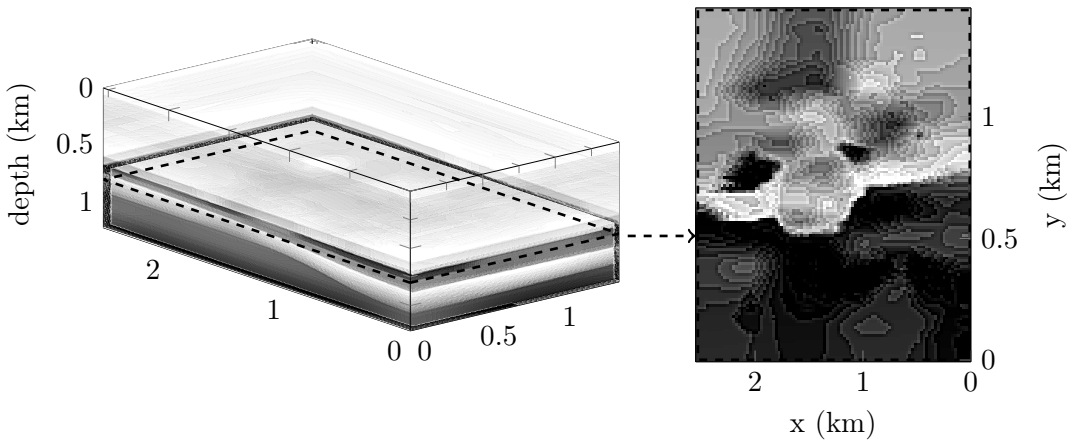
We consider the full data case where the Gaussian sources (see Figure A.1) are positioned on each surface following a regular map. For each source, the data are acquired all over the boundary. We introduce a total of 630 sources and 76538 data points for each.

At a selected partition (number of domains) and frequency, we simulate the data for the two media c_1 and c_2 and compute the difference, from which we deduce the stability constant following equation (A.11).The main difference with the standard seismic setup is that we consider data on all the boundary and not only at the top. This last case will be mentioned in Subsection A.3.2.

The numerical estimates for the stability constant $\mathcal{C}$ should depend on the number of domains N following the expression of the lower and upper bounds defined in Remark A.2,



(a) Partition using $N = 2,880$ domains.



(b) Partition using $N = 1,527,168$ domains.

Figure A.2: Three-dimensional representations and horizontal sections at 800m depth of the reference wave speed (c_1) using different partition, i.e. scales. Every scale has a structured (rectangular) decomposition using piecewise constant. The size of the rectangular boxes defines the scale of the wave speed.

equations (A.2) and (A.2). Thus we fix the frequency and estimate the stability for different partitions. The evolution of the estimates and underlying bounds are presented in Figure A.4 at two selected frequencies, 5 and 10Hz. We plot on a log log scale the function $\log(\mathcal{C}\omega^2)$ to focus on the power of N in the estimates, which is the slope of the lines (4/7 for the upper bound and 1/5 for the lower bound).

Regarding the different coefficients in the analytical bounds, K and K_1 remain undecided and are numerically approximated so that the bounds match the estimates at best.

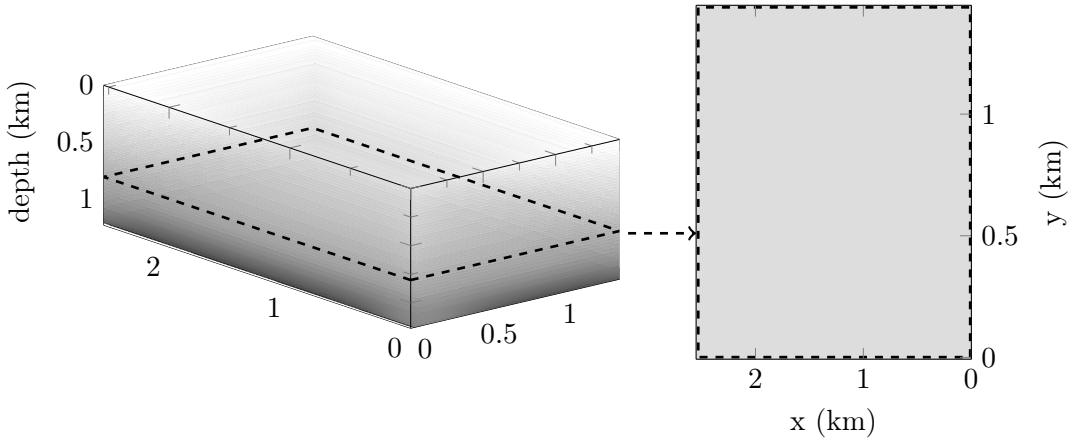
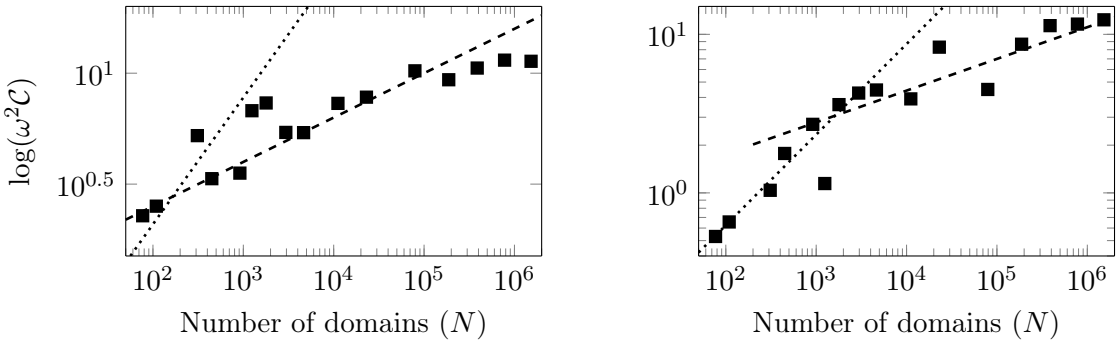


Figure A.3: Three-dimensional wave speed used for the successive estimation of the stability constant (c_2), 3D representation (left) and horizontal sections at 800m depth (right).



(a) Stability estimates at 5Hz frequency

(b) Stability estimates at 10Hz frequency

Figure A.4: The black squares represent the computational estimates of the stability constant (■) depending on the number of domains N at selected frequency. The dashed line (- - -) represents the analytical lower bound and the dotted line (·····) the upper bound, estimated with equation (A.3.1).

For instance the numerical value for K_1 is obtained from equation (A.2) by computing the average value based on the numerical stability estimates and K is approximated following the same principle:

$$K_1 = \frac{1}{n_{st}} \sum_{i=1}^{n_{st}} \frac{\log(4\omega^2 \mathcal{C}_i)}{N_i^{1/5}}, \qquad K = \frac{1}{n_{st}} \sum_{i=1}^{n_{st}} \frac{\log(\omega^2 \mathcal{C}_i)}{(1+\omega^2 B_2) N_i^{4/7}}.$$

Here, n_{st} is the number of numerical stability constant estimates and $\mathcal{C}_i$ the corresponding estimate for partitioning N_i. We actually limit the computation of K to use the first scales as it grows too rapidly. The numerical values obtained are given Table A.1. We also note that the term $\omega^2 B_2$ of the upper bound equation (A.2) is relatively small in the geophysical context as we have here $B_2 = 5.10^{-7}$.

We can see that the stability constant increases with the number of subdomains, as

expected. There are clearly two states in the evolution of the estimates at the highest frequency (10Hz, Figure A.4(b)). For a low number of partitions N the numerical estimates match particularly well the upper bound while at finer scale it follows accurately the lower bound. This is illustrated in Figure A.5 where we decompose the two parts of the estimates between the low and high number of domains.

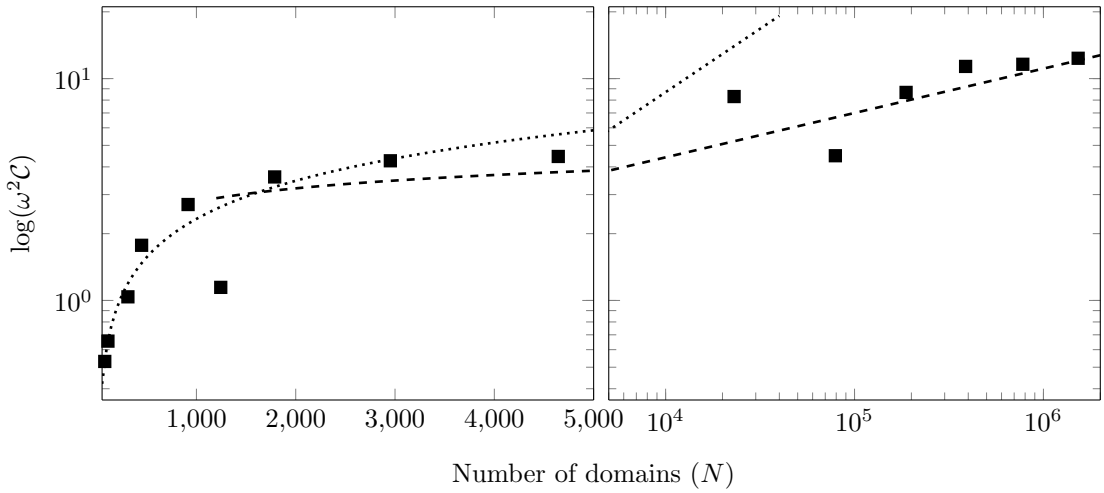


Figure A.5: The black squares represent the computational estimates of the stability constant (■) depending on the number of domains N at 10Hz. The left part shows the coarsest scales which match accurately the upper bound (dotted line, ·····). On the right the finer scale estimates are accurately anticipated by the lower bound (dashed line, - - -). The constants K and K_1 for the computation of the lower and upper bounds are numerically approximated with values given Table A.1, following equation (A.3.1).

	5Hz	10Hz
K_1	1	0.7
K	0.15	0.05

Table A.1: Numerical estimation of the constant in the analytical bounds formulation for the numerical estimates of the stability (Figure A.4, with $B_2 = (1/1400)^2$.

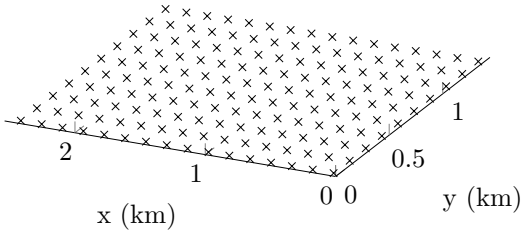
Alternatively for a lower frequency, i.e. 5Hz on Figure A.4(a), the upper bound appears to increase too rapidly while the lower bound matches accurately the evolution of the stability constant estimates. Hence the upper bound we have obtained here is particularly appropriate for coarse scale and high frequency: when the variation of model is much coarser compared to the wavelength.

A.3.2 Seismic inverse problem using partial data

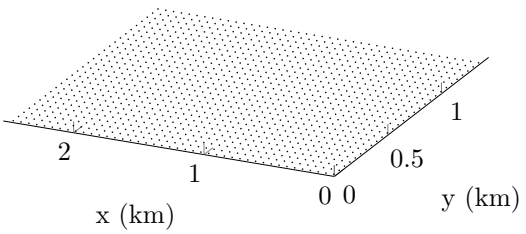
In realistic geophysical experiments for the reconstruction of subsurface area (seismic tomography), it is more appropriate not to consider the full data but partial data only located

on the upper surface. The data obtain from c_1 can be seen as field observation (sensor measurement of a seismic event at the surface). The data using c_2 are simulation using an 'initial guess'. For the reconstruction, we mention the full waveform inversion method, where the recovery follows an iterative minimization of the difference between the measurements and simulations, to successively update the initial guess (see Tarantola (1984); Pratt et al. (1998)). There is also the difference in the boundary conditions where perfectly matched layers (PMLs) or absorbing boundary conditions are invoked instead of the Dirichlet boundary condition for the lateral and bottom boundaries. However, the top boundary is a free surface and remains a Dirichlet boundary condition.

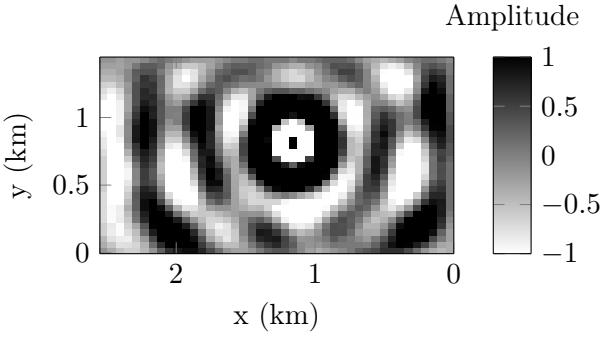
For this test case we reproduce the same experiments but limiting the set of sources and the collected data to be at the top boundary only. We define a set of sources at the surface, separated by 160m along the $x-$axis and 150m along $y-$axis to generate a regular map of 16×10 points. The receivers (data location) are positioned in the same fashion every 60m along the $x-$axis and 45m along $y-$axis and generate a regular map of 43×32 points, see Figures A.6(a) and A.6(b). The partial boundary data computed are illustrated for a single centered boundary shot at 5Hz frequency Figure A.6(c).



(a) the crosses represent the boundary sources locations.



(b) The lattice represents the discretization of the data, i.e. the receivers location.

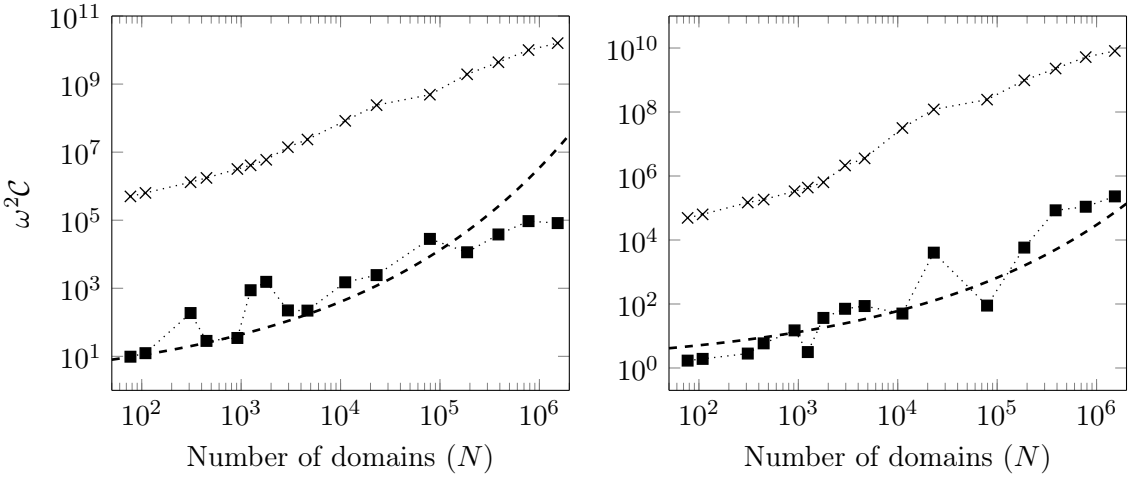


(c) Data recovered from a boundary centered shot, i.e. wavefield measured at the receivers location.

Figure A.6: Illustration of the seismic acquisition set.

In Figure A.7 we compare the stability constant estimates using partial data with the stability constant estimates obtained when considering the full Dirichlet-to-Neumann map as the data. We incorporate the analytical lower bound that was computed in the previous

test case.



(a) Stability estimates at 5Hz frequency

(b) Stability estimates at 10Hz frequency

Figure A.7: Comparison of the computational stability estimates using partial data only located on the top boundary (···×···) and using the full boundary data (···■···). The dashed line (- - -) represents the analytical lower bound as found in Figure A.4.

The numerical estimates of the stability constants for the full and partial data in a log log-scale differ by a constant. This leads us to our conjecture that the log log of the stability constants (as a function of N) of the full and partial data case in the continuous setting differ by a constant.

A.4 Acknowledgment

This research was supported in part by the members, BGP, ExxonMobil, PGS, Statoil and Total, of the Geo-Mathematical Imaging Group now at Rice University. We want to thank the referees for their valuable comments and suggestions, which significantly contributed to improving the quality and clarity of the paper. The research of O.S. is supported by the Austrian Science Fund (FWF), Project P26687-N25 Interdisciplinary Coupled Physics Imaging.

Chapter 4

Seismic inverse problem resolution: Full Waveform Inversion

Abstract

The seismic inverse problem aims the recovery of subsurface Earth parameters and is established depending on the type of the considered subsurface area (reconstruction of the bulk modulus and density in acoustic, reconstruction of the Lamé parameters and the density in elastic, etc). It is also characterized by the available data (field measurements). However, the general method for the reconstruction does not depend on the types of media and is based on an iterative minimization scheme using the seismograms measurements information. It is commonly referred to as Full Waveform Inversion (FWI) and targets a quantitative reconstruction of the area of interest. The optimization problem is naturally conducted using Newton types methods for the iterations. Nonetheless, the numerical cost of second order information (Hessian) might become overwhelming for large scale seismic models, and the gradient descent is an alternative for time-saving. We present here the options we have implemented in our toolbox to solve the harmonic FWI.

Contents

In Chapter 3, we have introduced the inverse problem in the mathematical framework and studied the associated stability. We now focus on the numerical method to reconstruct seismic subsurface. In regards to the theoretical framework, partial data and noise are usually difficult to handle. Furthermore, the observed data usually do not correspond to the same forward operator as the simulation (the Earth is anisotropic by nature). In our project, we reconstruct the subsurface properties following an iterative minimization algorithm, but several alternatives exist for seismic imaging. For example, we can mention the reverse time migration (RTM) where the recovery is based upon the cross-correlation of the forward and backward fields, and appropriate imaging conditions (Claerbout, 1971, 1985). We can also mention the popular Kirchoff migration. The interested reader is referred to the work of Biondi (2006), where different imaging techniques are given.

Iterative methods for the recovery of subsurface parameters allow a quantitative reconstruction of the parameters but requires intense numerical computations. They have been collectively referred to, in reflection seismology, as *Full Waveform Inversion* (FWI). It consists in the minimization of the residuals, defined as the difference between the observation and numerical simulations. The recovery of the subsurface coefficients based upon a least squares minimization scheme originates from the work of Lailly (1983) and Tarantola (1984, 1987b) and relies on the adjoint state method for the computation of the descent direction. We also mention the original work of Bamberger et al. (1977, 1979) in the one-dimensional case. Pan et al. (1988) supposedly introduced the term 'full waveform inversion', with reference to the use of full seismograms information. Lailly, in particular, noticed that the gradient of the cost function could be assimilated to an imaging procedure and computed with the adjoint state method. The use of adjoint state method to calculate the derivative with respect to the parameter is originally promoted by Chavent (1974). Numerical experiments in the two-dimensional case were initially carried out by Gauthier et al. (1986). Mora (1987) extended the formulation to elastic waves. He also noticed that both acoustic and elastic FWI can benefit from wide-aperture data to build the large and intermediate wavelengths in parameter variations in the subsurface. Crase et al. (1990) implemented elastic waveform inversion with field data using short offset and manage to solve the short wavelengths of both P- and S-impedances in the time domain.

This pioneering work was realized in the time domain, and the time-harmonic formulation was promoted by Pratt and other collaborators (Pratt and Worthington, 1990; Pratt and Goulty, 1991), specifically for wide-angle reflection data in Pratt et al. (1996). They also presented the features of the frequency domain formulation of FWI, based on the Newton method and using the discretized Helmholtz equation. Recently, Xu and McMechan (2014) utilize time domain modeling and frequency domain gradient formulation for 2D elastic FWI. This hybrid method is also used by Jun et al. (2013). The recovery of subsurface Earth materials has been extensively developed in the recent years with numerical

methods benefiting from the increase in the computational power.

Newton type algorithms provide a natural framework to solve this nonlinear minimization problem, with early applications in Pratt et al. (1998). However, second order information may be limited due to the underlying numerical cost involved. To avoid the computation (or approximation) of the Hessian, the nonlinear conjugate gradient is the method of choice in the early applications (Gauthier et al., 1986; Mora, 1987). It necessitates the computation of a step length that can typically be estimated by a simple line search. However, one needs to be careful as a wrong estimation may lead to a failure of convergence. For this purpose, different methods exist, see (Nocedal and Wright, 2006, Chapter 3). We also mention the maximum projected curvature depicted in Chavent (2010) and note that for the Landweber iteration scheme, de Hoop et al. (2012) give the analytical definition of this step. In addition, research based on the trust region is also carried out by Eisenstat and Walker (1994); Conn et al. (2000).

In the recent approaches, one accounts for the Hessian, or at least its diagonal, as Shin et al. (2001); Choi et al. (2008) in the Gauss-Newton framework. The second order part in the Hessian (for the Newton method) can be interpreted in terms of multiple scattering (Pratt et al., 1998). To avoid extensive computation, Akcelik et al. (2002) apply multi-scale Newton methods and Ma and Hale (2012) use projected Hessian matrix, so-called P-BFGS. Alternatively, Gélis et al. (2007) use a preconditioned gradient method where the Fréchet derivative is computed by Born and Rytov formulations for least-squares misfit. They note that the model anomalies are hard to reconstruct with the presence of surface waves. They improve the resolution and convergence via selecting the first body waves and surface waves. The conjugate gradient method also provides the possibility of inverting the Hessian without explicitly forming it, following an iterative procedure, see (Nocedal and Wright, 2006, Chapter 5). Application to FWI has been processed in Métivier et al. (2013) with the definition of Hessian vector product from the second order adjoint method (Wang et al., 1992). We refer to this method as *Hessian CG* (for Hessian Conjugate Gradient). In certain earthquake seismology applications, one can even build the Fréchet derivative or Jacobian (for sensitivity purpose, see, for example, Chen et al. (2007)) explicitly. Epanomeritakis et al. (2008) combine outer Gauss-Newton nonlinear iterations with inner conjugate gradient linear iteration, preconditioned by L-BFGS to solve large scale 3D elastic FWI problem in the time domain.

The choice of the norm of the cost function has also been investigated, in particular functions using a logarithmic norm were explored in Tarantola (1987a); Shin and Min (2006). Subsection 4.3.2 reviews some possibilities. Motivated by the lack of the low frequency component in the field data, the use of complex frequencies is also studied in Shin and Cha (2009); Ha et al. (2010). In the underlying Laplace domain, the damping

effects can be treated as time windowing, see Ha et al. (2010). The subsequent improvement of the algorithm following this use of complex frequencies is analyzed in Chapter 5, where we show its connection with the size of the radius of convergence. Another aspect of interest that we study in Chapter 5 is the parametrization, which is known to have crucial effect on the algorithm, and we refer to Section 5.4.

The nonlinearity of FWI has motivated the development of hierarchical multiscale strategies. In the time domain, Bunks et al. (1995) propose the inversion of successive subsets of increasing frequencies, following the intuition that low frequencies are less sensitive to cycle skipping. This multiscale approach can, for example, be related to the subspace search method advocated by Kennett et al. (1988). The time-harmonic formulation, however, is more natural for this purpose. The concept of frequency progression has been considered, amongst others, by Sirgue and Pratt (2004), where the selection of consecutive frequencies is also provided from wavenumber coverage. It is also studied in the problem associated with electromagnetic waves, see, for example, Bao and Li (2005a, 2009).

The application of wavelet bases to compress the successive models in the iteration has been considered by Loris et al. (2007, 2010) in wave-equation tomography, and in FWI by Lin et al. (2012) to reduce the size of the Jacobian. Multi-scale using wavelets is also employed for seismic reconstruction in Yuan and Simons (2014); Yuan et al. (2015). In our approach, the compression, in particular, the compression rate, is connected to the convergence of our multi-level scheme by acting on the stability constant. It provides a conditional well-posedness while reducing the number of unknowns, as depicted in Chapter 3.

Accurate seismic reconstruction requires accounting for the complexity of the subsurface Earth properties: attenuation, elasticity, anisotropy, etc. Malinowski et al. (2011) demonstrate high-resolution imaging of attenuation and phase velocity in the visco-acoustic case. In this context, we also mention the work of Askan et al. (2007). We note that accounting for attenuation is natural for the harmonic wave equation, with the representation complex valued velocity, depicted in Section 1.7. With the recent evolution of high performance computing, FWI becomes more and more popular and heads towards anisotropic media reconstruction, for acoustic we mention Plessix and Cao (2011); Alkhalifah and Plessix (2014) and Rusmanugroho et al. (2017); Guitton and Alkhalifah (2017) in elastic.

Elastic FWI has also been introduced in global seismology. Tromp et al. (2005) connect seismic tomography with adjoint methods for elastic waves, and they also select phase and smoothed models. Fichtner and Trampert (2011) explore both first-order and second-order influence zones of Hessian kernels, recalling the idea of Tarantola (2005). For tomographic model, they demonstrate the difference between Gauss-Newton and full-Newton Hessians, especially for off-diagonal elements. They state that the Gauss-Newton Hessian can lead to

erroneous inferences while full Hessian can correctly account for the effect of nonlinearity on model resolution. Trampert et al. (2012) provide a randomized method to update linear model iteratively, which works if the resolution operator is diagonal-dominant. However, the Hessian matrix may not always have such property. Based on gradient optimization scheme, Fichtner et al. (2013) apply the FWI method in a tomographic model and design a multigrid approach based on the decomposition of a multi-scale earth model.

Some of the major milestones related to FWI have been hereby given and we refer to Virieux and Operto (2009) for an extensive review. Regarding the more general framework of optimization and the Newton method, the literature is certainly voluminous, here we insist on Nocedal and Wright (2006) for a review of optimization algorithms and Chavent (2010); Allaire (2012); Kern (2016), among others.

In this chapter, we will proceed to the formulation of the reconstruction of subsurface properties as a minimization problem. We first review the data available from a seismic campaign in Section 4.2. We emphasize the non-intrusive process for which data are only collected in the near surface area. We present different possibilities regarding the seismic data and the mechanism involved in the two main situations: land and marine seismic. In order to solve the minimization problem, we need the computation of the gradient, and possibly the Hessian, which are conducted from the adjoint state method, see Sections 4.4 and 4.5. Several possibilities can be employed to conduct an optimization problem and we only focus on the ones that have been implemented during this project, in the precise context of FWI, in Section 4.6. Eventually, we provide numerical experiments to compare the effect of incorporating Hessian information (via the conjugate gradient method) with gradient descent methods in Section 4.7.

4.1 Notation

The general quantities related to the inverse problem have already been introduced in Subsection 3.1.1. We naturally follow the same notation and give some more details related to the seismic situations and numerical algorithm. In this chapter, we focus on the discretized representation of the quantities (field, models) so that the numerical framework comes into place naturally.

The essence of the FWI is to conduct the subsurface reconstruction following an iterative minimization of the *residual*. The residual is the difference between the observed measurements and simulations. The simulations employ a guess model and compute the quantities of interest at the receivers positions. The cost function $\mathcal{J}$ can be defined with

different choices of norms, and the standard least squares minimization gives

$$\mathcal{J}(m) = \frac{1}{2}\|\mathcal{F}(m) - d\|^2, \qquad \text{least squares cost function.}$$

In the following we detail the quantities that have already been introduced Subsection 3.1.1 and propose some numerical understanding associated with the inverse wave problem. In Table 4.1, we review the list of symbols.

4.1.1 Acquisition, data and models

Let us first re-give the notation of the problem regarding the domain, which is illustrated in Figure 4.1, for a two dimensional domain. $\Omega \subset \mathbb{R}^2$ or $\mathbb{R}^3$ is the area of interest with boundary Γ. The upper boundary, where it is natural to consider a free surface boundary condition is denoted by Γ_1, the lateral and bottom boundaries, are denoted by $\Gamma_2 = \Gamma/\Gamma_1$.

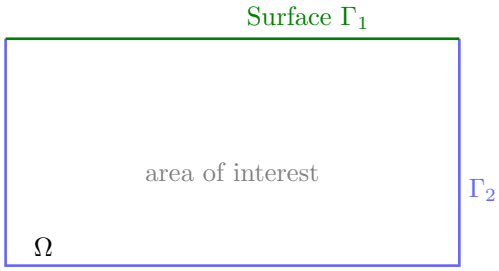


Figure 4.1: Notation for the domain of interest Ω, Γ_1 denotes the upper boundary, $\Gamma_2 = \Gamma/\Gamma_1$ the others. The upper boundary usually consists in a free surface characterizing the interface between the air and the medium. On the other hand the lateral and bottom boundaries have to reflect the natural continuation of the Earth.

The *models* m are the quantities of interest over the domain. In the context of seismic, we have identified the main parameters of interest as the wave speed, the density, the Lamé parameters, etc. Following the Equation (3.1) we have that $m \in \mathcal{M}$, where $\mathcal{M}$ is the model space.

The measurements data are denoted by $d \in \mathcal{D}$, where $\mathcal{D}$ is the data space, as indicated in the Equation (3.2). The data are obtained from a combination of sources and receivers, which are located at, or near, the upper surface for the seismic problem. We introduce

- n_{src} the number of source in the acquisition,
- n_{rcv} the number of receivers per source. In order to simplify the formulation, the number of receivers is assumed to be the same for all sources. However, the use of an independent number of receivers for every source does not change the process.

At selected frequency, the data are denoted d_ω and encompass $n_d = n_{src} \times n_{rcv}$ complex

measures. We further denote by Σ, the positions of the receivers, i.e., where the observations are acquired.

4.1.2 List of symbols

The list of symbols related to the quantities used is given Table 4.1. All symbols are also introduced at their first appearance.

Notation	Meaning	Detail
Ω	domain of interest	$\subset \mathbb{R}^2$ or $\mathbb{R}^3$
Γ	domain boundary	
Γ_1	free surface boundary	$\subset \Gamma$
Γ_2	absorbing boundary or PML	Γ/Γ_1
$\mathcal{P}_\omega$	general wave operator at frequency ω (acoustic, elastic)	
u_ω	general field of interest, solution of the wave equation $\mathcal{P}_\omega$ (pressure or displacement)	
Σ	receivers positions	
$\mathcal{M}$	model space	
$\mathcal{D}$	data space	
m	models of interest (density, wave speed, etc)	$\in \mathcal{M}$
n_{src}	number of sources	
n_{rcv}	number of receivers	
n_d	quantity of data per frequency	$n_{rcv} \times n_{src}$
d_ω	data available at selected frequency ω	$\in \mathcal{D}$
$\mathcal{F}_\omega$	forward operator at frequency ω	$\mathcal{F}_\omega : \mathcal{M} \to \mathcal{D}$
$\mathcal{R}$	linear restriction operator associated with the forward problem	$\mathcal{F}_\omega = \mathcal{R}u_\omega$
$\mathcal{J}$	cost function	

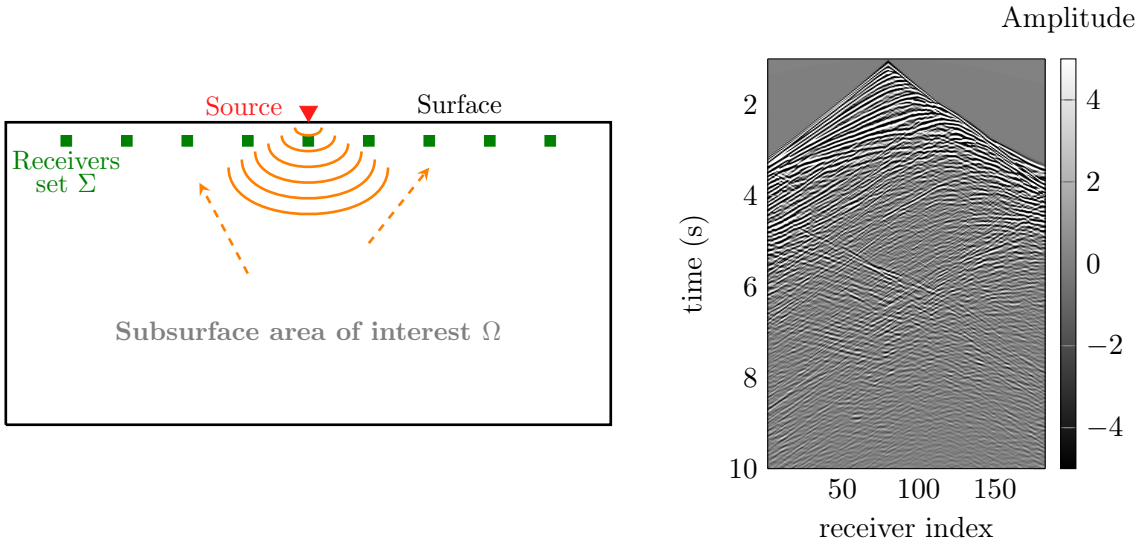
Table 4.1: List of symbols introduced for the seismic inverse problem.

4.2 Seismic data acquisition

In order to reveal subsurface materials, we rely on observations obtained from seismic acquisition on the area of interest. It consists of a set of sources and receivers localized at the surface or slightly underneath when it is possible (in the case of water for example). We have already mentioned the possibility of natural sources for the generation of waves but seismic exploration involves artificial sources for which the position and formulation are (relatively well) known. We recall that our problem is nonintrusive, that is why the

acquisition devices are only localized at (or near) the surface (i.e., we do not consider boreholes, nor any underground a priori data).

The seismic acquisition is naturally related to the phenomenon of wave propagation prescribed by the direct and forward problem. The source emits waves which propagate in the subsurface and receivers are positioned to record the resulting signal. Then the source (and possibly the receivers) is moved, and the operation is reproduced. In Figure 4.2(a) we portray the seismic setup in a simple two-dimensional illustration, for a single source positioned at the surface and where the receivers are located underneath. The wave propagates from the source and is recorded by the receivers. We shall specify the meaning of *record* in the following subsections. In Figure 4.2(b), we present the recorded information from the receivers in a synthetic experiment. We have considered a two-dimensional domain with a centered source. The seismic trace indicates the amplitude of the quantity recorded by the receivers with time.



(a) 2D seismic acquisition.

(b) Seismic trace.

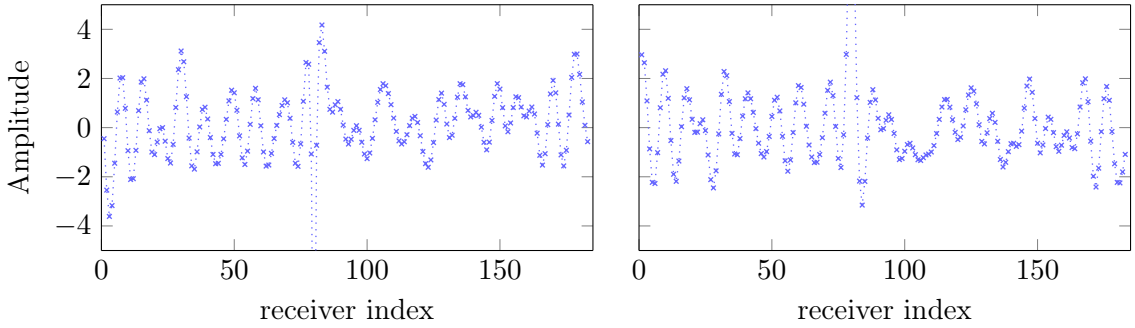
Figure 4.2: (a): Illustration of seismic acquisition for a two-dimensional domain for a single source positioned at the surface which is represented by the red triangle. The wave propagates through the area of interest, then subsurface structures reflect part of the wave (indicated with the dashed arrows). The receivers (represented by the squares) record the resulting signal (direct wave and reflections).
(b): Example of a recorded seismic trace associated with a centrally localized shot in a two-dimensional experiment. The y-axis is the time variable (here from 0 to 10 s). The x-axis denotes the receivers index, here we have considered 183 receivers. The amplitude of the recorded signal is given by the grey scale. This is an acoustic synthetic experiment where the amplitude is the measurement of the pressure at the position of the receiver.

We see that the seismic trace depends naturally on time. However, our method is designed in the frequency domain so that we need the discrete Fourier transform of the

original measurements. Let us denote $S_k = \{S_{k,i}\}$ the signal recorded by the receiver k, where $k = \{1, \dots, n_{rcv}\}$, $i = \{1, \dots, n_t\}$ with n_{rcv} the total number of receivers and n_t the number of time steps. For each i corresponds a local experimental time $t_i = (i-1)d_t$, where d_t is the time step between successive measures, assuming we start from $t_1 = 0$. Then the data at frequency ω are given by the discrete Fourier transform of the signal:

$$\mathfrak{F}(S_k) = \widehat{S}_k(\omega) = \sum_{i=1}^{n_t} S_{k,i} e^{-\mathrm{i}\omega t_i}. \tag{4.1}$$

Meanwhile the original measurements are real, the resulting frequency domain signal is complex. The discrete Fourier transform (Equation (4.1)) of the trace Figure 4.2(b) is shown Figure 4.3 at the selected 5 Hz frequency. We shall see that in the presence of noise in the original time domain trace, the Fourier transform is the less affected around the source frequency peak, which we illustrate as part of the experiments in Chapter 6.



(a) Real part.

(b) Imaginary part.

Figure 4.3: Discrete Fourier transform of the seismic trace Figure 4.2(b) at 5 Hz. For every receiver index k the Fourier domain complex value is given by $\widehat{S}_k$, as indicated in Equation (4.1)

Standardly, seismic acquisition is composed of several hundreds of shots, with even more receivers associated for each of them. The actual type of recorded data depends on the acquisition devices; we focus here on the pressure, displacement or velocity components, which are the usual quantities arising from seismic campaigns. In particular, it is adapted to the type of medium, and we distinguish marine and land acquisitions. In Chapter 1, we have presented the wave equation accordingly to these two types of propagation (acoustic or elastic), and the forward problem is associated with the pressure for acoustic and displacement for elastic.

4.2.1 Marine acquisition

Seismic marine acquisition aims at retrieving acoustic or acoustic-elastic subsurface media. The upper layer of water can be known and it is where the seismic acquisition is conducted.

The coupling of acoustic and elastic media refers to the consideration of solid matters underneath the water bottom, but it does not change the setup. Namely, the sources and receivers are carried out by a ship at the surface. These consist of towed streamers which are grouped via parallel cables. In this setup the receivers usually move for every source position, yet we disregard this possibility for simplicity.

In this situation an *air gun* is employed for the source and generates a pressure perturbation. In the case where the source is positioned at the surface it corresponds to a Dirichlet boundary condition. Following the acoustic wave equation (1.28) the resulting pressure, p, for a single air gun, solves

$$\begin{cases} -\omega^2 p(\mathbf{x}) - \kappa(\mathbf{x})\nabla \cdot \left(\dfrac{1}{\rho(\mathbf{x})}\nabla p(\mathbf{x})\right) = 0, & \mathbf{x} \in \Omega \\ p(\mathbf{x}) = f(\mathbf{x}), & \mathbf{x} \in \Gamma_1, \end{cases}$$

where f is the characterization of the air gun, Γ_1 denotes the free surface boundary, i.e. the interface between air and water. Here we omit boundary conditions to be implemented on Γ_2, which are numerically represented with PMLs or ABC .

4.2.1.1 Hydrophones measurements

Seismic marine acquisition standardly relies on a single type of waves observation obtained from *hydrophones*, which provide measurements of the pressure field. It records the information $p(\mathbf{x}_k)$, where $\mathbf{x}_k$ stands for the hydrophone position. The forward problem associated with a single source is

$$\mathcal{F}_\omega \ : \ m = (c, \rho) \quad \rightarrow \quad \Big\{ p(\mathbf{x}) \mid_\Sigma \Big\} \ = \ \Big\{ p(\mathbf{x}_1), \ldots, p(\mathbf{x}_{n_{rcv}}) \Big\},$$

where Σ denotes the receivers location set, i.e., n_{rcv} positions.

4.2.1.2 Dual sensors devices

It is relatively recently that PGS (Petroleum Geo-Services) developed new types of acquisition devices which allow multi components data recordings. The dual sensors allow the recording of the vertical velocity in addition to the pressure, see Carlson et al. (2007); Tenghamn et al. (2007). In this case we have the following forward operator:

$$\begin{aligned} \mathcal{F}_\omega \ : \ m = (c, \rho) \quad \rightarrow \quad & \Big\{ p(\mathbf{x}) \mid_\Sigma ; \ v_\nu(\mathbf{x}) \mid_\Sigma \Big\} \\ & = \Big\{ p(\mathbf{x}_1), \ldots, p(\mathbf{x}_{n_{rcv}}); \ v_\nu(\mathbf{x}_1), \ldots, v_\nu(\mathbf{x}_{n_{rcv}}) \Big\}, \end{aligned} \tag{4.2}$$

where v_ν denotes the normal velocity. Chapter 7 is devoted to the use of Cauchy data, which we relate to the dual sensors, and illustrates novel perspectives.

4.2.2 Land acquisition, Neumann-to-Dirichlet map

Land seismic campaigns make use of vibroseis trucks. The initial perturbation is generated from a *baseplate* that imposes pressure at the surface. For the measurement, *geophones* are employed and allow the recording of the particle displacement. All devices in land acquisition are usually exclusively located at the surface. This configuration is thoroughly studied in Baeten (1989). In the case of isotropic elastic wave propagation the displacement $\boldsymbol{u}$ solves the system with Neumann boundary condition,

$$\begin{cases} -\rho(\mathbf{x})\omega^2\boldsymbol{u}(\mathbf{x}) - \nabla\big(\lambda(\mathbf{x})\nabla\cdot\boldsymbol{u}(\mathbf{x})\big) - \nabla\cdot\big(\mu(\mathbf{x})\big[\nabla\boldsymbol{u}(\mathbf{x}) + (\nabla\boldsymbol{u}(\mathbf{x}))^T\big]\big) = 0, & \mathbf{x}\in\Omega, \\ \partial_\nu\boldsymbol{u}(\mathbf{x}) = \boldsymbol{g}(\mathbf{x}), & \mathbf{x}\in\Gamma_1. \end{cases}$$

Then the geophones record the information on the displacement $\boldsymbol{u}(\mathbf{x}_k)$ so that the forward problem is written for the single source $\boldsymbol{g}$,

$$\mathcal{F}_\omega \ : \ m=(\lambda,\mu,\rho) \quad \rightarrow \quad \Big\{\boldsymbol{u}(\mathbf{x})\mid_\Sigma\Big\} \ = \ \Big\{\boldsymbol{u}(\mathbf{x}_1),\dots,\boldsymbol{u}(\mathbf{x}_{n_{rcv}})\Big\}.$$

Here we have considered isotropic elastic medium but the consideration of the TTI equation is straightforward and simply modify the wave equation in Ω accordingly (see Equation (1.30)). Then it redefines $m=(\lambda,\mu,\rho,\epsilon,\delta,\gamma,\theta,\phi)$ in three dimensions: accounting for the Thomsen's parameters.

Elastic data are represented with the so called Neumann-to-Dirichlet (NtoD) map and we refer to Baeten (1989) for further details. We also mention that the geophones may not be able to record all directions, in particular the measure of normal direction to the surface may be the only available component.

4.3 Iterative minimization problem

4.3.1 Representation of the forward problem and cost function

The forward problem $\mathcal{F}$ describes the correspondence between the model and the data. It has been introduced in the Equation (3.3). It usually encompasses the wave operator, which depends on the type of medium (acoustic, elastic, etc) as introduced in Chapter 1. In order to remain general we represent the wave equation at selected frequency ω with the

operator $\mathcal{P}_\omega$, such that the wave equation is written as (omitting the spatial dependency)

$$\mathcal{P}_\omega(m)u = f. \tag{4.3}$$

Here $u = u(m)$ represents arbitrarily the displacement or pressure, depending on the nature of the propagation medium (elastic and acoustic respectively). Similarly, f represents the appropriate kind of source. The wave equation is here identified with $\mathcal{P}_\omega$. For example, taking the Helmholtz equation gives $\mathcal{P}_\omega = (-\Delta - \omega^2 c^{-2})$, where m is identified with the velocity c. This notation will be very practical to derive the different steps to conduct the resolution of the inverse problem, because those steps are identical for all types of medium.

Let us denote by $u_\omega^{(s)}$ the solution of the wave equation associated with $\mathcal{P}_\omega$ for the source $f^{(s)}$. Let us decompose the forward problem and the data in the same way such that we denote $\mathcal{F}_\omega^{(s)}$ and $d_\omega^{(s)}$ the quantities associated with a single source s at selected frequency ω. The essence of the forward operator $\mathcal{F}$ is to prescribe some quantity of interest at the position of the receivers. We have seen in Section 4.2 that for seismic, it is often directly the solution of the wave equation that is recorded (i.e. hydrophones or geophones measurement). Then it is common to represent the forward operator with the introduction of a linear *restriction operator*, $\mathcal{R}$, acting on the solution of the wave equation u such that

$$\mathcal{F}_\omega^{(s)}(m) = \mathcal{R}(u_\omega^{(s)}(m)). \tag{4.4}$$

Numerically, the creation of the numerical simulation for the cost function follows two steps: first we solve the appropriate wave problem Equation (4.3) to obtain the solution u. Second we restrict this solution to the receiver locations, applying the proper transform if needed (e.g., in the case of Cauchy data, the operator $\mathcal{R}$ encompasses a normal derivative), Equation (4.4).

The cost function can consequently be rewritten separating the set of sources s and frequencies ω:

$$\begin{aligned} \mathcal{J}(m) &= \frac{1}{2} \sum_{\omega=\omega_1}^{\omega_{n_\omega}} \sum_{s=1}^{n_{src}} \|\mathcal{F}_\omega^{(s)}(m) - d_\omega^{(s)}\|^2, \\ &= \frac{1}{2} \sum_{\omega=\omega_1}^{\omega_{n_\omega}} \sum_{s=1}^{n_{src}} \|\mathcal{R}(u_\omega^{(s)}(m)) - d_\omega^{(s)}\|^2. \end{aligned} \tag{4.5}$$

Eventually, the reconstruction of the subsurface parameters is cast as the resolution of the following minimization problem

$$\min_{m \in \mathcal{M}} \mathcal{J}(m).$$

The data d_ω and $\mathcal{F}_\omega(m)$ can clearly be seen as complex vectors following Section 4.2. d_ω is the list of measurements, and $\mathcal{F}_\omega(m)$ the simulation at the position of interest. The cost

function is naturally formulated with the complex inner product,

$$\begin{aligned}\mathcal{J}(m) &= \frac{1}{2}\sum_{\omega=\omega_1}^{\omega_{n_\omega}}\sum_{s=1}^{n_{src}}\left\langle \mathcal{F}_\omega^{(s)}(m) - d_\omega^{(s)}, \mathcal{F}_\omega^{(s)}(m) - d_\omega^{(s)}\right\rangle \\ &= \frac{1}{2}\sum_{\omega=\omega_1}^{\omega_{n_\omega}}\sum_{s=1}^{n_{src}}\left(\mathcal{F}_\omega^{(s)}(m) - d_\omega^{(s)}\right)^*\left(\mathcal{F}_\omega^{(s)}(m) - d_\omega^{(s)}\right),\end{aligned}$$

where * stands for the adjoint, i.e. the transposed complex conjugate such that for a vector $v \in \mathbb{C}$, $v^* = \overline{v}^T$. Note that the use of complex data is a specificity of the frequency domain algorithm.

The notation can become unnecessarily complicated because of the multiple indexes, on the other hand the sums do not affect in any way the steps we follow to solve the problem. That is why we omit the sum over the sources to simplify the notation. This can be seen as assuming one single source in the acquisition so that the cost function is written as

$$\mathcal{J}(m) = \frac{1}{2}\sum_{\omega=\omega_1}^{\omega_{n_\omega}} \|\mathcal{F}_\omega(m) - d_\omega\|^2. \tag{4.6}$$

The sum over the frequency can be omitted as well, for simplicity. Yet it requires a little more details, see Subsection 4.3.3.

4.3.2 Norm of the objective function

The misfit functional for the minimization problem can be defined in different ways regarding the norm. The traditional $L2$ norm has been widely employed and remains the most popular, and it is the one we have used to introduce the cost function, see Equation (4.6). However, one can easily imagine a wide variety of norms for the functional. For example Shin et al. (2007); Bednar et al. (2007); Pyun et al. (2007) render a comparison of misfit functions depending on the use of the phase and/or the amplitude of the signal, and also employ the logarithmic function. The phase criterion is also studied in Bozdağ et al. (2011). In the frequency domain, following the Fourier transform of the data, one can select the real or imaginary part of the signal to conduct the reconstruction as well. We also mention the work of Brossier et al. (2010b), which emphasize the use of the $L1$ norm.

Among those possibilities, the logarithmic function has been explored in the early work of Tarantola (1987a) and has been proven to be crucial in the case of complex frequency inversion, see Shin and Min (2006). The misfit function is written as

$$\mathcal{J}(m) = \frac{1}{2}\sum_{\omega=\omega_1}^{\omega_{n_\omega}} \|\log(\mathcal{F}_\omega(m)) - \log(d_\omega)\|^2. \tag{4.7}$$

Naturally the logarithm gives more importance on small signal variations. We will later use this cost function when running experiments using complex frequency, see Chapter 6.

4.3.3 Frequency progression

Following the frequency decomposition, the sum over frequencies appeared in the cost function Equation (4.5). We have mentioned the operation of filtering subset of frequency bands, in the time domain with the work of Bunks et al. (1995). This is obviously very natural in the frequency domain, see Sirgue and Pratt (2004). Hence the general sum over the frequencies do not cover all the frequencies but subset of one or several consecutive frequencies. After the iterative minimization has been conducted on a first subset, we update the set and pursue the inversion. We can also invert frequency sequentially. We propose an analysis and explanation for the frequency separation and progression in Chapter 5 where we study the effect of the frequency on the size of the basin of attraction.

In order to have clearer expression, we drop the ω indexes for now, as if we had a sequential frequency update in the reconstruction algorithm, and write the cost function

$$\begin{aligned} \mathcal{J}(m) &= \frac{1}{2}\|\mathcal{F}(m) - d\|^2 \\ &= \frac{1}{2}\|\mathcal{R}u(m) - d\|^2. \end{aligned} \tag{4.8}$$

Remark 4.1. *in the following we employ the least squares minimization to derive the successive expressions. However, the analysis is not restricted to the L2 norm and alternative misfits can be used, as given in Subsection 4.3.2. Regarding this choice of cost function, the use of Cauchy data certainly provides the major difference and we will elaborate in Chapter 7.*

4.3.4 Minimization using Newton type methods

FWI performs successive updates of the subsurface model so that the simulation can eventually match (with more or less accuracy) the measurements. It is a nonlinear iterative minimization problem. The Newton method naturally appears as the best candidate when resolving such optimization problems. It is based upon the first order Taylor expansion which states that, assuming a small model perturbation δ_m,

$$\mathcal{J}(m + \delta_m) = \mathcal{J}(m) + \mathcal{J}'(m)\delta_m + o(\delta_m^2),$$

where $o(\delta_m^2)$ stands for the terms of order higher than or equal to two. It is well known that when the minimum, say $m^\dagger$, of the cost function is reached, then the derivative $\mathcal{J}'(m^\dagger)$ is

zero. Hence we are looking for this critical point (the minimizer). The Taylor expansion is applied onto the derivative of the cost function and we obtain

$$\mathcal{J}'(m+\delta_m) = \mathcal{J}'(m) + \mathcal{J}''(m)\delta_m + o(\delta_m^2).$$

The Newton method aims in building a sequence of m_i converging towards the zero of the function where the iterative update is given from δ_m, neglecting the higher order terms. We have

$$m_{i+1} = m_i - \frac{\mathcal{J}'(m_i)}{H(m_i)}. \tag{4.9}$$

H is the second order derivative of the cost function, i.e. the Hessian (named after Ludwig Otto Hesse, 1811–1874).

Remark 4.2. *The global minimum of the cost function has its gradient to zero. However, the gradient is also null for any local minimum of the cost function. Hence the condition $\nabla\mathcal{J} = 0$ does not guarantee to find the global minimum, especially because of the large scale (in particular the number of parameters to reconstruct) of our inverse problem. This issue is addressed with the use of model partition to provide some regularization for the problem. Also in Chapter 5, we review the class of problems where local minima can be avoided. It allows us to define quantitative estimates of the model space size, see Chapter 5.*

In the general Algorithm 1, we present the iterative minimization process with the appropriate steps, accounting for frequency progression. The heart of the problem is the model update. In the next sections we further explicit some of the various possibilities available for iterative optimization; we refer to the work of Nocedal and Wright (2006) for an extensive review of optimization techniques.

4.3.5 Complications

We have already mentioned the nonlinearity and ill-posedness of the problem, which makes it complex to solve. Now we distinguish two categories of complication for the iterative reconstruction. The first is due to numerical capability. Indeed, due to the size of seismic domains, the resulting large scale optimization problem may suffer from numerical restriction. The cost of computing the Hessian can easily become too expensive and second order information becomes unavailable. This is probably the major limitation for the computations and it can justify the use of gradient descent algorithm (see Subsection 4.6.1). More generally, large scale optimization requires to use the computational memory wisely, such as some algorithms presented in Nocedal and Wright (2006). We also mention the

Data:
- Acquisition set (sources and receivers positions),
- frequency domain measurements,
- initial model m_1.

for $k \in \{1, \dots, n_\omega\}$ **do**
  for $i \in \{1, \dots, n_{iter}\}$ **do**
    $j = i + (k-1) \times n_{iter}$
    for $s \in \{1, \dots, n_{src}\}$ **do**
      Numerical resolution of the wave equation using m_j
    end
    Apply the appropriate restriction operator $\mathcal{R}$ to deduce $\mathcal{F}_{\omega_k}(m_j)$
    Compute the residuals and underlying cost function $\mathcal{J}(m_j)$
    if $\mathcal{J}(m_j)$ *is 'low enough'* **then**
      end iteration loop
    else
      Update the model with method of choice to obtain m_{j+1}
    end
  end
end

Algorithm 1 – General FWI algorithm: FWI iterative minimization in the frequency domain. The model update is typically performed using Newton type methods. When using direct solver for the resolution of the frequency domain wave equation, the loop over the shots can straightforwardly be replaced by a multi right-hand sides linear system solve, largely reducing the numerical cost involved.

cost of solving many forward simulations, which requires the use of advanced and efficient computing as well, in particular the linear algebra solver for our harmonic propagation.

The second category of complication is related to the general setup of the problem and can be associated with several other applications of inverse problem, in particular we can mention that

- few data are available compared to the number of coefficients investigated to define the area,
- data have relative accuracy, due to the devices precision,
- data are only located at the surface, no information can be obtained inside the domain, nor on the sides,
- no prior information on the subsurface structures is usually available.

When one takes into account those partial data with limited accuracy, very few mathematical results hold. For the seismic reconstruction, the problem can be non-unique in practice (Landa and Treitel, 2016). Hence, the frequency progression and multi-scale techniques are important to mitigate the issues arising from the nonlinearity and ill-posedness of the

problem.

4.4 Gradient computation with first order adjoint state

Following the Newton method, we seek the computation of the gradient of the cost function. Here we consider the formulation where the frequency and sources sums have been simplified, see Equation (4.8). We make use of the notation introduced for the forward problem, based on the general wave operator given in Equation (4.3).

Because we work with the frequency domain formulation of the wave equation, the quantity of interest which defines u (pressure, displacement, etc) is complex. This complex data set is a particularity of the frequency domain approach and does not hold in the time domain. This is an important difference with working in the time domain and we believe this aspect is too often omitted (we actually have not been able to find an appropriate reference of the adjoint state problem using complex data in a geophysical framework). Let us introduce the function G, obtained from the cost function definition of Equation (4.8) by

$$\mathcal{J}(m) = G(m, u(m)). \tag{4.10}$$

It is important to notice that here, G is not analytic (holomorphic) with respect to the field u. It may appear not important at first sight (we seek the derivative with respect to m and not u), but it is crucial because of the nonlinearity in m of the cost function. In the following subsection we detail the tools which are required to pursue the gradient computation.

4.4.1 Derivative and gradient

Let us first remind some standard definitions concerning the gradient. We refer to Kern (2016) for more details and discussion.

Definition 4.1 (Fréchet derivative)**.** *Let U and V be Banach spaces, the function $h : U \to V$ is Fréchet differentiable at $x \in U$ if there exists a bounded linear operator $A : W \to V$, with $U \subset W$ such that*

$$\lim_{\delta \to 0} \frac{\|h(x+\delta) - h(x) - A\delta\|_W}{\|\delta\|_V} = 0,$$

We can equivalently apply the first order expansion, which gives

$$h(x+\delta) = h(x) + A\delta + o(\|\delta\|),$$

If A exists, it is the Fréchet derivative of h at x, that we denote $A = h'(x)$.

Definition 4.2 (Gradient)**.** *We denote $\nabla h(x)$ the gradient of h at the point $x \in U$. It is defined such that*

$$< \nabla h(x), \delta >= h'(x)\delta, \qquad \forall \delta \in U.$$

This is also the Riesz-representation of the derivative f'.

We note that in the finite dimensional case, where $< \nabla h(x), \delta >= \nabla h(x)^T \delta$ in $\mathbb{R}$, the relation between the gradient and the differential is

$$\nabla h(x)^T = h'(x). \tag{4.11}$$

By considering the variable $x = \{x_1, \ldots, x_n\}$ we have

$$\nabla h(x)^T = h'(x) = \left(\frac{\partial h(x)}{\partial x_1}, \ldots, \frac{\partial h(x)}{\partial x_n} \right).$$

4.4.2 Complex derivation

In the frequency domain, we have seen that G is not analytic, thus not differentiable, with respect to u, due to the complex conjugation. A workaround is relatively standard for such functional, see for example the work of Brandwood (1983); Li and Adali (2006); Kreutz-Delgado (2009). It is based on elements of complex calculus such as the Wirtinger calculus. The main idea is to consider independently the complex variable and its conjugate. Respectively we denote z and $\overline{z}$, where $z = x + \mathrm{i}y$ and $\overline{z} = x - \mathrm{i}y$ with $\mathrm{i}^2 = -1$. We first review the results of Brandwood (1983).

Theorem 4.1. *(Brandwood, 1983, Theorem 1) Let $g : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ be a function of a complex number z and its conjugate $\overline{z}$ and let g be analytic with respect to each variable (z and $\overline{z}$) independently. Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{C}$ be the function of the real variables x and y such that $g(z, \overline{z}) = f(x, y)$ where $z = x + \mathrm{i}y$. Then the partial derivative $\partial_z g$ (treating $\overline{z}$ as a constant) gives the same result as $(\partial_x f - \mathrm{i}\partial_y f)/2$. Similarly, $\partial_{\overline{z}} g$ is equivalent to $(\partial_x f + \mathrm{i}\partial_y f)/2$.*

Corollary 4.1. *Following the statement of Theorem 4.1, we have*

$$\overline{\frac{\partial g}{\partial \overline{z}}} = \frac{\partial g}{\partial z}.$$

Proof. By direct application of Theorem 4.1,

$$\overline{\frac{\partial g}{\partial \overline{z}}} = \frac{1}{2}\left(\overline{\frac{\partial f}{\partial x} + \mathrm{i}\frac{\partial f}{\partial y}} \right) = \frac{1}{2}\left(\frac{\partial f}{\partial x} - \mathrm{i}\frac{\partial f}{\partial y} \right) = \frac{\partial g}{\partial z}.$$

□

Theorem 4.2. *(Brandwood, 1983, Theorem 2) Let $f : \mathbb{C} \to \mathbb{R}$ be a function of a complex variable z. Let $f(z) = g(z, \overline{z})$ where $g : \mathbb{C} \times \mathbb{C} \to \mathbb{R}$ is such that $g(z, a)$ and $g(b, \overline{z})$ are each analytic function of z (with a and $b \in \mathbb{C}$). Then a necessary and sufficient condition for f to have a stationary point is that $\partial_z g = 0$ where the partial derivative with respect to z treats $\overline{z}$ as constant. Similarly $\partial_{\overline{z}} g = 0$ is also a necessary and sufficient condition.*

Theorem 4.3. *(Brandwood, 1983, Theorem 3) Let $f : \mathbb{C}^N \to \mathbb{R}$ be a function of a complex vector z. Let $f(z) = g(z, \overline{z})$ where $g : \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{R}$ is analytic $\forall$ z and $\overline{z}$ considered independently. Then a necessary and sufficient condition for f to have a stationary point is that $\nabla_z g = 0$. Similarly $\nabla_{\overline{z}} g = 0$ is also a necessary and sufficient condition.*

We straightforwardly apply the theorems to the misfit function where we consider $u := z = x + \mathrm{i}y$. In this context the functional can be written as

$$G : (x, y) \quad \to \quad \frac{1}{2}\|\mathcal{R}(x + \mathrm{i}y) - d\|^2,$$

where x, y and d can be assimilated with vectors. Then by deriving independently with respect to x and y we obtain

$$\begin{cases} \dfrac{\partial G}{\partial x} = \dfrac{1}{2}\mathcal{R}^*(\mathcal{R}(u) - d) + \dfrac{1}{2}(\mathcal{R}(u) - d)^*\mathcal{R}, \\ \dfrac{\partial G}{\partial y} = -\dfrac{\mathrm{i}}{2}\mathcal{R}^*(\mathcal{R}(u) - d) + \dfrac{\mathrm{i}}{2}(\mathcal{R}(u) - d)^*\mathcal{R}. \end{cases}$$

With Theorem 4.1, we can further deduce the derivative of G with respect to u and $\overline{u}$, where they are considered independent such that $G = G(u, \overline{u})$,

$$\begin{cases} \dfrac{\partial G}{\partial u} = \dfrac{1}{2}(\mathcal{R}(u) - d)^*\mathcal{R}, \\ \dfrac{\partial G}{\partial \overline{u}} = \dfrac{1}{2}\left(\mathcal{R}^*(\mathcal{R}(u) - d)\right)^T = \dfrac{1}{2}(\mathcal{R}(u) - d)^T\overline{\mathcal{R}}. \end{cases} \tag{4.12}$$

The computation of the Hessian requires the second order derivative, here we straightforwardly see that

$$\begin{cases} \dfrac{\partial^2 G}{\partial u^2} = 0, & \dfrac{\partial^2 G}{\partial \overline{u}^2} = 0, \\ \dfrac{\partial^2 G}{\partial \overline{u}\partial u} = \dfrac{\partial^2 G}{\partial u \partial \overline{u}} = \dfrac{1}{2}\mathcal{R}^*\mathcal{R}. \end{cases}$$

Remark 4.3. *We can proceed similarly working with the forward problem instead of the wavefield, using a function T in a similar fashion as G was introduced in Equation (4.10)*

with $\mathcal{J}(m) = T(m, \mathcal{F}(m))$. *The complex derivation with respect to* $\mathcal{F}$ *gives analogously*

$$\begin{cases} \dfrac{\partial T}{\partial \mathcal{F}} = \dfrac{1}{2}(\mathcal{F}(m) - d)^*, \\ \dfrac{\partial T}{\partial \overline{\mathcal{F}}} = \dfrac{1}{2}(\mathcal{F}(m) - d)^T, \end{cases} \tag{4.13}$$

and

$$\begin{cases} \dfrac{\partial^2 T}{\partial \mathcal{F}^2} = 0, & \dfrac{\partial^2 T}{\partial \overline{\mathcal{F}}^2} = 0, \\ \dfrac{\partial^2 T}{\partial \overline{\mathcal{F}} \partial \mathcal{F}} = \dfrac{\partial^2 T}{\partial \mathcal{F} \partial \overline{\mathcal{F}}} = \dfrac{1}{2}. \end{cases} \tag{4.14}$$

We have presented how to derive a function with respect to complex variable, namely by considering independently the conjugate. For the update of the cost function, however, we need to derive with respect to a real parameter (the model) a function involving complex variable. The following theorems give the proper framework by introducing what can be seen as the chain rule for complex derivation.

Theorem 4.4. *Consider the complex-valued function* f *of a real parameter* m *and the real-valued functions* g_1 *and* g_2 *such that* $f(m) = g_1(z(m), \overline{z}(m)) + \mathrm{i} g_2(z(m), \overline{z}(m))$. *The derivative with respect to the real parameter* m *is defined by*

$$\frac{\partial f}{\partial m} = \frac{\partial g}{\partial z}\frac{\partial z}{\partial m} + \frac{\partial g}{\partial \overline{z}}\frac{\partial \overline{z}}{\partial m}.$$

Proof. From the definition of f we have

$$\begin{aligned} \frac{\partial f}{\partial m} &= \frac{\partial g_1(z(m), \overline{z}(m))}{\partial m} + \mathrm{i}\frac{\partial g_2(z(m), \overline{z}(m))}{\partial m} \\ &= \frac{\partial g_1}{\partial z}\frac{\partial z}{\partial m} + \frac{\partial g_1}{\partial \overline{z}}\frac{\partial \overline{z}}{\partial m} + \mathrm{i}\frac{\partial g_2}{\partial z}\frac{\partial z}{\partial m} + \mathrm{i}\frac{\partial g_2}{\partial \overline{z}}\frac{\partial \overline{z}}{\partial m} \\ &= \frac{\partial (g_1 + \mathrm{i} g_2)}{\partial z}\frac{\partial z}{\partial m} + \frac{\partial (g_1 + \mathrm{i} g_2)}{\partial \overline{z}}\frac{\partial \overline{z}}{\partial m} \\ &= \frac{\partial g}{\partial z}\frac{\partial z}{\partial m} + \frac{\partial g}{\partial \overline{z}}\frac{\partial \overline{z}}{\partial m} \end{aligned}$$

□

Theorem 4.5. *Consider the real-valued functions* f *and* g *defined by* $f(m) = g(z(m), \overline{z}(m))$. *From Theorem 4.4 we have*

$$\frac{\partial f}{\partial m} = 2\,\mathrm{Re}\left(\frac{\partial g}{\partial z}\frac{\partial z}{\partial m}\right) = 2\,\mathrm{Re}\left(\frac{\partial g}{\partial \overline{z}}\frac{\partial \overline{z}}{\partial m}\right).$$

Proof. Direct application of Theorem 4.4 gives

$$\frac{\partial f}{\partial m} = \mathrm{Re}\left(\frac{\partial g}{\partial z}\frac{\partial z}{\partial m} + \frac{\partial g}{\partial \overline{z}}\frac{\partial \overline{z}}{\partial m}\right). \tag{4.15}$$

We use Theorem 4.1 and Corollary 4.1, and take $z(m) = x(m) + \mathrm{i}y(m)$ to have

$$\mathrm{Re}\left(\frac{\partial g}{\partial \overline{z}}\frac{\partial \overline{z}}{\partial m}\right) = \mathrm{Re}\left(\overline{\frac{\partial g}{\partial z}}\frac{\partial \overline{z}}{\partial m}\right) = \mathrm{Re}\,\overline{\left(\frac{\partial g}{\partial z}\overline{\frac{\partial \overline{z}}{\partial m}}\right)} = \mathrm{Re}\left(\frac{\partial g}{\partial z}\overline{\frac{\partial \overline{z}}{\partial m}}\right),$$

where

$$\overline{\frac{\partial \overline{z}}{\partial m}} = \overline{\frac{\partial(x - \mathrm{i}y)}{\partial m}} = \overline{\left(\frac{\partial x}{\partial m} - \mathrm{i}\frac{\partial y}{\partial m}\right)} = \frac{\partial x}{\partial m} + \mathrm{i}\frac{\partial y}{\partial m} = \frac{\partial z}{\partial m}.$$

We inject in Equation (4.15) to obtain

$$\frac{\partial f}{\partial m} = \mathrm{Re}\left(\frac{\partial g}{\partial z}\frac{\partial z}{\partial m}\right) + \mathrm{Re}\left(\frac{\partial g}{\partial z}\frac{\partial z}{\partial m}\right) = 2\,\mathrm{Re}\left(\frac{\partial g}{\partial z}\frac{\partial z}{\partial m}\right).$$

The alternative expression is obtained similarly but by replacing $\partial_z g$ in Equation (4.15), instead of $\partial_{\overline{z}} g$. □

Theorem 4.5 provides the necessary tools to compute the cost function gradient with respect to the model m, which is the one we require. The functional depends on the field u which itself depends on m. We write the functional with respect to the forward problem first, following the notation of the Remark 4.3,

$$\begin{aligned}\frac{\partial}{\partial m}\Big(\mathcal{J}(m)\Big) = \frac{\partial}{\partial m}\Big(T(m, \mathcal{F}(m))\Big) &= 2\,\mathrm{Re}\left(\frac{\partial T}{\partial \mathcal{F}}\frac{\partial \mathcal{F}}{\partial m}\right) = \mathrm{Re}\Big((\mathcal{F}(m) - d)^* D\mathcal{F}(m)\Big) \\ &= 2\,\mathrm{Re}\left(\frac{\partial T}{\partial \overline{\mathcal{F}}}\frac{\partial \overline{\mathcal{F}}}{\partial m}\right) = \mathrm{Re}\Big((\mathcal{F}(m) - d)^T D\overline{\mathcal{F}}(m)\Big).\end{aligned}$$

Here $D\mathcal{F}$ stands for the Fréchet derivative of the forward problem with respect to m. The two formulations are identical because of the real part. we define $D\mathcal{F}^* = D\overline{\mathcal{F}}^T = \overline{D\mathcal{F}}^T$. We write the gradient from the derivative, following the formula of Equation (4.11),

$$\nabla_m \mathcal{J}(m) = \nabla \mathcal{J}(m) = \mathcal{J}'(m)^T = \mathrm{Re}\Big(D\mathcal{F}(m)^*(\mathcal{F}(m) - d)\Big). \tag{4.16}$$

Similarly, we can write the gradient with the use of the restriction operator using the function G of Equation (4.10),

$$\nabla \mathcal{J}(m) = \mathrm{Re}\left(\frac{\partial u(\mathbf{x})}{\partial m}^* \mathcal{R}^*\Big(\mathcal{R}\big(u(\mathbf{x})\big) - d\Big)\right). \tag{4.17}$$

The steps to formulate the gradient are general and can easily be specified depending on the data (acoustic, elastic, Cauchy, etc). Similarly it is easily adapted to alternative norms for the cost function.

4.4.3 Adjoint state method

The seismic inverse problem involves large scale media. The computation of the Fréchet derivative (or $\partial_m u$) is expensive in terms of numerical operations, which is related to the number of unknowns (i.e., the number of coefficients to represent the subsurface medium). Moreover, the underlying discretization of $D\mathcal{F}$ leads to a large matrix which may suffer from memory limitation. The computation of this Jacobian gives rise to the sensitivity kernel, which can, however, be useful for optimization schemes, allowing a more precise algorithm, (Chavent, 2010; Allaire, 2012; Kern, 2016).

Instead, the adjoint state method allows us to obtain the gradient of the cost function without the direct computation of the Fréchet derivative. This method has been introduced in the framework of control theory by Lions and Mitter (1971) and Chavent (1974) implemented it for the computation of the gradient of a functional with respect to a parameter. It is further detailed for the wave problem in the work of Lailly (1983); Tarantola (1987b, 1988). It has become a fundamental technique employed for many optimization problems, e.g., Allaire (2012); Chavent (2010); Kern (2016). For geophysical inverse problem, a review can be found in Plessix (2006) for time and frequency domain formulation of the wave propagation. However, the specificity implied by the complex derivation is less common in geophysics.

In order to compute the gradient of the cost function, we formulate the following minimization problem incorporating the solution of the wave equation as a constraint,

$$\min_m \mathcal{J}(m) = G(m, v(m)) \qquad \text{subject to } \mathcal{P}(v) = f. \tag{4.18}$$

The problem is recast into a formulation with Lagrangian to give

$$\mathcal{L}(m, v(m), \tilde{\gamma}) = G(m, v(m)) + \big\langle \mathcal{P}(v) - f, \tilde{\gamma} \big\rangle,$$

where $\tilde{\gamma}$ represents the Lagrange multipliers associated with v. Here we have that $m \in \mathbb{R}$ and $v \in \mathbb{C}$ so that we need to use the complex derivation techniques of Subsection 4.4.2. In particular, from Theorem 4.4, the derivation with respect to m gives

$$\frac{\partial}{\partial m}\Big(\mathcal{L}(m, v(m), \tilde{\gamma})\Big) = \mathrm{Re}\left(\frac{\partial \mathcal{L}}{\partial m} + \frac{\partial \mathcal{L}}{\partial \overline{v}}\frac{\partial \overline{v}}{\partial m} + \frac{\partial \mathcal{L}}{\partial v}\frac{\partial v}{\partial m}\right).$$

From Theorem 4.5, we have observed that this can be written as

$$\frac{\partial}{\partial m}\Big(\mathcal{L}(m, v(m), \tilde{\gamma})\Big) = \mathrm{Re}\left(\frac{\partial \mathcal{L}}{\partial m} + 2\frac{\partial \mathcal{L}}{\partial \overline{v}}\frac{\partial \overline{v}}{\partial m}\right). \tag{4.19}$$

The minimization of the Problem (4.18) is obtained when the derivative is equal to

zero (Allaire, 2012). In particular, the adjoint state γ is selected to verify $\mathrm{Re}(\partial_{\overline{v}}\mathcal{L}+\partial_v\mathcal{L}) = 2\,\mathrm{Re}(\partial_{\overline{v}}\mathcal{L}) = 0$. We have

$$2\,\mathrm{Re}\left(\frac{\partial\mathcal{L}}{\partial\overline{v}}\right) = 2\,\mathrm{Re}\left(\frac{\partial G}{\partial\overline{v}} + \mathcal{P}^*\gamma\right).$$

Let us now substitute the derivative of the cost function with respect to the complex field, given in the Equation (4.12); so that γ verifies

$$\mathrm{Re}\left(\mathcal{R}^*(\mathcal{R}(v)-d) + \mathcal{P}^*\gamma\right) = 0.$$

Hence the adjoint state γ is selected to solve the problem

$$\mathcal{P}^*(\gamma) = -\mathcal{R}^*(\mathcal{R}(v)-d). \tag{4.20}$$

By taking u solution of $\mathcal{P}u = f$, we observe that Equation (4.19) becomes

$$\frac{\partial}{\partial m}\Big(\mathcal{L}(m,u(m),\gamma)\Big) = \frac{\partial}{\partial m}\Big(\mathcal{J}(m)\Big) = \mathrm{Re}\left(\frac{\partial\mathcal{L}}{\partial m} + 2\frac{\partial\mathcal{L}}{\partial\overline{u}}\frac{\partial\overline{u}}{\partial m}\right) = \mathrm{Re}\left(\frac{\partial\mathcal{L}}{\partial m}\right),$$

where the latest equality is obtained by taking γ solution of the problem (analogous to Equation (4.20))

$$\mathcal{P}^*(\gamma) = -\mathcal{R}^*(\mathcal{R}(u)-d). \tag{4.21}$$

In this context, the gradient of the cost function is given by

$$\nabla_m\mathcal{J}(m) = \nabla_m\mathcal{L} = \mathrm{Re}\left(\big\langle\partial_m(\mathcal{P})u,\gamma\big\rangle\right)^T. \tag{4.22}$$

The adjoint state method provides the gradient of the cost function at the cost of an additional problem (for the computation of γ). This additional problem, Equation (4.21) is the adjoint of the forward problem where the right-hand side basically consists in the residuals. Eventually, it requires the derivative of the operator with respect to the model parameters, which we have referred to as $\partial_m\mathcal{P}$. We further illustrate the computation in the following.

Remark 4.4. *We have formulated the adjoint state without considering the boundary conditions for the wave equation. This should appear as additional constraints in the Problem 4.18. Dirichlet or Neumann boundary conditions do not affect the steps, simply generating analogous conditions in the definition of the adjoint state. However, the use of absorbing boundary conditions or PMLs may be delicate due to the introduction of complex numbers.*

4.4.4 Numerical cost

We have shown that the computation of the gradient of the misfit function requires the resolution of one additional problem, to obtain the adjoint state, Equation (4.21). The problem stands for every source in the seismic acquisition so that there are as many additional problems as the number of forward problems to generate the simulation. The gradient is the trivially given with the sum over the sources,

$$\nabla \mathcal{J}(m) = \sum_s \mathrm{Re}\left(\left\langle \partial_m(\mathcal{P}) u^{(s)}, \gamma^{(s)} \right\rangle \right)^T.$$

Once again we emphasize the importance of having multi-right-hand sides solvers which allow the resolution of linear system for many independent right-hand sides at a reduced cost (this is less straightforward with the time domain formulation). Moreover, there is no need in matrix factorization, which is the computationally expensive (time and memory) step of the direct solvers. Indeed the adjoint state solves the adjoint of the direct problem and we have $\mathcal{P}^*\gamma = f$ equivalent to $\mathcal{P}\gamma^* = f^*$ so that the same information can be used for the forward and the adjoint problem (for example LU factors). Hence the cost of obtaining the adjoint state is relatively low in terms of time, and the memory does not require much more than what is already used for the forward problem.

Finally, once the adjoint field has been retrieved, there is still the need to derive the operator with respect to the models. Yet this operation can usually remain relatively cheap following a little thinking, as illustrated in the following subsection.

4.4.5 Derivation of the operator

After the appropriate discretization of the wave equation has been carried out, $\mathcal{R}$ and $\mathcal{P}$ can be seen as a matrix. All of the ingredients to compute the gradient of the cost function through the first order adjoint state are known but the derivative of the operator, $\partial_m \mathcal{P}$. This quantity depends obviously on the model but also on the method used for the discretization. Here we illustrate the computation in the context of Finite Differences, for an acoustic case with constant density, with one single parameter to invert: the velocity.

4.4.5.1 Example for finite differences discretization

The Helmholtz equation defines the propagation of waves from a single medium parameter, the wave speed c. In such media, the inverse problem aims at reconstructing this velocity. The general wave operator $\mathcal{P}$ we have used for the adjoint state method is defined by

$$\mathcal{P} := (-\Delta - \omega^2 c^{-2}),$$

see Equation (1.27). For simplicity we derive $\mathcal{P}$ with respect to c^{-2}. The FD stencil standardly gives one coefficient per nodes in the discretization, let us say we have a total of n coefficients. Then the gradient of the cost function with respect to the model is a vector of size n such that every component is given by

$$\big[\nabla_m \mathcal{J}(m)\big]_k = \mathrm{Re}\left(\sum_j \Big[\overline{\partial_{c_k^{-2}}(\mathcal{P})u}\Big]_j [\gamma]_j\right).$$

Here we have used the notation $[\cdot]_k$ for the k^{th} component of the vector.

For FD method, it is common to consider the velocity with one coefficient per node (where the structured grid has n nodes). Then, those coefficients only appear in the diagonal of the discretization matrix, unlike Galerkin methods where fluxes terms can give different patterns. The derivation with respect to one of those coefficients, say c_k^{-2} gives a matrix with only one non-zero value at the diagonal position $[k,k]$ and this value is $-\omega^2$. The resulting matrix is a squared matrix with size the total number nodes. In this configuration the quantity $\Big[\partial_{c_k^{-2}}(\mathcal{P})u\Big]_j$ is not zero only when $k = j$ and we have

$$\Big[\partial_{c_k^{-2}}(\mathcal{P})u\Big]_j = \begin{cases} 0 & \text{if } k \neq j, \\ -\omega^2 u_k & \text{if } k = j. \end{cases}$$

The components of the gradient are straightforwardly given by

$$\big[\nabla_m \mathcal{J}(m)\big]_k = \mathrm{Re}\Big((\overline{-\omega^2 u_k})\gamma_k\Big).$$

We have obtained an important simplification of the formulation as the derivative of the discretized operator does not require the computation of an additional matrix. Instead, we end up with a multiplication of two vectors with an appropriate weight.

Remark 4.5. *This formulation holds in the case where we have assumed a piecewise constant representation of the velocity with one coefficients per nodal value, which is similar to the number of coefficients for the discretized solution u. This assumption appears natural for the finite differences case, where nodal disretization is performed. It is important to notice that defining a different representation for the velocity will impact the gradient formula, which must be rewritten accordingly.*

4.4.5.2 Extension

The derivation of the discretized operator must be conducted carefully depending on the situation. In particular some aspects must be taken into account when deriving the discretized wave operator from a Galerkin type discretization. Here fluxes involving the models may appear in the discretization and must be dealt with, as well as penalty terms. Moreover, as mentioned in the Remark 4.5, the way the model is represented generates specific derivation.

4.4.6 Parametrization

The parametrization refers to the choice of parameter to invert in the algorithm (e.g., c^{-2} or c for acoustic inversion). It naturally influences the gradient and it is proven to be crucial, especially in the context of multi-parameters reconstruction. Tarantola (1986) already mentioned the importance of an appropriate choice by selecting parameters as uncorrelated as possible. Since then, various possibilities have been tested. (Tarantola, 2005, p.126) also decides the parametrization to incorporate a priori information based on Gaussian functions. In Chapter 5, Sections 5.4, 5.5 and 5.6 we compare several parametrization in the point of view of the Fréchet derivative and illustrate the differences with application to seismic reconstruction.

4.4.7 Adjoint state for the HDG formulation

The adjoint state formulation has been derived assuming the forward operator $\mathcal{P}$ acts directly on the field u when solving the wave equation, with $\mathcal{P}u = f$. However, depending on the discretization method, the field can be obtained from an alternative problem resolution. This is in particular the case of the HDG method, where the field is expressed from multipliers Λ, which are solutions of a specific linear system. The method is for example given in Bonnasse-Gahot (2015). Here the point is not to review the method but simply to show how it affects the computation of the gradient through the adjoint state method. We follow the formulation of Bonnasse-Gahot (2015) which states the wave problem in the HDG framework as

$$\begin{cases} A_k U_k + C_k \mathcal{A}_k \Lambda = S_k & \forall k, \\ B_k U_k + L_k \mathcal{A}_k \Lambda = 0 & \forall k, \end{cases} \tag{4.23}$$

where A, B, C and L are matrices defined for the individual cell k (following the domain discretization, see Subsection 2.2.1), U encompasses the wavefield while Λ represents the

appropriate Lagrange multipliers for the system. S is the source. Note that A and C depend on the medium parameter, but neither B and L. Here the field is expressed locally on each cell with U_k whereas the multipliers Λ are defined for the interface of the cells only. Thus, $\mathcal{A}$ can be seen as the map from interface (global) to the cell (local) index. To finalize the HDG formulation we can sum over all the cells, and use the transposed of the index matrix on the second equation of the System (4.23) to define

$$\sum_k \mathcal{A}_k^T \big(B_k U_k + L_k \mathcal{A}_k \Lambda\big) = 0. \tag{4.24}$$

U is then expressed from Λ with the first equation of the System (4.23):

$$A_k U_k = S_k - C_k \mathcal{A}_k \Lambda \qquad \forall k,$$

injecting in the Equation (4.24) gives the following linear system for the resolution of the multipliers Λ:

$$\sum_k \mathcal{A}_k^T \big(L_k - B_k A_k^{-1} C_k\big) \mathcal{A}_k \Lambda = -\sum_k \mathcal{A}_k^T B_k A_k^{-1} S_k. \tag{4.25}$$

For simplicity we refer to the linear system (4.25) with

$$\mathcal{P}_{HDG} \Lambda = \mathcal{S}_{HDG},$$

where $\mathcal{P}_{HDG} = \big(\sum_k \mathcal{A}_k^T \big(L_k - B_k A_k^{-1} C_k\big) \mathcal{A}_k\big)$ and $\mathcal{S}_{HDG} = -\sum_k \mathcal{A}_k^T B_k A_k^{-1} S_k$.

Hence, the resolution of the wave equation follows two steps. First we retrieve the global multipliers Λ from the Equation (4.25) and then we deduce the wavefield U on each cell. Once again, the reader is referred to Bonnasse-Gahot (2015) and the references therein for the details. Eventually, the total field u can be expressed from the local one,

$$u = \sum_k \mathcal{W}_k^T U_k,$$

introducing $\mathcal{W}$ to link the local (cell) to the global (domain) indexes.

The computation of the cost function gradient follows the adjoint state method and the underlying minimization problem becomes

$$\min_m \mathcal{J}(m) \quad \text{subject to} \quad A_k U_k + C_k \mathcal{A}_k \Lambda = S_k \text{ and } \mathcal{A}_k^T \big(B_k U_k + L_k \mathcal{A}_k \Lambda\big) = 0, \quad \forall k,$$

where the system of equations (4.23) has been included for the constraints. The formulation with Lagrangian multipliers becomes

$$\mathcal{L}(m, \tilde{U}, \tilde{\Lambda}, \tilde{\gamma}_1, \tilde{\gamma}_2) = \mathcal{J}(m) + \sum_k \big\langle A_k \tilde{U}_k + C_k \mathcal{A}_k \tilde{\Lambda} - S_k, \tilde{\gamma}_{1,k} \big\rangle + \big\langle \mathcal{A}_k^T \big(B_k \tilde{U}_k + L_k \mathcal{A}_k \tilde{\Lambda}\big), \tilde{\gamma}_2 \big\rangle,$$

so that two multipliers are introduced, $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$, for each constraint. $\tilde{\gamma}_1$ is defined over the cell and $\tilde{\gamma}_2$ is global for the cells interfaces, following the nature of the field U and Λ

respectively. Furthermore, we define

$$\mathcal{L}_k(m,\tilde{U}_k,\tilde{\Lambda},\tilde{\gamma}_{1,k},\tilde{\gamma}_2)=\mathcal{J}(m)+\left\langle A_k\tilde{U}_k+C_k\mathcal{A}_k\tilde{\Lambda}-S_k,\tilde{\gamma}_{1,k}\right\rangle+\left\langle\mathcal{A}_k^T\left(B_k\tilde{U}_k+L_k\mathcal{A}_k\tilde{\Lambda}\right),\tilde{\gamma}_2\right\rangle.$$

The derivative of the function is given by

$$\frac{\partial}{\partial m}\left(\mathcal{L}(m,\tilde{U},\tilde{\Lambda},\tilde{\gamma}_1,\tilde{\gamma}_2)\right)=\mathrm{Re}\left(\frac{\partial\mathcal{L}}{\partial m}+\frac{\partial\mathcal{L}}{\partial\tilde{\Lambda}}\frac{\partial\tilde{\Lambda}}{\partial m}+\sum_k\frac{\partial\mathcal{L}_k}{\partial\tilde{U}_k}\frac{\partial\tilde{U}_k}{\partial m}\right).$$

Note that here, the complex derivation is hidden for $\tilde{\Lambda}$ and $\tilde{U}$ for the sake of notation but the derivation must be carried out with respect to the field and its conjugate, similarly to what has been done in Subsection 4.4.3.

We now take U and Λ to solve the Problem (4.24), and we select the adjoint states γ_1 and γ_2 such that they are the critical point of the Lagrangian, with $\partial_U\mathcal{L}=0$ and $\partial_\Lambda\mathcal{L}=0$. We consider the derivative with respect to U_k at fixed k and use the complex derivation introduced in Section 4.4.2 to obtain the conditions

$$\begin{cases}\partial_{U_k}\mathcal{L}=0=\mathcal{A}_k\mathcal{R}^*(\mathcal{R}(u)-d)+A_k^*\gamma_{1,k}+B_k^*\mathcal{A}_k\gamma_2, & \forall k,\\ \partial_\Lambda\mathcal{L}=0=\displaystyle\sum_k\mathcal{A}_k^TC_k^*\gamma_{1,k}+\mathcal{A}_k^TL_k^*\mathcal{A}_k\gamma_2,\end{cases}$$

where we have used that the index matrix is real, $\overline{\mathcal{A}}=\mathcal{A}$. The formulation of the adjoint states equations follow

$$\begin{aligned}&A_k^*\gamma_{1,k}=-\mathcal{A}_k\mathcal{R}^*(\mathcal{R}(u)-d)-B_k^*\mathcal{A}_k\gamma_2, \qquad \forall k,\\ &\sum_k\mathcal{A}_k^TC_k^*\gamma_{1,k}+\mathcal{A}_k^TL_k^*\mathcal{A}_k\gamma_2=0.\end{aligned}\tag{4.26}$$

Replacing γ_1 in the second equation gives

$$\begin{aligned}\sum_k\mathcal{A}_k^TC_k^*A_k^{-*}\left(\mathcal{A}_k\mathcal{R}^*(\mathcal{R}(u)-d)-B_k^*\mathcal{A}_k\gamma_2\right)+\mathcal{A}_k^TL_k^*\mathcal{A}_k\gamma_2=0,\\ \sum_k\mathcal{A}_k^T\left(L_k^*-C_k^*A_k^{-*}B_k^*\right)\mathcal{A}_k\gamma_2=\sum_k\mathcal{A}_k^TC_k^*A_k^{-*}\left(\mathcal{A}_k\mathcal{R}^*(\mathcal{R}(u)-d)\right).\end{aligned}$$

We recognize on the left hand side the adjoint of the operator $\mathcal{P}_{HDG}$ introduced for Equation (4.25), so that γ_2 solves the problem

$$\mathcal{P}_h^*\gamma_2=\sum_k\mathcal{A}_k^TC_k^*A_k^{-*}\left(\mathcal{A}_k\mathcal{R}^*(\mathcal{R}(u)-d)\right).$$

From the computation of γ_2, one deduces the expression of γ_1 with the first equation of (4.26). The gradient is eventually given by

$$\nabla_m\mathcal{J}(m)=\frac{\partial}{\partial m}\left(\mathcal{L}(m,U,\Lambda,\gamma_1,\gamma_2)\right)^T=\mathrm{Re}\left(\sum_k<\partial_mA_kU_k+\partial_mC_k\mathcal{A}_k\Lambda,\gamma_{1,k}>\right)^T,$$

because L and B do not depend on the models. Eventually, we see that the adjoint state

can easily be adapted for other situations, following the same steps with caution. Here, the two steps necessary to obtain the wavefield (computation of the multipliers Λ and then the computation of U on each cell) are retrieved for the adjoint, with the computation of the global γ_2 first and the resolution of γ_1 locally on each cell from γ_2 next.

4.5 Hessian computation

The computation of a dense Hessian is usually unrealistic for optimization in seismic, due to the dimension of the problem and numerical restrictions (regarding the computational time and memory). Furthermore, the Newton algorithm requires the computation of the inverse of the matrix. Pratt et al. (1998) recall the full and Gauss-Newton Hessian computation in the time-harmonic domain, feasible for relatively small area of interest. More generally Hessian kernels for wave problem are detailed in Tarantola (1984); Fichtner and Trampert (2011). In order to avoid large computation, some applications simply approximate the Hessian or account for the diagonal, references have already been given in introduction.

In the Newton method, the descent direction, say s, is given by the Equation (4.9), which can be written as

$$H(m)s = -\nabla\mathcal{J}(m), \tag{4.27}$$

where H designates the Hessian. In order to solve the system, the conjugate gradient method, see Subsection 4.5.4, provides an iterative approximation of the direction only requiring the matrix vector multiplication $H(m)v$, $\forall v$. The problem is then recast to compute the product Hessian vector for any vector at reduced cost, see Wang et al. (1992); Nocedal and Wright (2006); Métivier et al. (2013). This computation can rely on the adjoint state approach. In the context of seismic, the method is detailed and applied in Métivier et al. (2013).

We review the computation of Hessian vector multiplication using the second order adjoint state in Subsection 4.5.2. We consider the two main cases for Hessian utilization: *full* and *Gauss-Newton* approximation. Then we detail the conjugate gradient iterations in Subsection 4.5.4, in order to define the appropriate descent direction for the iterative minimization.

4.5.1 Hessian formulations

The second order derivative of the cost function employs the complex derivation techniques introduced in Subsection 4.4.2. We can apply Theorem 4.4 to derive the cost function twice, using the function T introduced in Remark 4.3,

$$\nabla^2 \mathcal{J}(m) = H(m) = \mathrm{Re}\left(\frac{\partial}{\partial m}\left(\frac{\partial T}{\partial \mathcal{F}}\frac{\partial \mathcal{F}}{\partial m} + \frac{\partial T}{\partial \overline{\mathcal{F}}}\frac{\partial \overline{\mathcal{F}}}{\partial m}\right)\right).$$

From Theorem 4.5, we know that this is equivalent to

$$\begin{aligned} H(m) &= 2\,\mathrm{Re}\left(\frac{\partial}{\partial m}\left(\frac{\partial T}{\partial \mathcal{F}}\frac{\partial \mathcal{F}}{\partial m}\right)\right) \\ &= 2\,\mathrm{Re}\left(\frac{\partial^2 T}{\partial \mathcal{F}^2}\left(\frac{\partial \mathcal{F}}{\partial m}\right)^2 + \frac{\partial^2 T}{\partial \mathcal{F}\partial \overline{\mathcal{F}}}\frac{\partial \overline{\mathcal{F}}}{\partial m}\frac{\partial \mathcal{F}}{\partial m} + \frac{\partial T}{\partial \mathcal{F}}\frac{\partial^2 \mathcal{F}}{\partial m^2}\right). \end{aligned}$$

We have shown in Remark 4.3, Equation (4.14), that the second order derivatives of T with respect to $\mathcal{F}$ and $\overline{\mathcal{F}}$ vanish and the cross derivative gives $1/2$ so that we have

$$H(m) = \mathrm{Re}\left(2\frac{\partial T}{\partial \mathcal{F}}\frac{\partial^2 \mathcal{F}}{\partial m^2} + \frac{\partial \overline{\mathcal{F}}}{\partial m}\frac{\partial \mathcal{F}}{\partial m}\right).$$

Let us introduce $D\mathcal{F} = \partial_m \mathcal{F}$, and inject Equation (4.13), using the real part, we can finally obtain

$$H(m) = \mathrm{Re}\left(D^2\mathcal{F}(m)^*(\mathcal{F}(m) - d) + D\mathcal{F}(m)^* D\mathcal{F}(m)\right). \tag{4.28}$$

The Hessian expression defined in the above Equation (4.28) may also be referred to as the '*full Hessian*' and consists in all the terms of the Hessian. Alternatively, the Gauss-Newton approximation H_{GN} can be employed and neglect the second order term $D^2\mathcal{F}(m)$, such that

$$H_{GN}(m) = D\mathcal{F}(m)^* D\mathcal{F}(m). \tag{4.29}$$

The approximation may in particular be justified by the complexity and relative weak effect of second order term for seismic application, as explained in Tarantola (2005); Pratt et al. (1998).

Remark 4.6 (Regularization)**.** *In our application we also incorporate a regularization parameter to be added on the diagonal of the Gauss-Newton Hessian (i.e. Levendberg–Marquardt, see Chavent (2010); Kern (2016)). the Hessian becomes $H = H + \beta I_d$, where I_d is the identity matrix and $\beta \geq 0$ is a scalar. For seismic problem, Pratt et al. (1998) explain that it stabilizes the method, because of the generally ill-conditioned Hessian matrix. More generally, several techniques exist in term of regularization, such as the Tikhonov regularization for example. The main problem is to select efficiently the damping parameter β, which is a difficult step. For this purpose, the trust region that we mentioned in introduction*

may be applied (Kern, 2016).

Remark 4.7. *From the complex derivation we can detail the Fréchet derivative with the restriction operator as it was realized for the gradient, see Equation (4.17). We have*

$$H(m) = \mathrm{Re}\left(\left(\frac{\partial^2 u}{\partial m^2}\right)^* \mathcal{R}^*\Big(\mathcal{R}\big(u\big) - d\Big) + \left(\frac{\partial u}{\partial m}^* \mathcal{R}^*\right)\left(\mathcal{R}\frac{\partial u}{\partial m}\right)\right).$$

The full and Gauss-Newton Hessians are identified similarly.

4.5.2 Second order adjoint state

Let us take a real vector v, we seek the computation of its multiplication with the Hessian, $H(m)v$ (note that the Hessian is a real symmetric matrix), in the full and Gauss-Newton senses. In order to illustrate the different steps we consider the general framework introduced for the first order adjoint state computation. The misfit function corresponds with Equation (4.8). Here we follow the steps defined in Métivier et al. (2013) but we adapt them to the use of the complex derivation framework introduced before. The adjoint state follows the traditional three steps:

1. define the minimization problem with constraints and underlying formulation with Lagrangian,
2. the Lagrange multipliers give the adjoint sates when imposing the derivative to zero,
3. the gradient is computed.

In Table 4.2, we review the different functions introduced so that one can follow more easily, they are also introduced at their first appearance.

Notation	**Detail**
v	real vector to multiply with the Hessian
u	wavefield solution of the wave equation
γ	first order adjoint state
$\mathcal{J}_D(m,u,\gamma) = \big\langle \partial_m\big(\mathcal{P}\big)u, \gamma\big\rangle$	'complex' gradient of the misfit functional $\mathcal{J}$
$h_1(m,u,\gamma) = \mathcal{J}_D(m,u,\gamma)v$	function minimized for the full Hessian vector product case
$h_2(m,u(m),w) =< u(m), w >$	function minimized for the Gauss-Newton Hessian vector product case

Table 4.2: List of notation used for the second order adjoint state.

4.5.2.1 Full Hessian multiplication

For the computation of the full Hessian multiplication, we consider the following minimization problem,

$$\min_m h_1(m,u,\gamma) \quad \text{subject to } \mathcal{P}(u)=f; \quad \mathcal{P}^*(\gamma)=-\mathcal{R}^*(\mathcal{R}(u)-d).$$

This minimization problem encompasses two constraints: for the forward wave problem (u) and the first order adjoint state equation (γ). h_1 is defined such that

$$h_1(m,u,\gamma)=\mathcal{J}_D(m,u,\gamma)v$$

where $\mathcal{J}_D(m,u,\gamma)$ can be viewed as the 'complex' gradient from the first order adjoint state definition of Equation (4.22), without the real part,

$$\mathcal{J}_D(m,u,\gamma)=\big\langle \partial_m\big(\mathcal{P}\big)u,\gamma\big\rangle.$$

Direct derivation gives the expression we seek with

$$\frac{\partial}{\partial m}\bigg(h_1(m,u,\gamma)\bigg)=H(m)v.$$

Once again, because u and γ are complex, the derivation of $\mathcal{J}_D$ must be considered carefully. By definition, we have

$$\mathcal{J}_D(m,u,\gamma)=\big(\partial_m(\mathcal{P})u\big)^*\gamma.$$

We apply Theorem 4.1 considering u and $\overline{u}$ are independent, as well as γ and $\overline{\gamma}$, let us denote $\mathcal{J}_D(m,u,\gamma)=G_D(m,u,\overline{u},\gamma,\overline{\gamma})$, we have

$$\begin{cases} \dfrac{\partial G_D}{\partial \gamma}=\big(\partial_m(\mathcal{P})u\big)^*, & \dfrac{\partial G_D}{\partial \overline{\gamma}}=0,\\ \dfrac{\partial G_D}{\partial u}=0, & \dfrac{\partial G_D}{\partial \overline{u}}=\big(\partial_m(\mathcal{P})\big)^*\gamma. \end{cases}$$

Similarly, we compute the derivative for $G_h(m,u,\overline{u},\gamma,\overline{\gamma})=h_1(m,u,\gamma)$ with respect to γ and u,

$$\begin{cases} \dfrac{\partial G_h}{\partial \gamma}=0, & \dfrac{\partial G_h}{\partial \overline{\gamma}}=\big(\partial_m(\mathcal{P})u\big)v,\\ \dfrac{\partial G_h}{\partial u}=\big(\partial_m(\mathcal{P})^T\overline{\gamma}\big)^Tv=\big\langle\partial_m(\mathcal{P})^*\gamma,v\big\rangle, & \dfrac{\partial G_h}{\partial \overline{u}}=0. \end{cases} \tag{4.30}$$

Let us now formulate the problem with Lagrange multipliers, introducing two adjoint variables, $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$, for the two constraints,

$$\mathcal{L}(m,\tilde{u},\tilde{\gamma},\tilde{\gamma}_1,\tilde{\gamma}_2)=h_1(m,\tilde{u},\tilde{\gamma})+\big\langle\mathcal{P}(\tilde{u})-f,\tilde{\gamma}_1\big\rangle+\big\langle\mathcal{P}^*(\tilde{\gamma})+\mathcal{R}^*(\mathcal{R}(\tilde{u})-d),\tilde{\gamma}_2\big\rangle.$$

We now take u to be the actual solution of $\mathcal{P}(u) = f$ and γ the solution of $\mathcal{P}^*(\gamma) = -\mathcal{R}^*(\mathcal{R}(u) - d)$. Using the complex derivation Theorem 4.4, we have

$$\frac{\partial}{\partial m}\Big(\mathcal{L}(m, u, \gamma, \tilde{\gamma}_1, \tilde{\gamma}_2)\Big) = H(m)v = \mathrm{Re}\left(\frac{\partial \mathcal{L}}{\partial m} + \frac{\partial \mathcal{L}}{\partial u}\frac{\partial u}{\partial m} + \frac{\partial \mathcal{L}}{\partial \overline{u}}\frac{\partial \overline{u}}{\partial m} + \frac{\partial \mathcal{L}}{\partial \gamma}\frac{\partial \gamma}{\partial m} + \frac{\partial \mathcal{L}}{\partial \overline{\gamma}}\frac{\partial \overline{\gamma}}{\partial m}\right). \tag{4.31}$$

The adjoint states γ_1 and γ_2 are then selected to verify

$$\begin{cases} \mathrm{Re}\left(\dfrac{\partial \mathcal{L}}{\partial u}\dfrac{\partial u}{\partial m} + \dfrac{\partial \mathcal{L}}{\partial \overline{u}}\dfrac{\partial \overline{u}}{\partial m}\right) = 0, \\ \mathrm{Re}\left(\dfrac{\partial \mathcal{L}}{\partial \gamma}\dfrac{\partial \gamma}{\partial m} + \dfrac{\partial \mathcal{L}}{\partial \overline{\gamma}}\dfrac{\partial \overline{\gamma}}{\partial m}\right) = 0. \end{cases}$$

We can replace using the appropriate expression of Equation (4.30),

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial u} = \big\langle \partial_m(\mathcal{P})^*\gamma, v\big\rangle, \\ \dfrac{\partial \mathcal{L}}{\partial \overline{u}} = \mathcal{P}^*\gamma_1 + \mathcal{R}\mathcal{R}^*\gamma_2, \\ \dfrac{\partial \mathcal{L}}{\partial \gamma} = 0, \\ \dfrac{\partial \mathcal{L}}{\partial \overline{\gamma}} = \big(\partial_m(\mathcal{P})u\big)v + \mathcal{P}\gamma_2. \end{cases}$$

The last equation straightforwardly gives the adjoint state γ_2. For γ_1 we use the Corollary 4.1 to obtain

$$\mathrm{Re}\left(\frac{\partial \mathcal{L}}{\partial u}\frac{\partial u}{\partial m} + \frac{\partial \mathcal{L}}{\partial \overline{u}}\frac{\partial \overline{u}}{\partial m}\right) = \mathrm{Re}\left(\Big(\overline{\frac{\partial \mathcal{L}}{\partial u}} + \frac{\partial \mathcal{L}}{\partial \overline{u}}\Big)\frac{\partial \overline{u}}{\partial m}\right).$$

Therefore, the adjoint states have to satisfy

$$\begin{cases} \mathcal{P}^*\gamma_1 = -\mathcal{R}\mathcal{R}^*\gamma_2 - \big\langle \partial_m(\mathcal{P})^*\gamma, v\big\rangle, \\ \mathcal{P}\gamma_2 = -\big(\partial_m(\mathcal{P})u\big)v. \end{cases} \tag{4.32}$$

From this definition of the adjoint states, we have that $\partial_u\mathcal{L} + \partial_{\overline{u}}\mathcal{L} = \partial_\gamma\mathcal{L} + \partial_{\overline{\gamma}}\mathcal{L} = 0$, and from Equation (4.31) we deduce the Hessian vector product formulation:

$$H(m)v = \mathrm{Re}\left(\big\langle \big(\partial_m^2(\mathcal{P})u\big)^*\gamma, v\big\rangle + \big\langle \partial_m(\mathcal{P})u, \gamma_1\big\rangle + \big\langle \partial_m(\mathcal{P}^*)\gamma, \gamma_2\big\rangle\right).$$

4.5.2.2 Gauss-Newton Hessian multiplication

The Gauss-Newton approximation disregards the second order derivative of the forward problem, see the Equation (4.29). For the calculation of the Gauss-Newton Hessian vector product, we proceed similarly to the full Hessian with some adjustments for the definition

of the minimization problem. We consider

$$\min_m \quad h_2(m,u,w) \quad \text{subject to} \quad \mathcal{P}(u) = f.$$

Here h_2 is defined such that

$$h_2(m,u(m),w) = \langle u(m), w\rangle.$$

The derivation with respect to m gives

$$\frac{\partial}{\partial m}\Big(h_2(m,u(m),w)\Big) = \mathrm{Re}\left(\frac{\partial u(m)}{\partial m}^* w\right).$$

We write the minimization problem with Lagrange multiplier as

$$\mathcal{L}(m,\tilde{u},\tilde{\gamma}_1) = \langle \tilde{u}, w\rangle + \langle \mathcal{P}(\tilde{u}) - f, \tilde{\gamma}_1\rangle.$$

We take u solution of $\mathcal{P}u = f$ so that

$$\frac{\partial}{\partial m}\Big(\mathcal{L}(m,u,\tilde{\gamma}_1)\Big) = \frac{\partial}{\partial m}\Big(h_2(m,u,w)\Big) = \mathrm{Re}\left(\frac{\partial \mathcal{L}}{\partial m} + \frac{\partial \mathcal{L}}{\partial u}\frac{\partial u}{\partial m} + \frac{\partial \mathcal{L}}{\partial \overline{u}}\frac{\partial \overline{u}}{\partial m}\right). \tag{4.33}$$

The adjoint state γ_1 is selected to have the derivative with respect to u equal to zero,

$$\mathrm{Re}\left(\frac{\partial \mathcal{L}}{\partial u}\frac{\partial u}{\partial m} + \frac{\partial \mathcal{L}}{\partial \overline{u}}\frac{\partial \overline{u}}{\partial m}\right) = 0.$$

We note that

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial u} = 0, \\ \dfrac{\partial \mathcal{L}}{\partial \overline{u}} = \mathcal{P}^*\gamma_1 + w. \end{cases}$$

Hence the adjoint state solves the problem

$$\mathcal{P}^*\gamma_1 = -w. \tag{4.34}$$

We can finally inject back in the Equation (4.33) to get

$$\frac{\partial}{\partial m}\Big(h_2(m,u,w)\Big) = \mathrm{Re}\left(\langle \frac{\partial u}{\partial m}, w\rangle\right) = \mathrm{Re}\left(\langle \partial_m(\mathcal{P})u, \gamma_1\rangle\right), \tag{4.35}$$

where γ_1 depends on w from the problem (4.34).

Let us take w such that $w = \mathcal{R}^* D\mathcal{F}(m)v$, where v is a real vector. We recall that $D\mathcal{F} = \mathcal{R}\partial_m u$ and we see that

$$\begin{aligned} \langle \frac{\partial u}{\partial m}, w\rangle &= \langle \mathcal{R}^{-1} D\mathcal{F}(m), w\rangle = \langle \mathcal{R}^{-1} D\mathcal{F}(m), \mathcal{R}^* D\mathcal{F}(m)v\rangle \\ &= \langle D\mathcal{F}(m), D\mathcal{F}(m)v\rangle = D\mathcal{F}(m)^* D\mathcal{F}(m)v. \end{aligned}$$

The latest is the Gauss-Newton Hessian vector product,

$$H_{GN}(m)v = \mathrm{Re}\left(\langle \frac{\partial u}{\partial m}, \mathcal{R}^* D\mathcal{F}(m)v \rangle\right).$$

From the Equation (4.35) the Gauss-Newton Hessian can be expressed as

$$H_{GN}(m)v = \mathrm{Re}\left(\langle \partial_m(\mathcal{P})u, \gamma_1 \rangle\right), \tag{4.36}$$

where γ_1 is given by Equation (4.34) where w is now specified,

$$\mathcal{P}^*\gamma_1 = -\mathcal{R}^* D\mathcal{F}(m)v = -\mathcal{R}^*\mathcal{R}\frac{\partial u}{\partial m}v. \tag{4.37}$$

We still need to compute the right-hand side of Equation (4.37). We remind that u solves the problem $\mathcal{P}u = f$ so that the derivation with respect to m is

$$\mathcal{P}\partial_m u = -\partial_m \mathcal{P}u.$$

Here we recognize that $(\partial_m u)v$ is actually identical to the adjoint state γ_2 defined for the full Hessian vector product in the Equation (4.32) so that

$$-\mathcal{R}^* D\mathcal{F}(m)v = -\mathcal{R}^*\mathcal{R}\frac{\partial u}{\partial m}v = -\mathcal{R}^*\mathcal{R}\gamma_2.$$

Finally, in order to compute the Gauss-Newton Hessian vector product, two additional problems are required, where one is in common with the full Hessian situation,

$$\begin{cases} \mathcal{P}^*\gamma_1 = -\mathcal{R}^*\mathcal{R}\gamma_2, \\ \mathcal{P}\gamma_2 = -\big(\partial_m(\mathcal{P})u\big)v. \end{cases}$$

The matrix vector multiplication is then obtained from Equation (4.36).

4.5.2.3 Summary of the methods

The computation of Hessian vector product, or Gauss-Newton Hessian vector product, requires two additional forward problems to be solved, where one of the two is common to both situations. The computation of the gradient has already required the solution of two problems, in order to find the wavefield u and the first order adjoint state, γ (see Equation (4.21)). The Hessian vector multiplication generates the additional problems

$$\begin{aligned} \mathcal{P}\gamma_2 &= -\big(\partial_m(\mathcal{P})u\big)v, \\ \mathcal{P}^*\gamma_1^{FH} &= -\mathcal{R}\mathcal{R}^*\gamma_2 - \langle \partial_m(\mathcal{P})^*\gamma, v \rangle, \\ \mathcal{P}^*\gamma_1^{GN} &= -\mathcal{R}^*\mathcal{R}\gamma_2. \end{aligned} \tag{4.38}$$

Here γ_1^{FH} is for the full Hessian and γ_1^{GN} for the Gauss-Newton version. After the appropriate two second order adjoint states have been computed, the matrix vector product is appropriately chosen between

$$\begin{aligned} H(m)v &= \text{Re}\left(\left\langle \left(\partial_m^2(\mathcal{P})u\right)^*\gamma, v\right\rangle + \left\langle \partial_m(\mathcal{P})u, \gamma_1^{FH}\right\rangle + \left\langle \partial_m(\mathcal{P}^*)\gamma, \gamma_2\right\rangle\right), \\ H_{GN}(m)v &= \text{Re}\left(\left\langle (\partial_m\mathcal{P})u, \gamma_1^{GN}\right\rangle\right). \end{aligned} \tag{4.39}$$

4.5.3 Multi-parameters Hessian

We are able to compute the Hessian vector product for the full and Gauss-Newton Hessian when deriving with respect to the model parameter. However, one should take carefully the formulation (4.38) in the context of multi parameters. Indeed the number of parameters affects directly the derivative of the operator, hence the actual number of adjoint equations. Let us specify the Hessian for the acoustic and elastic seismic situation, decomposing the models of interest

$$\begin{aligned} H_{\text{acoustic}} &= \begin{pmatrix} H_{\lambda\lambda}(m) & H_{\rho\lambda}(m) \\ H_{\lambda\rho}(m) & H_{\rho\rho}(m) \end{pmatrix}, && m = \{\lambda, \rho\}, \\ H_{\text{elastic}} &= \begin{pmatrix} H_{\lambda\lambda}(m) & H_{\mu\lambda}(m) & H_{\rho\lambda}(m) \\ H_{\lambda\mu}(m) & H_{\mu\mu}(m) & H_{\rho\mu}(m) \\ H_{\lambda\rho}(m) & H_{\mu\rho}(m) & H_{\rho\rho}(m) \end{pmatrix}, && m = \{\lambda, \mu, \rho\}. \end{aligned} \tag{4.40}$$

Generally by denoting $m = \{m_1, \dots, m_{n_m}\}$ the n_m models of interest we have

$$H = \begin{pmatrix} H_{m_1m_1}(m) & H_{m_2m_1}(m) & \dots & H_{m_{n_m}m_1}(m) \\ H_{m_1m_2}(m) & H_{m_2m_2}(m) & \dots & H_{m_{n_m}m_2}(m) \\ \vdots & \vdots & \ddots & \vdots \\ H_{m_1m_{n_m}}(m) & H_{m_2m_{n_m}}(m) & \dots & H_{m_{n_m}m_{n_m}}(m) \end{pmatrix},$$

where every block is a squared matrix of size the number of coefficients to represent one of the models. Similarly, the vector that multiplies H can be decomposed by $v = \{v_1, \dots, v_{n_m}\}$ (here v_i is a vector of size the number of coefficients used to define the model). We obtain

$$H(m)v = \begin{pmatrix} \sum_i H_{m_im_1}v_i \\ \sum_i H_{m_im_2}v_i \\ \vdots \\ \sum_i H_{m_im_n}v_i \end{pmatrix}.$$

The Hessian vector multiplication that we obtained in the Equations (4.38) and (4.39) correspond to the computation of a single $H_{m_i,m_j}v_i$, where $m = m_i = m_j$. Let us specify

the case when $i \neq j$, for which the adjoint problems Equation (4.38) become

$$\begin{aligned}
\mathcal{P}\gamma_2 &= -\big(\partial_{m_j}(\mathcal{P})u\big)v_i, \\
\mathcal{P}^*\gamma_1^{FH} &= -\mathcal{R}\mathcal{R}^*\gamma_2 - \big\langle \partial_{m_j}(\mathcal{P})^*\gamma, v_i\big\rangle, \\
\mathcal{P}^*\gamma_1^{GN} &= -\mathcal{R}^*\mathcal{R}\gamma_2.
\end{aligned}$$

That is, γ_2 must be computed for every model m_j and vector v_i, giving n_m^2 possibilities, similarly for γ_1. The final formulation Equation (4.39) is

$$\begin{aligned}
< H_{m_i,m_j}(m), v_i > &= \mathrm{Re}\left(\big\langle \big(\partial^2_{m_i,m_j}(\mathcal{P})u\big)^*\gamma, v_i\big\rangle + \big\langle \partial_{m_i}(\mathcal{P})u, \gamma_1^{FH}\big\rangle + \big\langle \partial_{m_i}(\mathcal{P}^*)\gamma, \gamma_2\big\rangle\right), \\
< H_{GN,m_i,m_j}(m), v_i > &= \mathrm{Re}\left(\big\langle (\partial_{m_i}\mathcal{P})u, \gamma_1^{GN}\big\rangle\right).
\end{aligned}$$

Remark 4.8. *We see that the adjoint states hold for every model of interest and associated vector v. Thus the multi-parameter situation multiplies accordingly the number of adjoint problems to be solved. This is not the case for the first order adjoint state where the adjoint equation (4.22) is unique, whatever the number of models is.*

4.5.4 Conjugate gradient

The conjugate gradient method is an iterative algorithm to approximate the solution of $Ax = b$ without forming explicitly the matrix A, nor its inverse; it is a *matrix free* algorithm. The algorithm is based upon the use of matrix vector multiplications Av only, where A is symmetric positive definite matrix. The method is in particular reviewed in (Nocedal and Wright, 2006, Chapter 5) and well known in association with the Newton method, e.g., Le Dimet and Shutyaev (2000).

In the context of the Newton optimization, the search direction is selected to solve Equation (4.27), so that using the conjugate gradient avoids the formulation and inversion of the Hessian, but only requires Hessian vector multiplication, which we have established with the second order adjoint state formulation for the computation of (full or Gauss-Newton) Hessian vector multiplication. The method is well known in this context of optimization and is often referred to as *Hessian CG*. It is detailed for example, in (Nocedal and Wright, 2006, Section 5.1), and applied for seismic reconstruction in Métivier et al. (2013). We do not aim at providing much details but simply review the iterative method in the Algorithm 2, which refers to (Nocedal and Wright, 2006, Algorithm 5.2).

The criterion η is incorporated to define an acceptable approximation. Otherwise the number of iterations is limited to, say, n_{iter}. Note that the method gives the exact solution when the number of iterations matches the size of the matrix. However, one commonly

Objective: We aim to find x solution of the linear system $Hx = b$.
Initial guess: x_0
compute the matrix vector product $Hx_0 =: v_0$
compute $r_0 = v_0 - b$ and set $p_0 = -r_0$
for $k = 1, \ldots, n_{iter}$ **do**
  if $\|r_{k-1}\| < \eta$ *('low enough')* **then**
    exit k loop
  else
    compute the matrix vector product $Hp_k =: w_k$
    $\alpha_k = \dfrac{r_k^T r_k}{p_k^T w_k}$
    $x_{k+1} = x_k + \alpha_k p_k$
    $r_{k+1} = r_k + \alpha_k w_k$
    $\beta_{k+1} = \dfrac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$
    $p_{k+1} = -r_{k+1} + \beta_{k+1} p_k$
  end
end
end of algorithm with $x := x_k$

Algorithm 2 – Conjugate gradient (CG) algorithm: approximate the solution of linear systems avoiding the explicit formulation and inversion of the matrix. The method only needs matrix vector multiplication. The tolerance η acts as a threshold for acceptable approximation, as an alternative, a maximum number of iterations n_{iter} is performed.

processes only a few iterations.

4.5.5 Numerical cost

The Hessian vector multiplication requires two problems to be solved per model of interest and per vector associated with each of them, which gives a total of $2n_m^2$ adjoint states, with n_m the number of models of interest. Similarly to the first order adjoint state, it has to be repeated for every source of the acquisition but direct solvers, using the factorization obtained for the forward problem and benefiting from multi right-hand sides options, can reduce the numerical cost.

However, when we incorporate the method in the conjugate gradient framework, with, say n_{iter} iterations, the number of system to be solved is now $2n_{iter}n_m^2$. We see that multi-parameters inversion has drastic effect on the numerical cost. For example, assume an elastic isotropic inverse problem where the medium is defined with three parameters (the density and the Lamé parameters), say we want to perform twenty iterations of the conjugate gradient (which is not incredibly high), it already gives an additional 360 problems. At this point, even if the factorization has not to be conducted and multi right-hand sides provides

a profound benefit, the cost may not worth the effort. Namely, increasing the amount of gradient descent iterations may be a better substitution, as suggested by Tarantola (2005). On the other hand, the Hessian can give major insight into the subsurface properties, Pratt et al. (1998), and one should be careful when reducing the computational time at all cost.

4.5.6 Higher order adjoint state

We have extended the adjoint state formulation to arbitrary order, in a joint work with Jia Shi, Maarten V. de Hoop and Henri Calandra, see Shi et al. (2014); particularly when the data are represented with the Neumann-to-Dirichlet map, following land seismic situation. The reader is referred to this additional work for more information.

4.6 Full waveform inversion algorithm options

Many methods exist to conduct iterative minimization algorithm and many can be found, for example, in Nocedal and Wright (2006). The computation of the gradient of the cost function at each iteration seems a minimal requirement (we do not account for the *gradient free* techniques arising in the stochastic approach here). Second order information can provide tremendous help but its access may be limited due to numerical restriction. As an alternative the family of Quasi-Newton method can be employed to conduct the iterative algorithm using approximation of the Hessian. Several methods have been designed to approximate the Hessian while avoiding the computation and inversion of the full matrix, and we have given the Hessian CG in Subsection 4.5. In the context of seismic reconstruction, it is mentioned in (Tarantola, 2005, p. 158) that the benefit of Newton method is not obvious and one can prefer performing more iterations of steepest descent algorithm. Yet, Pratt et al. (1998) advocate the necessity of the Hessian and gives an understanding on the information it carries in seismic.

Here we review the methods that have been employed during this project for FWI, where some of the first and second order techniques have also been compared. In particular, we detail the methods that require only the computation of the gradient in Subsection 4.6.1. The formulation of the Hessian approximation using conjugate gradient (Hessian CG), given in Subsection 4.5, is briefly reminded.

4.6.1 Nonlinear conjugate gradient, limited-BFGS

The limited use of first order derivative information via the gradient is primarily motivated by the large size of the seismic inverse problems where the computational time becomes critical. Indeed when one considers large three-dimensional media and multi-parameters inversion, the computational cost is understandably vastly increasing. That is why the gradient descent algorithm (and its variations) may be preferred instead of more standard Newton method. It consists in updating the model using the gradient only such that the updated model is defined by

$$m_{i+1} = m_i - \alpha \nabla \mathcal{J}(m_i). \tag{4.41}$$

Here, α is the step length, typically approached using line search algorithm which we later depict in Subsection 4.6.3. This method can also be interpreted as a Landweber iteration scheme, Landweber (1951); Kirsch (1996); Hanke (2014); de Hoop et al. (2012).

4.6.1.1 Nonlinear conjugate gradient

Following the gradient descent, an alternative approach can be developed, still being based on the gradient only. In particular, nonlinear conjugate gradient (NLCG) methods use the gradient computed at previous iterations to improve the direction of minimization. It relies on an additional parameter which definition can follow different formulations. Here we have used the four main formulations: Fletecher–Reeves (FR, Fletcher and Reeves (1964)), Polak–Ribière (PR, Polak and Ribiere (1969)), Hestenes–Stiefel (HS, Hestenes and Stiefel (1952)) and more recently Dai–Yuan (DY, Dai and Yuan (1999)). In the Algorithm 3, we show the modification of the descent direction generated by the NLCG method.

Numerically speaking, the NLCG method does not require heavy computation because it only requires simple operations. However, it necessitates to keep the gradient at previous iterations, which can be expensive memory-wise. A simple alternative is to store the gradients on the disk, in order to save the computational memory.

4.6.1.2 Limited-BFGS

The BFGS method aims to estimate the inverse of the Hessian and corresponds with a quasi-Newton method. It is named after Broyden, Fletcher, Goldfarb and Shanno and their series of papers in 1970's. In order to reduce the memory required, Nocedal (1980) introduces a version which is now referred to as limited-memory BFGS, limited-BFGS, or

- Gradient for iteration i, $g_{m_i} = -\nabla_m \mathcal{J}(m_i)$
- Compute parameter β following one of the following definition:

$$\left|\begin{array}{rl} \text{Fletcher–Reeves formula} & \beta_i = \dfrac{(g_{m_i})^T g_{m_i}}{(g_{m_{i-1}})^T g_{m_{i-1}}}, \\ \text{Polak–Ribière formula} & \beta_i = \dfrac{(g_{m_i})^T (g_{m_i} - g_{m_{i-1}})}{(g_{m_{i-1}})^T g_{m_{i-1}}}, \\ \text{Hestenes–Stiefel formula} & \beta_i = -\dfrac{(g_{m_i})^T (g_{m_i} - g_{m_{i-1}})}{s_{i-1}^T (g_{m_i} - g_{m_{i-1}})}, \\ \text{Dai–Yuan formula} & \beta_i = -\dfrac{(g_{m_i})^T g_{m_i}}{s_{i-1}^T (g_{m_i} - g_{m_{i-1}})}. \end{array}\right.$$

- Compute the conjugate direction $s_i = g_{m_i} + \beta_i s_{i-1}$
- Perform line search algorithm to find the step $\alpha = \arg\min_\alpha \mathcal{J}(m_i + \alpha s_i)$
- Update the model $m_{i+1} = m_i + \alpha s_i$

Algorithm 3 – Nonlinear conjugate gradient (NLCG) algorithm: it requires the computation of the gradient to update the model in an iterative optimization framework. The first conjugate direction is given by $s_1 = -\nabla_m \mathcal{J}(m_1)$ and the NLCG can be employed for $i >= 2$.

L-BFGS method. In the Algorithm 4, we give the steps for the model update.

4.6.2 Hessian method

In our application, we only employ the Hessian CG method to account for the second order information. The method has been depicted in Subsection 4.5.4, following the formulation of Hessian vector product. We also employ a regularization term on the diagonal, see Remark 4.6. In Section 4.7, we experimentally compare the effect of reconstruction using the gradient and the Hessian.

Several possibilities have been proposed in the literature to account for the Hessian or its approximation and we have mentioned some applications in the introduction already. The implementation of these promising techniques is clearly an evolution we need to consider for the short term future of our code.

- We define at iteration i,

$$\begin{aligned} g_{m_i} &= \nabla_m \mathcal{J}(m_i), \\ \Delta_{m_i} &= m_{i+1} - m_i, \\ \Delta_{G_i} &= g_{m_{i+1}} - g_{m_i}, \\ \gamma_i &= \frac{1}{\Delta_{G_i}^T \Delta_{m^i}}. \end{aligned}$$

- Set $q = g_{m_i}$

for $k = (i-1), (i-2), \ldots, (i-l)$ **do**

  $\alpha_k = \gamma_k \Delta_{m_k}^T q$

  $q = q - \alpha_k \Delta_{G_k}$

end

- $W_i = \dfrac{\Delta_{G_{i-1}}^T \Delta_{m^{i-1}}}{\Delta_{G_{i-1}}^T \Delta_{G_{i-1}}}$
- $s_i = W_i q$

for $k = (i-l), (i-l+1), \ldots, (i-1)$ **do**

  $\beta_k = \gamma_k \Delta_{G_k}^T s_i$

  $s_i = s_i + \Delta_{m_k}(\alpha_k - \beta_k)$

end

- Perform line search algorithm to find the step $\alpha = \arg\min_\alpha \mathcal{J}(m_i - \alpha s_i)$
- Update the model $m_{i+1} = m_i - \alpha s_i$

Algorithm 4 – Limited-Memory BFGS (L-BFGS) algorithm: we introduce l, a chosen integer. W acts as the approximate of the Hessian inverse.

4.6.3 Line search method

When the update direction has been computed with the method of choice (gradient descent, nonlinear conjugate gradient, Hessian CG, etc), it is common to adjust the direction with an appropriate step length. If we refer by s the update direction, the updated model is written

$$m_{i+1} = m_i - \alpha_i s_i,$$

and α is the step length (as in Equation (4.41) for the gradient descent). It is formally defined by

$$\alpha_i = \arg\min_\alpha \mathcal{J}(m_i - \alpha s_i).$$

For simplicity and to reduce the number of evaluation of the misfit function $\mathcal{J}$, one can limit the search to the first α that decreases the cost function instead of looking for the actual minimum. However, some conditions must be respected by the step length: these are the Wolfe and Armijo conditions (Armijo, 1966; Wolfe, 1969, 1971). Those conditions guarantee in some sense that the step α would not be too large or too small.

– The Armijo condition is

$$\mathcal{J}(m_i - \alpha s_i) \leq \mathcal{J}(m_i) + K_1 \alpha s_i^T \nabla \mathcal{J}(m_i).$$

– The Wolfe condition is

$$s_i^T \nabla \mathcal{J}(m_i - \alpha s_i) \geq K_2 s_i^T \nabla \mathcal{J}(m_i).$$

As it is mentioned in Chavent (2010), the latest Wolfe condition can be replaced by the Goldstein condition.

– The Goldstein condition (as an alternative to the Wolfe condition) is

$$\mathcal{J}(m_i - \alpha s_i) \geq \mathcal{J}(m_i) + K_2 \alpha s_i^T \nabla \mathcal{J}(m_i).$$

In the above relations, K_1 and K_2 are constants such that $K_1 < K_2 < 1$.

In our iterative scheme the line search is conducted via a backtracking algorithm for simplicity. We start with a relatively high initial step $\alpha_i^{(0)}$ and it is decreased until the cost function decreases correctly. Hence every iteration for $\alpha_i^{(l)}$ only requires the evaluation of the cost function. This gives us a computationally inexpensive formula. However, better alternatives are possible, yet they require more computations. We mention in particular the Maximum Projected Curvature (MPC) step of Chavent (2010). Finally, de Hoop et al. (2012) give the analytic step when considering for Landweber iteration scheme. For more details about the line search, we refer to (Nocedal and Wright, 2006, Chapter 3).

4.6.4 Source inversion

The experimental source(s) employed to generate the wave is not exactly known. One can expect to have the time domain source wavelet with relative accuracy but one cannot expect the exact behavior. This is due to noise, limited accuracy of the devices, etc. That is why it is necessary to approximate the source simultaneously to the model update. Naturally defined in the time domain, the initial frequency domain source corresponds with the Fourier transform of the wavelet, giving a complex number associated with the current frequency. The problem is then to recover the source amplitude (and not the spatial position). It is relatively easy to handle because it is a linear inverse problem. Indeed, the source acts typically at the right-hand side of the linear system describing the wave phenomenon. In geophysics, the source update has been given in particular by Pratt (1999); Virieux and Operto (2009) who define

$$g_{i+1} = g_i \frac{\langle \mathcal{R}u(m_i), d \rangle}{\langle \mathcal{R}u(m_i), \mathcal{R}u(m_i) \rangle},$$

where g_i stands for the source at iteration i. Hence we update simultaneously the model and the source value. Even if the initial source wavelet, is accurate, it is 'safer' to adjust the source with the iterations.

4.6.5 Complex frequency

In our algorithm the angular frequency is denoted by ω. Let us introduce the notation for complex frequency, $\tilde{\omega}$, such that the Helmholtz equation becomes

$$-\big(\frac{\tilde{\omega}^2}{c^2(\mathbf{x})} - \Delta u(\mathbf{x})\big) = f(\mathbf{x}).$$

We define $\tilde{\omega}$, following the introduction made in the Remark 1.7, such that

$$-\tilde{\omega}^2 = (\sigma + 2\mathrm{i}\pi f)^2,$$

where f is the frequency in Hz. We see that in the case where $\sigma = 0$ we have the traditional angular frequency and $-\tilde{\omega}^2 = -(2\pi f)^2 = -\omega^2$.

This approach has been popularized for the purpose of seismic inversion in the work of Shin and Cha (2009); Ha et al. (2010) with reference to the Laplace and Laplace-Fourier domain. Indeed, if $f = 0$, the resulting wave equation can be understood as a Laplace transform of the time domain formulation. Yet, in the following of this manuscript we prefer to use the term complex frequency which is more general. Several applications have been carried out since then, we mention for example the work of Shin et al. (2010); Petrov and Newman (2014) where the benefits of using such frequencies are illustrated. We will later study the effect of using complex frequencies, which are particularly useful for the reconstruction of subsurface salt body of high velocity contrast, see Chapter 6. We also justify the usefulness of such approach by the study of the size of the basin of attraction of the optimization problem in Chapter 5. This justification gives a clear understanding on why complex frequency can provide benefits. However, the data at such frequencies suffer from the noise an may be inaccessible in seismic.

Remark 4.9. *By abuse of notation we will keep ω as the general frequency notation in the equations for the rest of the manuscript. It is justified for simplicity, but always represents a possibly complex frequency.*

4.6.6 Detailed algorithm

We review the general scheme to conduct the iterative minimization in the Algorithm 5, where the model update is detailed. In the Appendix B we further give the major options

that have been implemented in the FWI toolbox and how the software works. This toolbox have been created in `Fortran90` and designed for parallel architecture using `MPI` and `OpenMP`.

Data:
- Acquisition set (sources and receivers positions),
- frequency domain measurements,
- selection of options (norms of the cost function, search direction method, etc),
- initial model m_1.

for $k \in \{1, \ldots, n_\omega\}$ **do**
  for $i \in \{1, \ldots, n_{iter}\}$ **do**
    $j = i + (k-1) \times n_{iter}$
    for $s \in \{1, \ldots, n_{src}\}$ **do**
      Numerical resolution of the wave equation using m_j
    end
    Apply the appropriate restriction operator $\mathcal{R}$ to deduce $\mathcal{F}_{\omega_k}(m_j)$
    Compute the residuals and underlying cost function $\mathcal{J}(m_j)$
    if $(\mathcal{J}(m_j) < \eta_1)$ **then**
      end iteration loop
    else
      Solve the adjoint state problem
      Compute the derivative of the wave operator with respect to the model(s)
      Compute the gradient with respect to the model(s)
      Compute the search direction s_i:
      if *Hessian CG method* **then**
        for $l \in \{1, \ldots, n_H\}$ **do**
          Apply the conjugate gradient algorithm from Hessian vector product method
        end
      else
        Non-linear conjugate gradient formula of choice
      end
      Line search to estimate the step length α_j
      Update the model $m_{j+1} = m_j + \alpha_j s_j$
      if $(\mathcal{J}(m_{j+1}) - \mathcal{J}(m_j) < \eta_2)$ **then**
        end iteration loop
      end
    end
  end
end

Algorithm 5 – Detailed FWI algorithm: FWI iterative minimization in the frequency domain. The model update is conducted using the gradient descent or Hessian CG, $\eta_2 > 0$ prevent excessive iterations in case of the stagnation of the cost function; $\eta_1 \geq 0$ provides a tolerance for an acceptable convergence. Appendix B provides the detailed FWI toolbox created in `Fortran 90` according to seismic inversion.

4.7 Seismic reconstruction from gradient and Hessian

In this section, the iterative minimization algorithm is carried out using synthetic experiments. Among the options that have been presented, only the use of first order information (through gradient descent and NLCG variation), or second order (with the Hessian conjugate gradient method) are compared. The Hessian should naturally provide the wanted framework for the resolution of optimization problem following the Newton method, however, the numerical cost involved by the large scale seismic problem handled may undermine the benefits. That is why it is important to identify the situations in which the Hessian gives a major breakthrough in the recovery.

Here we design experiments for acoustic media with constant or variable density. The study should be pursued for elastic media as well, which certainly requires heavier computation but where the benefits may be crucial. More generally, we hereby illustrate one of the central features of our seismic inverse problem software. We review the different possibilities of nonlinear conjugate gradient methods for gradient descent algorithm and integrate a regularization term in the Hessian computation. For the latest we clearly detect its impact, but it requires more meticulous consideration and a precise definition.

Remark 4.10. *In this section we have used the slowness squared parametrization ($1/c^2$) to conduct the acoustic iterative minimization with constant density. For the variable density, we have selected to invert $(1/\kappa, 1/\rho)$. The choice of parametrization is later analyzed, in the Chapter 5, Sections 5.4, 5.5 and 5.6.*

4.7.1 The acoustic Marmousi 2D model with constant density

We consider the two-dimensional acoustic Marmousi model, of size 9.2×3 km. It has been synthetically designed by the Institut Français du Pétrole (IFP) in 1988 and has since become the most popular geophysical model for experiments. It has been thoroughly studied, as in the work of Martin (2004); Martin et al. (2006), whom further extend the model to elastic media (we will later use this elastic version for experiments). The wave speed varies from 1450 to 5500 $\mathrm{m\,s^{-1}}$ and is shown in Figure 4.4. It consists of different sedimentary structures with different celerities in every layer. We further assume a constant density with value 1000 $\mathrm{kg\,m^{-2}}$. Hence, the propagation of wave follows the Helmholtz equation and the inverse problem aims the recovery of the wave speed.

The seismic acquisition is designed with sources and pressure measurements located in the near surface area. We impose 91 sources that are generated one by one. For every

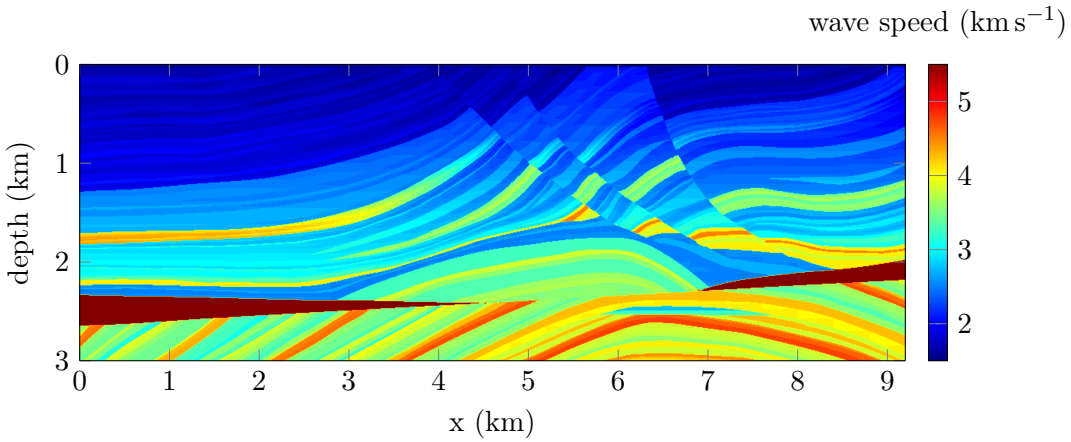


Figure 4.4: Wave speed for the acoustic Marmousi medium of size 9.2×3 km, the velocity varies between 1450 and 5500 m s^{-1}.

source, 183 receivers record the corresponding pressure data. The sources are equally spaced by 100 m and the receivers by 50 m. In this experiment, synthetic data are generated in the frequency domain using a different discretization and order of approximation compared to the one used during the inverse algorithm. Namely, the FWI algorithm employs a Discontinuous Galerkin discretization with order 2 polynomials while the data use an order 3, also the mesh is refined for the data generation. The data are generated using frequencies from 2 to 10 Hz.

From the data (pressure recorded at the receivers location for the frequencies) we aim at reconstructing the velocity without knowing any information on the subsurface structures. The model used to initiate the iterative minimization is presented in Figure 4.5. We simply consider a one-dimensional variation of velocity, with increasing value with depth.

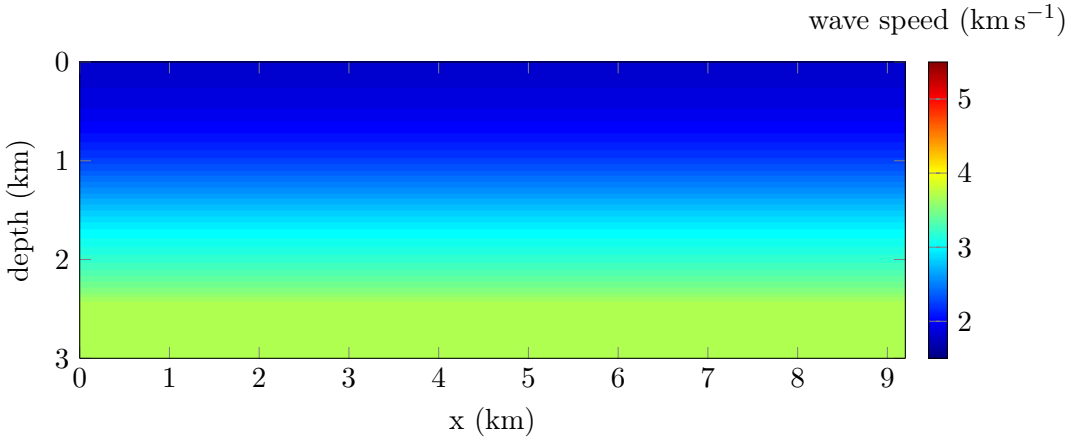
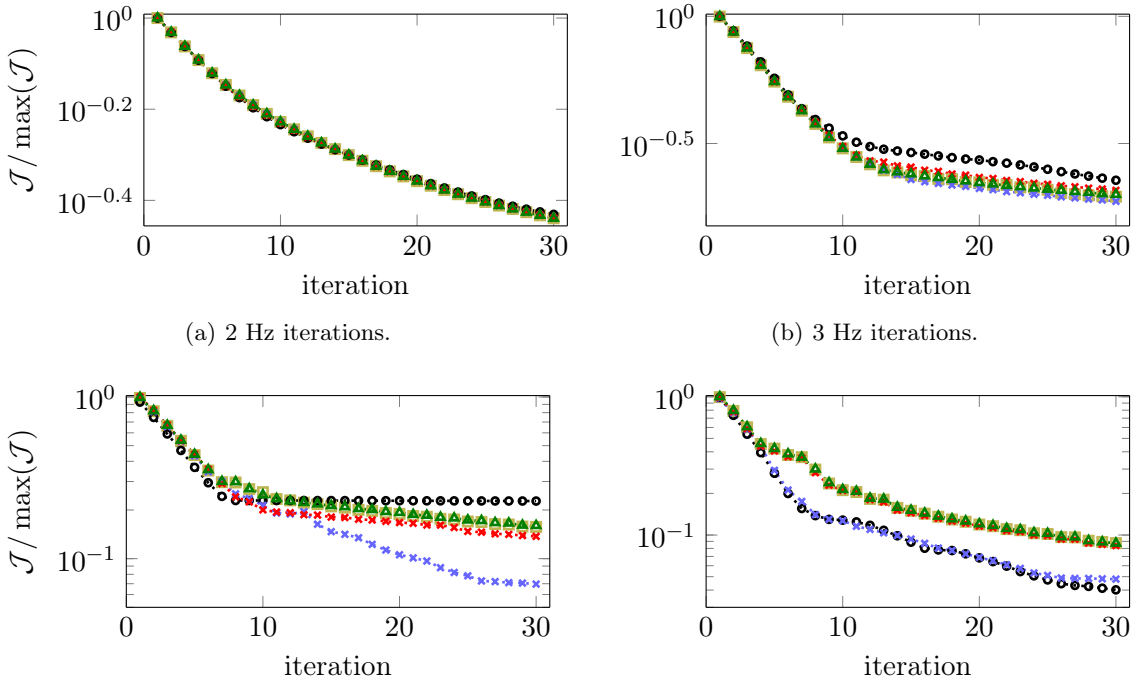


Figure 4.5: Initial wave speed for the iterative minimization algorithm and reconstruction of the Marmousi model. The model has no information on any of the true structures and consists of a one-dimensional variation of velocity with depth.

4.7.1.1 Gradient descent and nonlinear conjugate gradient

We iterate following the Algorithm 5. The set of frequency varies from 2 to 10 Hz with 1 Hz increment. We perform 30 iterations per frequency. We first review the reconstruction using only first order information (i.e. the gradient). In particular, we compare the gradient decent with the four nonlinear conjugate gradient formulas: Fletcher–Reeves (FR), Polak–Ribière (PR), Hestenes–Stiefel (HS) and Dai–Yuan (DY), see Subsection 4.6.1.

In the Firgure 4.6, we compare the evolution of the cost function with iterations for different frequencies. It allows a quantification of the methods efficiency in terms of residual diminution and convergence. We provide the results at the beginning of the inversion scheme (2 and 3 Hz), and for later iterations (6 and 8 Hz).



(a) 2 Hz iterations.

(b) 3 Hz iterations.

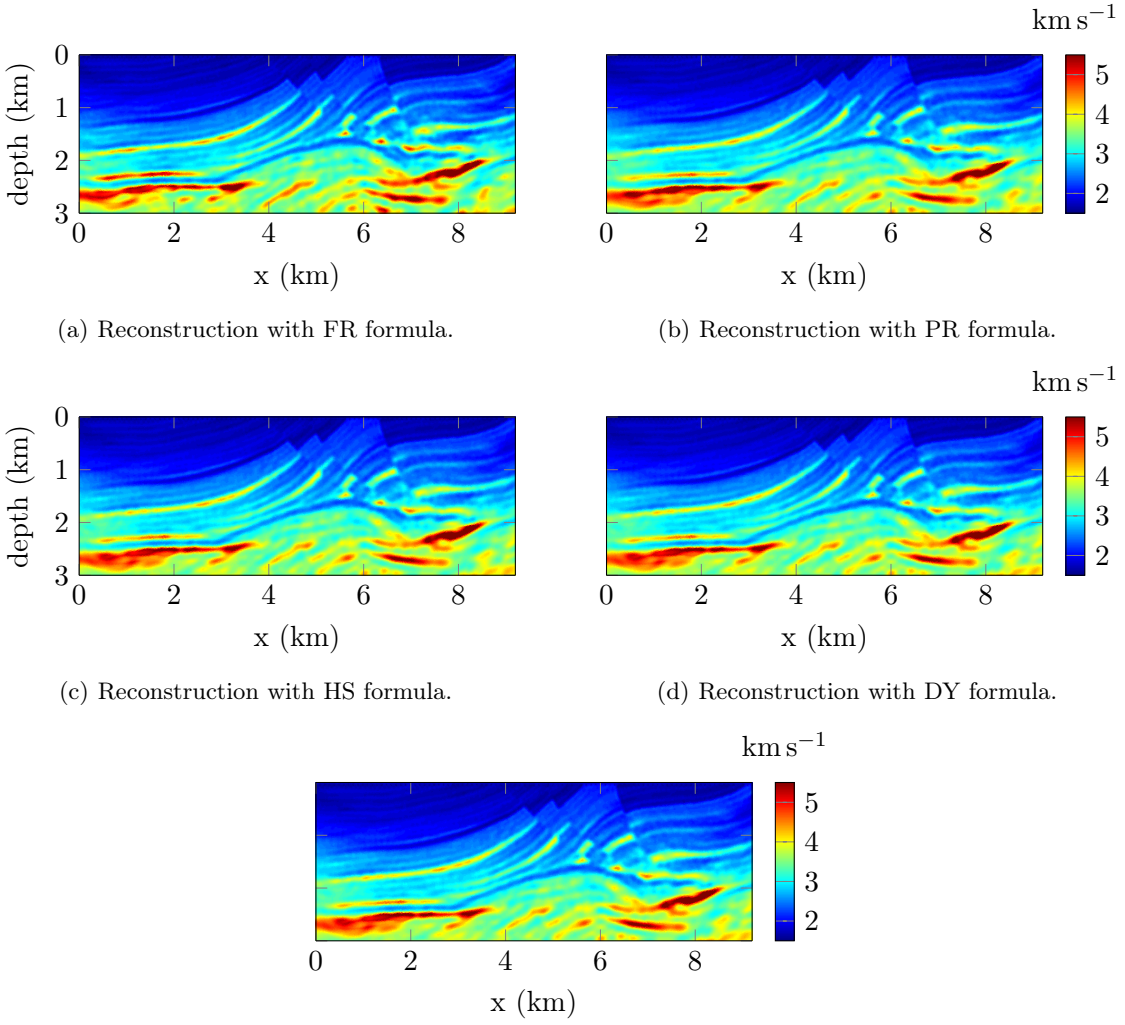
(c) 6 Hz iterations.

(d) 8 Hz iterations.

Figure 4.6: Evolution of the cost function for different frequencies using the nonlinear conjugate gradient method. The black circles (··o··) use the Fletcher–Reeves (FR) formula, the blue crosses (··×··) the Polak–Ribière (PR) formula, the red crosses (··×··) the Hestenes–Stiefel (HS) formula and the green triangles (··△··) the Dai–Yuan (DY) formula. The yellow squares (··□··) uses the gradient descent algorithm. At every frequency, the cost function progression is scaled with the value of the first iteration.

The initial frequency iterations (2 Hz) do not show any difference between the formulations. The second set of iterations (3 Hz) shows a slight improvement in the convergence rate with the PR formulation of the NLCG method. For the latest frequencies, it appears that the tendency is confirmed with the PR formulation (··×··) that gives the best

convergence. Namely, at 6 Hz, while the other formulations start to stagnate after about 15 iterations, the PR formula keeps improving the cost function drastically. It is interesting to notice that the FR formulation gives the worst convergence at 6 Hz but shows major improvement for the 8 Hz iterations (without being impacted by the previous slow convergence). The gradient descent presents very similar results between the HS and YD formulations. Overall, all methods seem to converge relatively nicely for the first 10 to 20 iterations and then the decrease of the cost function slows down. In Figure 4.7, we compare the final wave speed reconstructed from the different methods.



(a) Reconstruction with FR formula.

(b) Reconstruction with PR formula.

(c) Reconstruction with HS formula.

(d) Reconstruction with DY formula.

(e) Reconstruction with gradient descent.

Figure 4.7: Final reconstructions of the acoustic Marmousi wave speed assuming constant density and using frequencies from 2 to 10 Hz. 30 iterations are performed per frequency (for a total of 270 iterations) and we compare the methods involving only the gradient of the cost function.

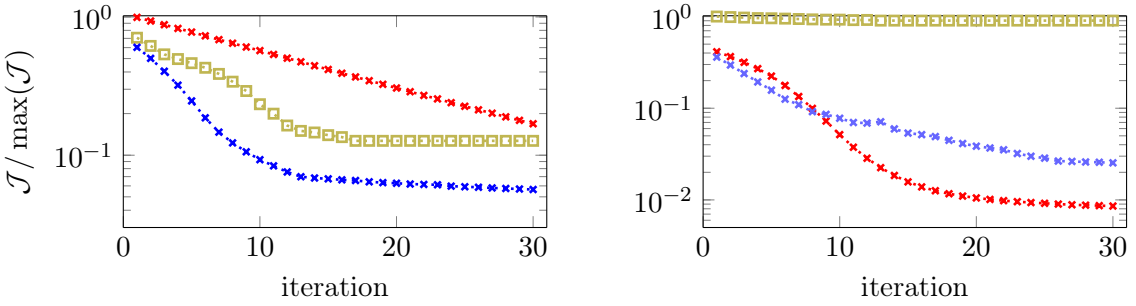
The iterative minimization algorithm accurately recovers the Marmousi wave speed with comparable accuracy for all the methods involving the gradient of the cost function

only. The different layers are well retrieved with the appropriate velocity values, except for the deepest parts of the model. As it was expected by the cost function evolution presented Figure 4.6, the PR formula slightly improves the reconstruction, in particular with the deepest central structures. However, in this first experiment, all methods behave pretty well.

4.7.1.2 Hessian approximation using conjugate gradient

Following the result using the first order information, we proceed to the same experiment but using techniques that involved the second order information (Hessian). In particular, we employ the Newton CG where the search direction given by the (full or Gauss-Newton) Hessian is approximated by the conjugate gradient method, see Section 4.5. Compared to the gradient descent (and the NLCG alternatives), the general FWI Algorithm 5 incorporates a second loop for the conjugate gradient approximation of the Hessian inversion. We take a reasonable number of iterations to limit the computational time, i.e., 10 (and exact resolution is not the objective of the method anyway).

We compare the evolution of the cost function at selected frequencies using the full or Gauss-Newton Hessian approximation. We also show the results where the NLCG is used with the Polak–Ribière formula. Figure 4.8 presents the evolution of the residuals for two frequencies, and we remind that the complete set is a sequential progression from 2 to 10 Hz.



(a) 4 Hz iterations.

(b) 6 Hz iterations.

Figure 4.8: Evolution of the cost function for different frequencies using different methods for the model update. The blue crosses (···×···) use the Polak–Ribière (PR) nonlinear conjugate gradient. The red crosses (···×···) use the Gauss-Newton Hessian CG approximation. The yellow crosses (···□···) use the full Hessian CG approximation. The number of iterations for the conjugate gradient is set to 10.

During the initial iterations (at low frequencies) the NLCG provides the best results, while the Gauss-Newton approximation shows a surprisingly slow convergence rate. The

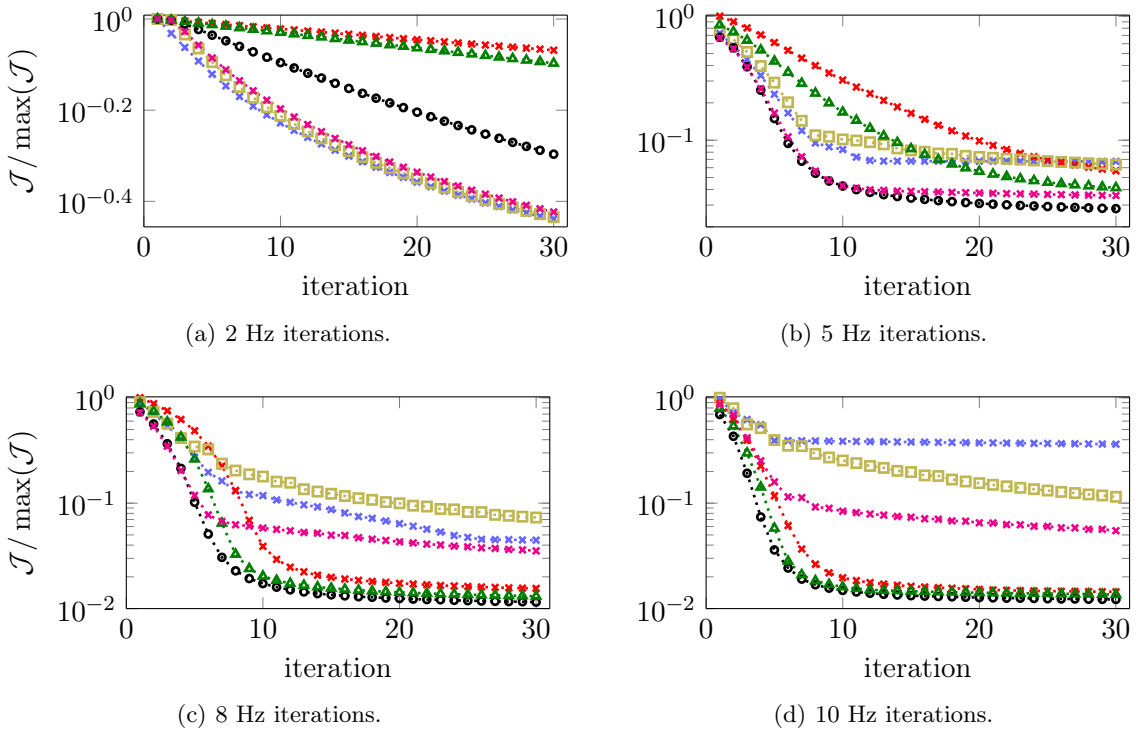
later iterations give a different situation where the nonlinear conjugate gradient stagnates meanwhile the Gauss-Newton approximation keeps converging. In this case, the method based on the full Hessian does not converge for the higher frequencies. Hence the benefit of using Hessian information is not straightforward for this acoustic situation. Because of the numerical cost involved, one can imagine to iterate first using the gradient only and then the approximation of the Hessian for the latest iterations.

Regarding the Hessian, it can be important to consider regularization to improve the conditioning of the Hessian, as explained in the Remark 4.6. Here we decide to incorporate a term on the diagonal, following a Levenberg–Marquardt approach. We recall that the Hessian 'becomes' $H + \beta I_d$, where I_d is the identity matrix and β a scalar. Hence, we simply consider a constant damping on the diagonal for every iteration, with value

$$\beta_i = \frac{\|H_i(m_i - m_{i-1}) - (\nabla \mathcal{J}_i - \nabla \mathcal{J}_{i-1})\|}{\|m_i - m_{i-1}\|}, \tag{4.42}$$

where $\nabla \mathcal{J}_i$ indicates the gradient of the cost function at iteration i, m is the model of interest. It is certain that a more profound study is required to optimize the use of the preconditioner, benefiting from the existing theory. However, the choice of an appropriate value for β is not by any means a simple task, see Kern (2016), especially for large scale optimization problems. The point is here to simply illustrate its importance. In Figure 4.9, we compare the residuals evolution for the Gauss-Newton Hessian for different values of the preconditioner.

The value of the damping parameter on the diagonal of the Hessian (Equation (4.42)) plays an important role, as expected. If this coefficient is too low, then the results are naturally close to the original Hessian approximation. Increasing the damping generates a variety of possibilities. It is important to notice that the well chosen preconditioner eventually gives better convergence than either the natural Hessian or the gradient, which indicates its usefulness. However, the choice of this diagonal coefficient may be complex, here it is clear that to obtain the best convergence rate, it needs to be adapted at each frequency. At early frequencies, using the value β directly seems to be sufficient for the convergence, yet when increasing the frequency, the preconditioner with value $10^{-3}\beta$ or $10^{-4}\beta$ provides a much better decrease of the misfit functional. From these comparisons we conclude that the preconditioner is clearly needed to improve the convergence of the method, however, it would require a much better understanding to clarify how to choose it and make it evolved with frequency.



(a) 2 Hz iterations.

(b) 5 Hz iterations.

(c) 8 Hz iterations.

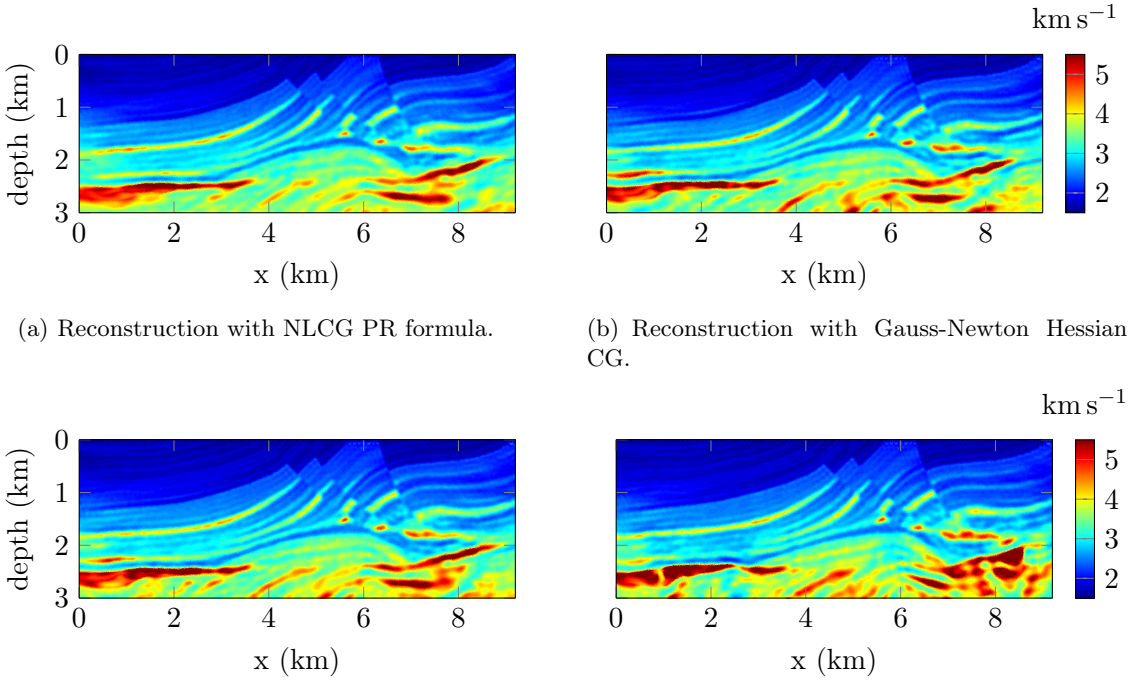
(d) 10 Hz iterations.

Figure 4.9: Evolution of the cost function for different frequencies using different methods. The blue crosses (···×···) use the non-linear conjugate gradient Polak–Ribière formula. The red crosses (···×···) use the Gauss-Newton Hessian and 10 conjugate gradient iterations. The yellow squares (···□···) use a diagonal preconditioner with value β of Equation (4.42). The black circles (···○···) use a diagonal preconditioner with value $10^{-3}\beta$. The green triangles (···△···) use a diagonal preconditioner with value $10^{-4}\beta$. The magenta crosses (···×···) use a diagonal preconditioner with value $10^{-2}\beta$. All methods using preconditioner use the Gauss-Newton Hessian and 10 conjugate gradient iterations.

4.7.1.3 Final reconstruction using 75 iterations per frequency

We continue the Marmousi experiments increasing the number of iterations per frequency, from 30 to 75. We compare the final results of the FWI algorithm, where the frequency still varies from 2 to 10 Hz, with 1 Hz increment. The reconstructions obtained from the NLCG, Gauss-Newton and full Hessian CG are illustrated in Figure 4.10. We also incorporate a preconditioner for the Hessian.

The general accuracy of the reconstruction is relatively good and some differences clearly appear in the deepest structures. The use of Hessian information captures deeper information better than when we restrict the algorithm to the use of the gradient. Here the preconditioner provides a more precise resolution. In particular, for the Gauss-Newton Hessian CG method. Finally, the full Hessian gives larger values than expected in this example.



(a) Reconstruction with NLCG PR formula.

(b) Reconstruction with Gauss-Newton Hessian CG.

(c) Reconstruction with Gauss-Newton Hessian CG and preconditioner.

(d) Reconstruction with full Hessian CG and preconditioner.

Figure 4.10: Final reconstruction of the acoustic Marmousi wave speed assuming constant density and using frequencies from 2 to 10 Hz. 75 iterations are performed per frequency (for a total of 675 iterations), and we compare the methods involving the gradient and the Hessian of the cost function.

In this experiment, the starting model has no prior information on the velocity and yet the algorithm is capable of recovering the main velocity structures. The acoustic Marmousi test case may be a simple one but it still reveals how the use of Hessian information can improve the reconstruction. Naturally, the numerical cost can easily become extreme, even in this simple acoustic isotropic situation. In Table 4.3, we provide the total time to run the algorithm for the different methods. The computational time given should be mitigated by the fact that we have used only 4 processors[2] and could certainly be reduced for this reason. However, there is still an intrinsic cost in time for approximating the Hessian. A mixed algorithm with initial iterations using the gradient and the use of Hessian information for a few set of final iterations would be an appropriate compromise.

Remark 4.11. *The Marmousi acoustic reconstruction is continued in Chapter 6, Section 6.2, in particular with the use of time domain noisy data.*

[2]The experiments have been conducted on the Purdue Conte clusters, https://www.rcac.purdue.edu/compute/conte/, specifications: 16 cores per node with two 8-Core Intel Xeon-E5; 64 GB per node; Infiniband: 40 Gbps FDR10.

Number of processors	Iterations	Nd.o.f.	NLCG	GNH CG	GNH CG preconditioner
4	675	734 022	8h	60h	75h

Table 4.3: Computational time for the FWI algorithm in the reconstruction of the acoustic Marmousi model with constant density. The experiments have been running on the Purdue Conte Cluster. Nd.o.f. indicates the number of degrees of freedom for the forward problem.

4.7.2 The acoustic Overthrust 2D model with constant density

We now consider the two-dimensional Overthrust (OT) model. It is actually originated from the SEG-EAGE Overthrust 3D velocity model of Aminzadeh et al. (1994), where a two-dimensional section has been extracted. We will proceed similarly to the Marmousi test case, in order to verify the experimental observations we have on the different methods of inversion. This model is of size 20×4.65 km and the velocity ranges from 2000 to 6000 $\mathrm{m\,s^{-1}}$. It is pictures in Figure 4.11.

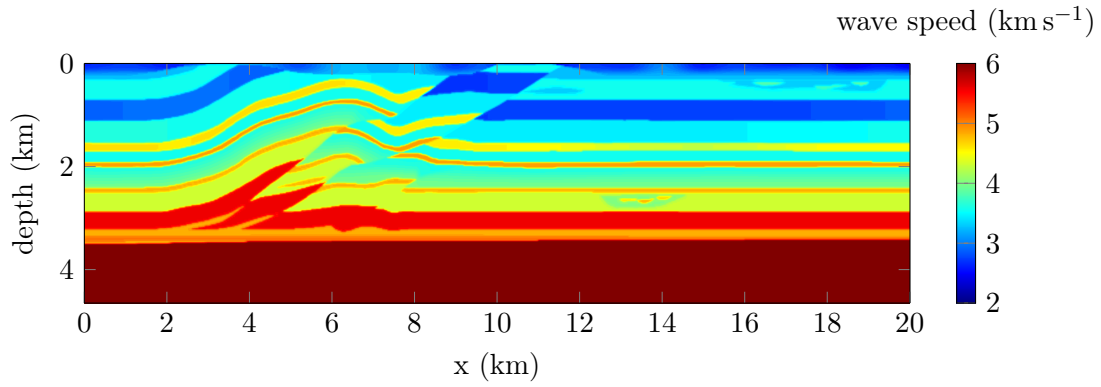


Figure 4.11: Wave speed for the acoustic Overthrust medium of size 20×4.65 km, the velocity varies between 2000 and 6000 $\mathrm{m\,s^{-1}}$.

This model mainly consists of horizontal layers of constant speed with some faults in the upper area. It is larger than the Marmousi experiment. We generate the seismic acquisition with sources and receivers located in the near surface area. We positioned 199 sources and 399 receivers per source. The sources are equally spaced by 100 m and the receivers by 50 m. The data are generated using a finite differences discretization of the wave equation and the iterative minimization is conducted using Discontinuous Galerkin discretization. Furthermore, the source value used to generate the data is assumed not to be precisely known and we recover it following the methodology of Subsection 4.6.4. The set of frequency varies from 1 to 10 Hz, where we use only sequential integer frequencies.

In Figure 4.12 is shown the initial model taken to start the iterative reconstruction. We do not assume any knowledge of the structures and the velocity is lower compared to

the original scale. As in the previous experiment, it is simply a one-dimensional variation of velocity with depth.

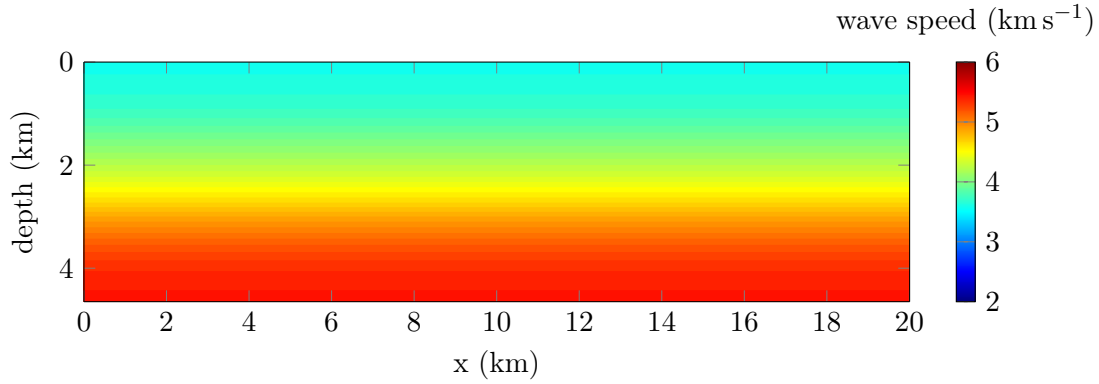
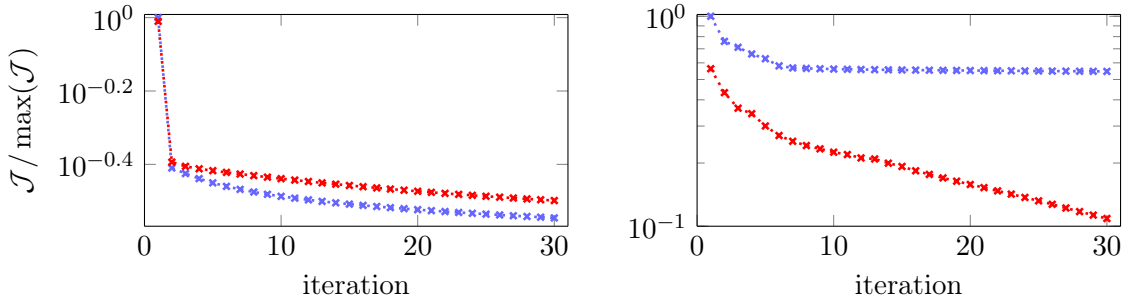


Figure 4.12: Initial wave speed for the iterative minimization algorithm and reconstruction of the Overthrust model. The model has no information on any of the true structures and consists of a one-dimensional variation of velocity with depth.

The FWI algorithm proceeds to 30 iterations per frequency. Following the Marmousi experiment, we have selected two methods to conduct the inversion: the NLCG using the Polak–Ribière formula and the Gauss-Newton Hessian CG with a diagonal preconditioner, where the conjugate gradient algorithm to approximate the Hessian inverse uses 10 iterations. (We have actually run more complete tests, as in Subsection 4.7.1, but because the results are relatively similar we focus on the main methods for clarity). The residuals evolution (cost function) using these two methods are presented in Figure 4.13, for different frequencies.



(a) 2 Hz iterations.

(b) 10 Hz iterations.

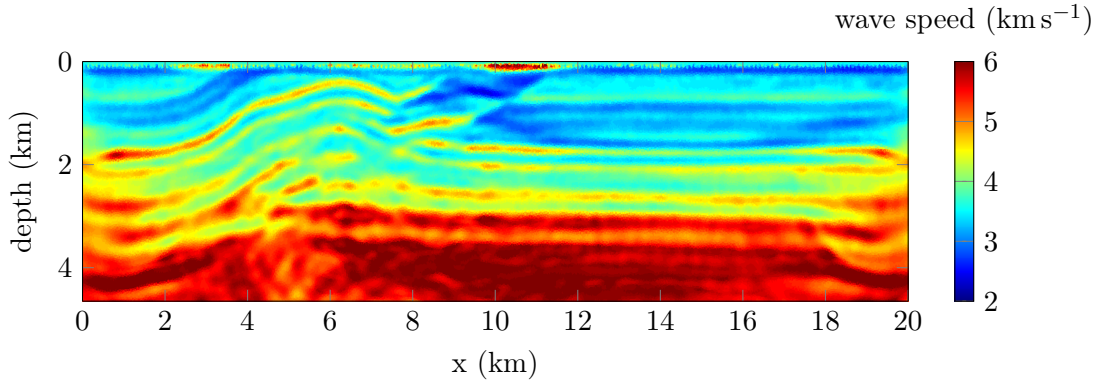
Figure 4.13: Evolution of the cost function for different frequencies. The blue crosses (···×···) involve the Polak–Ribière (PR) formula, the red crosses (···×···) use the Gauss-Newton Hessian with 10 conjugate gradient iterations and diagonal preconditioner with value $10^{-4}\beta$ (see Equation (4.42)).

The initial frequency shows few differences between the two methods. However, for the latest iterations, the benefit of using the Gauss-Newton Hessian approximation is very clear. In particular the nonlinear conjugate gradient stagnates very early, in less than 10 iterations in Figure 4.13(b), but the Gauss-Newton Hessian CG keeps improving the cost

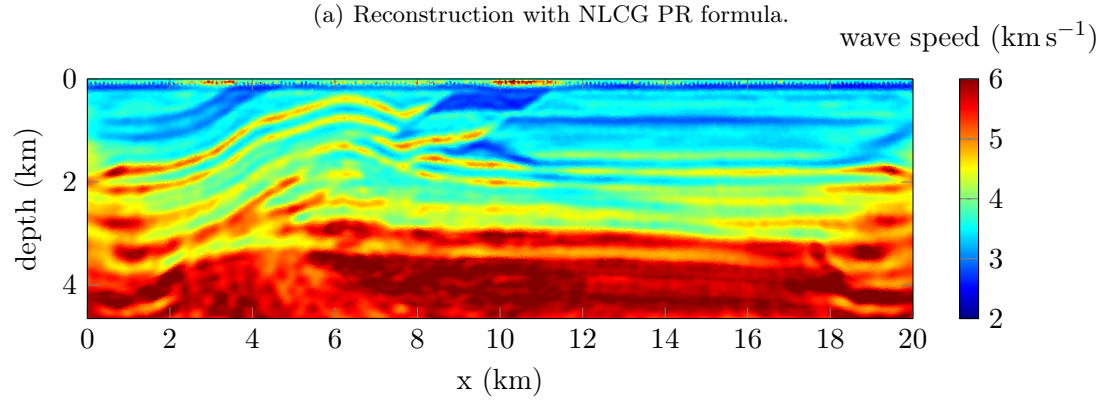
function, providing a major advantage. It confirms the observation of the Marmousi test case reconstruction of Figure 4.9.

Remark 4.12. *In Figure 4.13(a) there is a drastic improvement after the first iteration. This is due to the reconstruction of the source. Indeed, in this experiment we have assumed the source not to be precisely known. Because the source reconstruction is a linear problem (see Subsection 4.6.4), the appropriate value is easily recovered. It has great impact on the residual, as it can be observed.*

The velocities reconstructed for the two methods are presented in Figure 4.14. We perform 30 iterations per frequency from 1 to 10 Hz with 1 Hz step, leading to a total of 300 iterations.



(a) Reconstruction with NLCG PR formula.



(b) Reconstruction with Gauss-Newton Hessian CG and preconditioner.

Figure 4.14: Final reconstruction of the acoustic OT wave speed assuming constant density and using frequencies from 1 to 10 Hz. 30 iterations are performed per frequency (for a total of 300 iterations) and we compare the methods involving the gradient and the Hessian of the cost function.

The benefit of using the Gauss-Newton Hessian CG method appears in the reconstruction, as revealed by the residuals evolution. The different layers are slightly less accurate when using the NLCG method, and the upper area is better designed when employing the Hessian. Yet, the deeper part is hardly recovered, with both methods. Especially, the

sides suffer from limited illumination. Nevertheless, in this experiment as well, the reconstructions are relatively close and the extra cost involved by using the Hessian may not be justified, or should be restricted to a final set of iterations for each frequency. Because we do not exactly know the source value, the very near surface area remains perturbed but it does not impact the rest of the medium.

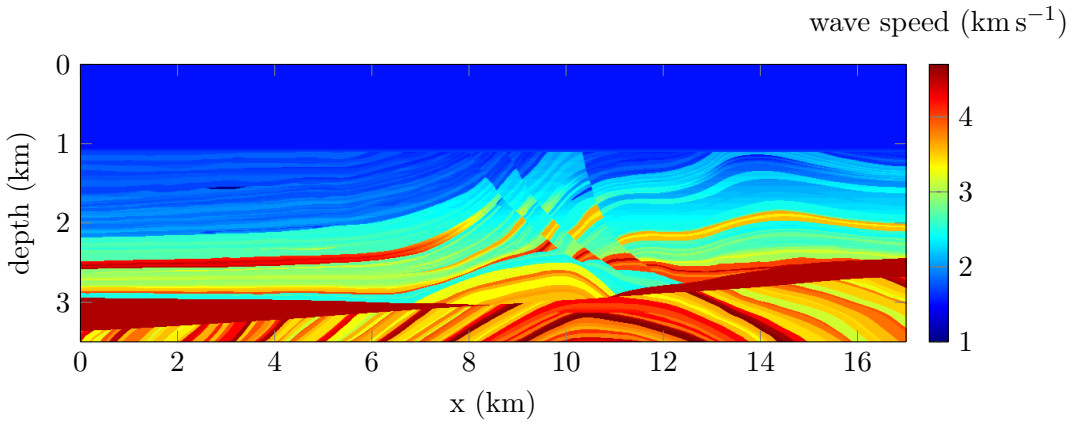
Remark 4.13. *The OT acoustic reconstruction is continued in Chapter 6, Section 6.3, where we illustrate in particular how the velocity recovery is improved by the use of complex frequencies.*

4.7.3 The acoustic Marmousi 2D model with variable density

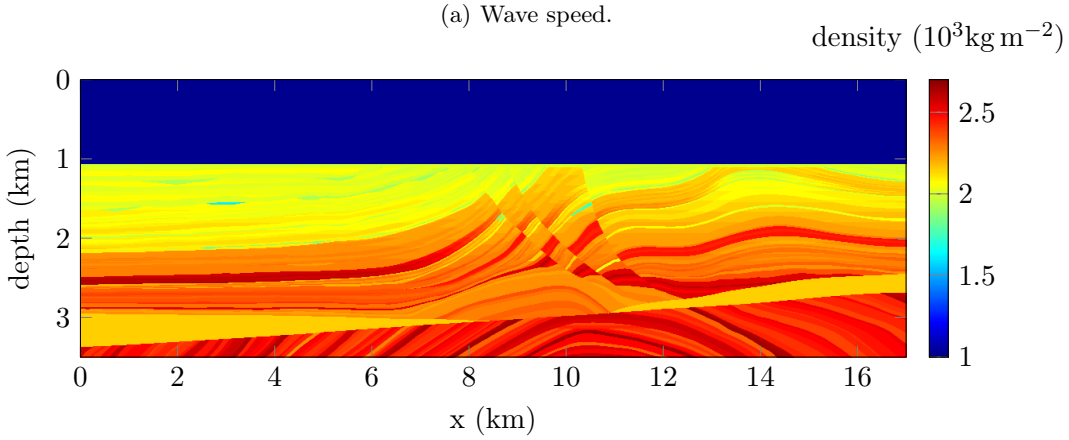
In the previous two experiments, we have considered a single parameter reconstruction where the velocity is the only unknown, meanwhile the density is homogeneous. We now define a multi-parameters inverse problem where, in addition to the velocity, the density depends on the space, and is not known. For this purpose we consider the elastic version of the Marmousi model, which was introduced by Martin et al. (2006), and is also referred to as Marmousi2. We extract the P-wave speed and density only to generate our acoustic test case with variable density. The new model is of size 17×3.5 km. The wave speed varies from 1000 to 4700 $\mathrm{m\,s^{-1}}$, the density varies from 1000 to 2700 $\mathrm{kg\,m^{-2}}$. The models are presented in Figure 4.15. The structures are analog to the Marmousi ones.

In order to perform subsurface reconstruction, the seismic acquisition is conducted from 141 sources (the greater amount compared to the initial Marmousi test is motivated by the increase of dimension). As usual, the sources emit a wave one by one and receivers are positioned to record the resulting pressure field. There are 281 receivers for each source. The data are generated in the frequency domain using a discontinuous Galerkin discretization, as for the initial experiment, the order and mesh differ compared to the inversion algorithm, to avoid inverse crime. The initial models for the reconstruction do not assume any knowledge of the structures: neither for the wave speed, nor for the density. The starting models are presented in Figure 4.16.

For the reconstruction of acoustic media with variable density, it is actually more convenient to invert the bulk modulus and the density instead of the velocity and the density. This is motivated by the wave equation, see Equation (1.28), which decouples the bulk modulus and the density whereas the velocity would depend on the density. A discussion on the parametrization and comparison of experiments will later be processed in Chapter 5, Sections 5.4, 5.5 and 5.6. Let us remind the definition of the bulk modulus

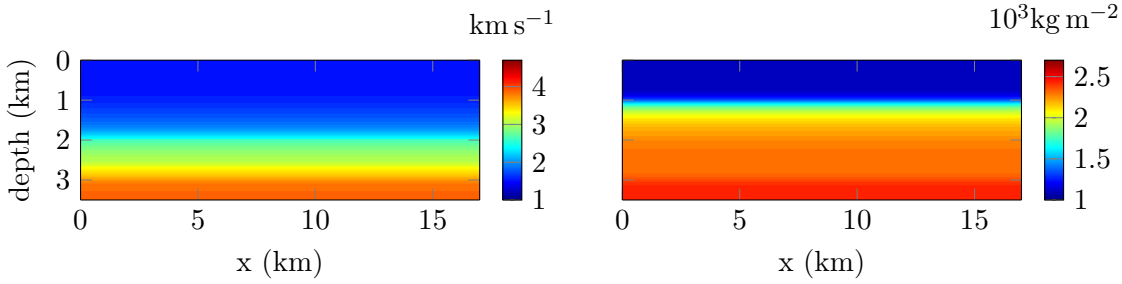


(a) Wave speed.



(b) Density.

Figure 4.15: Wave speed and density for the acoustic Marmousi medium of size 17×3.5 km with variable density.



(a) Wave speed.

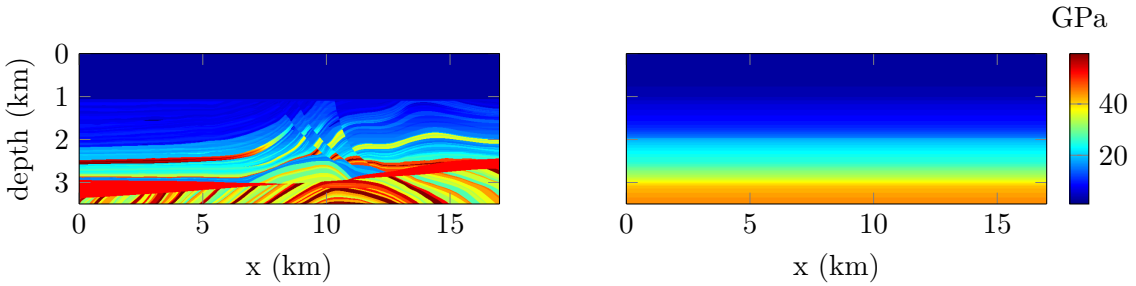
(b) Density.

Figure 4.16: Initial wave speed and density for the reconstruction of the acoustic Marmousi medium with variable density.

from the velocity and density:

$$c = \sqrt{\frac{\kappa}{\rho}}\,, \qquad\qquad c^2\rho = \kappa.$$

In Figure 4.17, we show the bulk modulus extracted from the true and starting Marmousi media. It obviously encompasses the same structures as the velocity and density but with the appropriate unit.



(a) Marmousi bulk modulus.

(b) Starting bulk modulus.

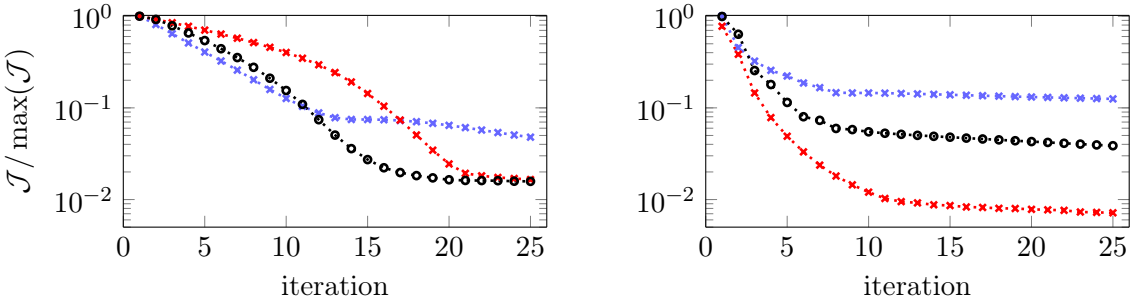
Figure 4.17: True and initial bulk modulus related to the acoustic Marmousi medium with variable density, the corresponding velocity and density have been shown in Figures 4.15 and 4.16.

The reconstruction of the density is known to be more complicated (Virieux and Operto, 2009) than the bulk modulus (once again we refer to the complete set of experiments of Chapter 5, Sections 5.4, 5.5 and 5.6). That is why we do not perform the iterative minimization for the density parameter but only for the bulk modulus (hence reducing the computational time). Nonetheless, we do not assume the knowledge of the density, we simply keep the initial representation (see Figure 4.16(b)), and we do not modify it through the process.

We present three methods that have been used to conduct the iterative scheme: the nonlinear conjugate gradient with Fletcher–Reeves formula and the Gauss-Newton Hessian CG with and without the diagonal preconditioner. We perform 25 iterations per frequency. In Figure 4.18, we show the evolution of the cost function with iterations for the first and last frequency.

From the residuals evolution, we observe a similar behavior as for the single parameter reconstruction: the difference between the methods is more important for the higher frequency iterations. Moreover, the use of Hessian information allows a better convergence and the NLCG stagnates rapidly. In this experiment, the use of a preconditioner seems not to improve the convergence but this should be complimented by a thorough study concerning the variety of possibilities for the damping factor. In Figure 4.19 are presented the final reconstructions of the bulk modulus using NLCG and Gauss-Newton CG with no preconditioner.

The benefit of using Hessian, which appears in the plot of the cost function evolution,



(a) 1 Hz iterations.

(b) 10 Hz iterations.

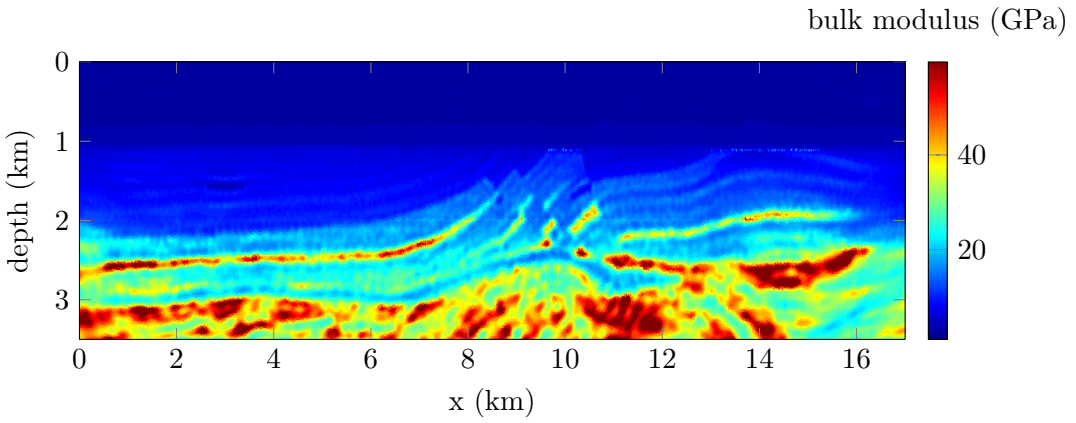
Figure 4.18: Evolution of the cost function for different frequencies using different methods for the model update. The blue crosses (···×···) use the Fletcher–Reeves (FR) nonlinear conjugate gradient. The red crosses (···×···) use the Gauss-Newton Hessian CG approximation. The black circles (···o···) use the Gauss-Newton Hessian CG approximation with preconditioner. The number of iterations for the conjugate gradient is set to 10.

is very clear in the final reconstruction. The upper part of the model is well recovered by both methods, however, the deepest structures are only recovered with the Hessian. In particular, the circular layers in the central deepest part of the model appear very well with the Hessian, but are not shown when using the NLCG method. Even the lateral borders seem slightly more accurate with the Hessian, despite the limited illumination. We detail this gain with a one-dimensional section of the bulk modulus at fixed position $x = 10$ km in Figure 4.20. Both methods recover well the upper part of the medium but it is clear that only the Gauss-Newton Hessian CG approximates accurately the deepest variations, and the NLCG remains at very low values in comparison. It is also important to notice that despite the absence of knowledge in the density, we are still able to recover the bulk modulus precisely.

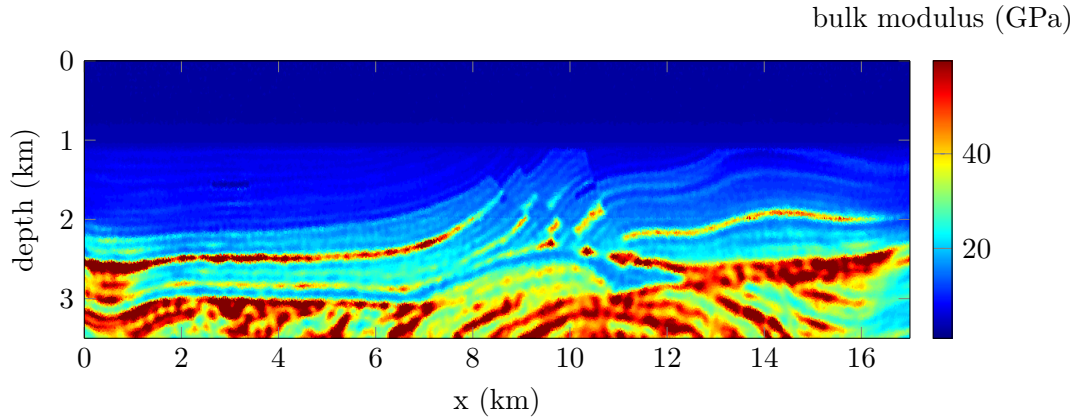
4.7.4 Concluding remark

The benefits of using the Hessian were relatively limited in our initial experiments involving a single parameter reconstruction. However, when two models are unknown, the gain of using the Hessian is much more relevant. There is a need to pursue the comparison and analysis of the methods, in particular to clarify the following aspects:

1. choice of preconditioner for the Hessian,
2. use of full or Gauss-Newton Hessian,
3. mixed algorithm with initial iterations using the gradient and a few set of final iterations involving the Hessian to optimize the computational time,
4. the effect of the noise in the data.



(a) Reconstruction using NLCG FR.



(b) Reconstruction using GNH CG.

Figure 4.19: Final reconstruction of the acoustic Marmousi bulk modulus assuming variable density and using frequencies from 1 to 10 Hz. 25 iterations are performed per frequency (for a total of 250 iterations) and we compare the methods involving the gradient and the Gauss-Newton Hessian of the cost function.

More generally a wide variety of regularization techniques exists in the literature and a thorough study for the seismic application could be engaged, in particular to quantify the gain in accuracy in relation with the numerical cost, which remains a limitation in seismic exploration.

4.8 Conclusion

In this chapter, we have presented the seismic imaging for subsurface media from an optimization point of view. This reconstruction method, relying on an iterative minimization, is the essence of FWI. We have reviewed the specificity of the geophysical situation with the types of data available depending on land or marine area of interest. Newton type

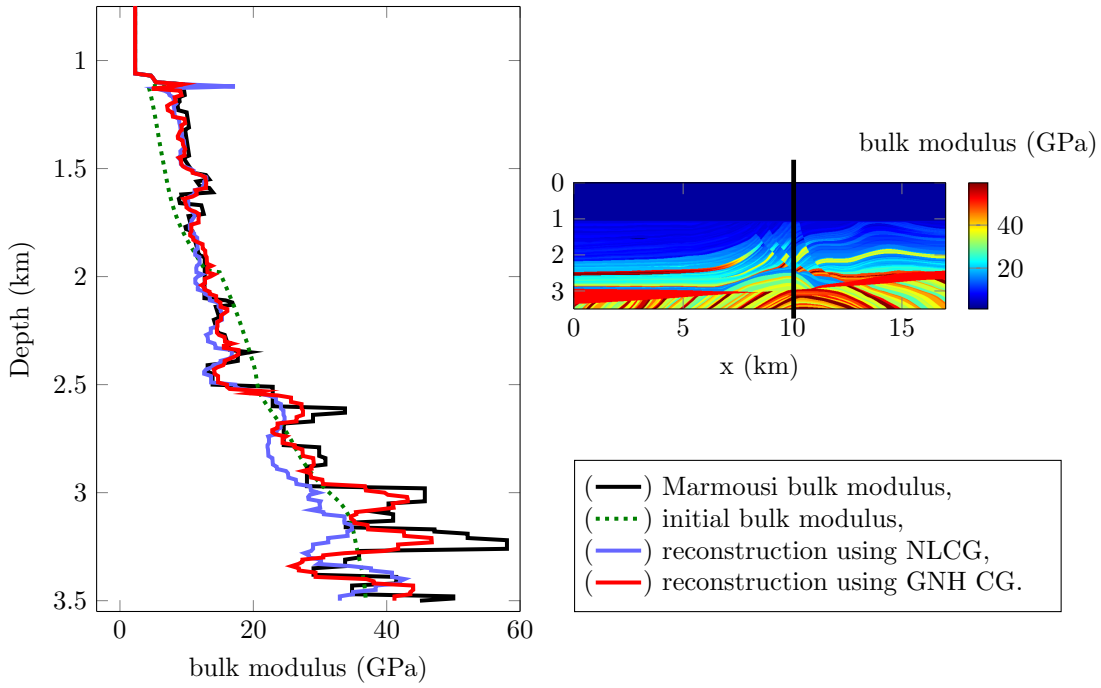


Figure 4.20: One-dimensional profile of the acoustic Marmousi bulk modulus for a fixed position $x = 10$ km. The position is indicated with the black line on the original Marmousi bulk modulus (right). The reconstructed profiles are extracted from the results presented Figure 4.19.

methods provide the natural framework to solve the minimization problem, and we have given the main features that have been implemented in this project. In particular, the cost of computing the Hessian may be numerically overwhelming, and we have presented some techniques based on gradient descent. Finally, we have compared the efficiency of the FWI with and without computing the Hessian. We have illustrated that in our preliminary experiments, the benefits of the Hessian appear in the multi-parameters inversion situation. The proper analysis of the methods should be pursued, in particular with the preconditioner, also because of the many optimization techniques existing.

Chapter 5

Quantitative convergence of FWI with geometry, frequency and parametrization

Abstract

We study the convergence of the iterative minimization algorithm deployed for the reconstruction of subsurface coefficients. The associated inverse problem is nonlinear and ill-posed, providing complications in the process. Here we propose a quantification of the convergence properties based on least squares optimization theory, which matches the Full Waveform Inversion (FWI) method. We define numerical estimates to indicate the nonlinearity of the problem depending on several aspects: the model space size, the selection of (complex) frequencies and the geometry of the subsurface structures. It allows us to precisely characterize the scheme and define the proper procedure to follow. It also indicates how the reconstruction is affected by the type of experiments. Our estimates are general and could as well be employed in other inverse problems solved via optimization methods. Eventually, we study the effect of the choice of parameters (the parametrization) and how it acts on the Fréchet derivative. The impact of parametrization is further illustrated with synthetic numerical reconstructions for acoustic and elastic media.

Disclaimer: *This is an ongoing work in part of collaboration with Profs. Hélène Barucq, Henri Calandra and Guy Chavent.*

Contents

The convergence of the minimization algorithm, coupled with the stability, are the keys to obtaining the appropriate recovery of subsurface parameters based on iterative minimization algorithms. We have already studied the stability associated with the inverse problem in Chapter 3, which legitimizes the minimization of surface information in order to obtain the medium coefficients. The stability directly relates the minimization of the cost function with the accuracy of the parameters reconstruction, and we have shown (Chapter 3) that under some advantageous situations, we have a Lipschitz-type stability and ways to quantify the underlying stability constant. For the convergence, we are more interested in the optimization algorithm itself (namely iterative minimization), to guarantee that our iterations 'head' to the appropriate solution. The seismic inverse problem is nonlinear and ill-posed, and the amount of information at our disposal is relatively limited (surface measurements only). Usually, as the available data further suffer from reduced accuracy, the uniqueness is not guaranteed, e.g., Landa and Treitel (2016), and local minima can lead to the failure of the minimization process. Hence, it is fundamental to understand the convergence of the scheme, and possibly extract the quantity of importance for its success.

In this chapter, we consider least squares minimization algorithm and follow the work of Chavent and Kunisch (1996); Chavent (2010). In particular, we use the notation and framework introduced in Chavent (2010). Then we extend it with numerical estimates applied to the geophysical situation to investigate the quantities reflecting the nonlinearity of the problem. In our approach, we first explore two aspects: the basin of attraction indicates how large (or restricted) is the research (model) space for the targeted parameters. It is particularly important as we aim to use minimal prior information in the experimental test cases. Then, the maximal distance allowed between the observed measurements and the attainable set is accessed with the radius of curvature, and validates the optimization scheme. In some sense it may be seen as the robustness to noise and model error of the problem.

Those two aspects are numerically quantified, and their behavior is inspected depending on several aspects, such as the frequency, the geometry of the subsurface models, etc; see Sections 5.2 and 5.3. It allows us to suitably define the (complex) frequency progression to follow during the procedure of iterative minimization.

We review how the choice of parametrization influences the conduction of FWI. By parametrization, we refer here to the selection of a parameter to target. For example, elastic media reconstruction can focus on the Lamé parameters, the velocities (P- and S-wave speeds) or the impedances. This choice may appear harmless but is actually crucial. As early applications suggested, Tarantola (1986), the combination must use as uncorrelated parameters as possible. Besides the choice of parameters, the application of functions (inverse, logarithm) onto the parameter also affects the reconstruction. In Section 5.4 we

give references and depict some of the possible parametrizations, depending on acoustic or elastic media. Here we study their influence onto the condition number and singular values of the associated Fréchet derivatives. FWI algorithm is finally conducted to compare the impact on the accuracy of the reconstruction, using acoustic and elastic synthetic experiments, see Sections 5.5 and 5.6.

5.1 Convergence of least squares minimization

In this section, we review convergence properties of minimization problems. In particular, we give an estimation of the radius of curvature of the problem. We also introduce the deflection which is related to the model space. The notation follows the work of Chavent (2010), where the reader is referred to for more details. Let us consider the $L2$ minimization of the cost function (as we introduced for the seismic reconstruction problem in Equation (4.8)),

$$\min_{m\in\mathcal{M}} \mathcal{J}(m) = \frac{1}{2}\|\mathcal{F}(m) - d\|_F^2. \tag{5.1}$$

Following the already used notation, $\mathcal{F}$ represents the forward problem and d stands for the observed data. We further introduce the requirements of Chavent (2010):

- E is a Banach space with norm $\| \, . \, \|_E$,
- $\mathcal{M}$ is the parameter set, $\mathcal{M} \subset E$ and $\mathcal{M}$ is convex and closed,
- $d \in F$, where F is a Hilbert space with norm $\| \, . \, \|_F$,
- the forward operator $\mathcal{F} : \mathcal{M} \to F$ is differentiable along segments of $\mathcal{M}$ and $\mathcal{F}(\mathcal{M})$ represents the *attainable set*,
- $\exists \ \mathcal{C} \geq 0$ such that $\forall m_1, m_2 \in \mathcal{M}, \ \forall t \in [0,1], \quad \|D_t\mathcal{F}\big((1-t)m_1 + tm_2\big)\|_F \leq \mathcal{C}\|m_2 - m_1\|_E$, where D_t stands for the derivative with respect to the parameter t.

The parametric curve P is defined such that

$$\forall m_1, m_2 \in \mathcal{M}, \qquad P \ : \ t \in [0,1] \to \mathcal{F}\big((1-t)m_1 + tm_2\big) \quad \in \quad \mathcal{F}(\mathcal{M}).$$

Assuming P is twice differentiable, the velocity V and acceleration A are given by the derivatives of the curve,

$$\begin{cases} V(t) = P'(t), & A(t) = P''(t), \\ v(t) = \dfrac{V(t)}{\|V(t)\|}, & a(t) = \dfrac{A(t) - \big\langle A(t), v(t)\big\rangle v(t)}{\|V(t)\|^2}. \end{cases}$$

5.1.1 Finite curvature/limited deflection problem

We define the finite curvature/limited deflection (FC/LD) framework for the least squares minimization problem. Then we explicit the underlying advantages of such problems with Proposition 5.1. For this review, we follow the work of (Chavent, 2010, pp. 167–171).

Definition 5.1 (Finite Curvature (FC)). *The minimization Problem (5.1) is a finite curvature least squares problem if the following condition is satisfied:*
there exists $R > 0$ such that $\forall m_1, m_2 \in \mathcal{M}$, the curve P satisfies $P \in W^{2,\infty}([0,1], F)$ and $\|A(t)\|_F \leq \frac{1}{R}\|V(t)\|_F^2$ for a.e. $t \in [0,1]$.

Definition 5.2 (Radius of curvature). *Assuming a finite curvature least squares problem for which $V(t) \neq 0$ $\forall t \in [0,1]$, the radius of curvature $R_\rho(t)$ along the curve P satisfies*

$$\frac{1}{R_\rho(t)} = \|a(t)\| \leq \frac{\|A(t)\|_F}{\|V(t)\|_F^2} \qquad \text{for a.e. } t \in [0,1].$$

Hence, the radii of curvature of the curve P and of the attainable set $\mathcal{F}(\mathcal{M})$, verify

$$\frac{1}{R_\rho(P)} = \sup_{t\in[0,1]} \frac{1}{R_\rho(t)} \leq \frac{1}{R_\rho(\mathcal{F}(\mathcal{M}))} = \sup_{m_1,m_2\in\mathcal{M};t\in[0,1]} \frac{1}{R_\rho(t)} \leq \frac{1}{R} \leq +\infty.$$

Definition 5.3 (Deflection). *The deflection Θ of the curve P is defined as the largest angle $\theta(t_a, t_b) \in [0, \pi]$ between any two tangent vectors $V(t_a)$ and $V(t_b)$ for any two points t_a and t_b of P. An infinitesimal variation of the deflection $d\theta$ satisfies*

$$d\theta \leq \frac{\|A(t)\|_F}{\|V(t)\|_F} dt.$$

Denoting t_1 and t_2 the values of t where the deflection is maximum, the maximum deflection along the curve P, $\Theta(P)$, is given by

$$\Theta(P) = \int_{t_1}^{t_2} d\theta \leq \int_0^1 \frac{\|A(t)\|_F}{\|V(t)\|_F} \, \mathrm{d}t.$$

Definition 5.4 (Limited Deflection (LD)). *A finite curvature problem is a limited deflection least squares problem if it satisfies the deflection condition*

$$\Theta \leq \frac{\pi}{2}.$$

Proposition 5.1 (FC/LD properties). *Let us consider a finite curvature limited deflection problem with finite curvature $1/R$ and let $\mathcal{V}$ be the enlargement neighborhood of $\mathcal{F}(\mathcal{M})$ defined by*

$$\mathcal{V} = \big\{ z \in F \quad | \quad \mathrm{dist}\,\big(d, \mathcal{F}(\mathcal{M})\big) < R \big\}.$$

Then the projection on the attainable set $\mathcal{F}(\mathcal{M})$ has the following properties:

- *Uniqueness: $\forall d \in \mathcal{V}$, there exists at most one projection of d on $\mathcal{F}(\mathcal{M})$.*

- *Unimodality: if $d \in \mathcal{V}$ admits a projection $\widehat{x}$ on $\mathcal{F}(\mathcal{M})$, the function* $\mathrm{dist}(d, \cdot)$ *has no parasitic stationary point on* $\mathcal{F}(\mathcal{M})$.
- *Existence: if $z \in \mathcal{V}$, any minimizing sequence $X_n \in \mathcal{F}(\mathcal{M})$ of the function $d(z, \cdot)$ is a Cauchy sequence for both distance $\|X - Y\|_F$ and the arc length distance $\delta(X, Y)$. Hence X_n converges in F to the unique projection $\widehat{x}$ of z onto $\mathcal{F}(\mathcal{M})$.*

Local stability is also obtained, see (Chavent, 2010, Proposition 4.2.7), but we do not insist on it as it has been studied in Chapter 3.

We have reviewed the main components of the FC/LD problem, and the proofs can be found in Chavent (2010). Those type of problems were also referred to as *weakly nonlinear* inverse problem in Chavent and Kunisch (1996). If the original minimization Problem 5.1 is cast to have a FC/LD property, it shows rightful dispositions so that one can legitimately aim at obtaining the appropriate minimizer. Finite curvature and limited deflection are the two aspects of importance to give such guarantees on the convergence of the minimization algorithm. First, the deflection must be lower than $\pi/2$ (Definition 5.4). Then we must assure that the distance between the measurements and the attainable set is lower than the curvature R (Proposition 5.1). We illustrate some quantities with a one-dimensional forward problem setup in Figure 5.1. The deflection gives information on the attainable structure and the radius of curvature ensures the data are 'not too far' from the attainable set.

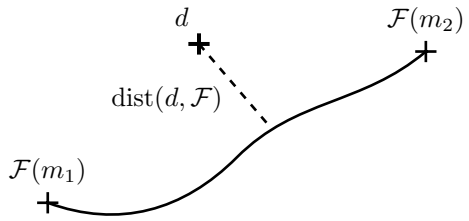


Figure 5.1: Illustration of forward operator in one dimension. The FC condition imposes m_1 and m_2 to be such that the deflection along the path $[\mathcal{F}(m_1), \mathcal{F}(m_2)]$ is less that $\pi/2$. the LD condition imposes the distance between the data d and their projection to attainable set, $\mathrm{dist}(d, \mathcal{F})$, to be bounded by the radius of curvature. The data generally do not belong on the attainable set because of model error, accuracy of acquisition devices and noise in the measurements.

5.1.2 Quantitative estimates

In general, the seismic inverse problem, without special assumptions, is impossible to be cast as an FC/LD problem. In particular, when no prior information is known on the subsurface, the model space can be a 'large' subset of $\mathbb{R}$. Our idea here is to control how

much the model space has to be reduced to eventually guarantee the deflection condition around a center point. We also give an estimation of the radius of curvature.

5.1.2.1 Model space size through estimation of the deflection

In the case of limited deflection problems, the deflection on the model space is less than or equal to $\pi/2$. Here, we study the restriction of the model space, $\mathcal{M}_\theta^{m_0} \subset \mathcal{M}$, for which we assume that the property is validated. Namely, the idea is to reduce the size of the model space so that the condition is eventually obtained. In the following, we only focus on all paths that pass by the central (initial) parameter m_0 to extract the estimates.

Definition 5.5. *For any unit vector δ_{m_k}, we define $\mathcal{M}_\theta^{m_0,\delta_{m_k}}$ as the interval $[m_0 - \Delta_{m_0}^{\delta_{m_k}}\delta_{m_k}, m_0 + \Delta_{m_0}^{\delta_{m_k}}\delta_{m_k}]$ where $\Delta_{m_0}^{\delta_{m_k}} > 0$ is chosen such that the deflection along the interval is smaller than $\pi/2$.*

We can extract a lower bound for the size of the subspace,

Definition 5.6. *$\mathcal{M}_\theta^{m_0}$ is the ball made of all intervals of the form $[m_0 - \Delta_{m_0}\delta_{m_k}, m_0 + \Delta_{m_0}\delta_{m_k}]$ where δ_{m_k} is a unit vector and $\Delta_{m_0} > 0$ is chosen such that the deflection along all intervals passing by m_0 is smaller than $\pi/2$.*

Hence we require the original Problem (5.1) restricted to $\mathcal{M}_\theta^{m_0}$ to have limited deflection. More precisely we are interested in the estimation of Δ_{m_0} which can be seen as the size of the basin around m_0 which satisfies radially the limited deflection property. The path P for an interval of size $2\Delta_{m_0}$ in the direction δ_{m_k} (of norm 1) is

$$P(t) = \mathcal{F}\big(m_0 + (2t-1)\Delta_{m_0}^{\delta_{m_k}}\big),$$

where a simple change of variable has been carried out, as illustrated in Figure 5.2.

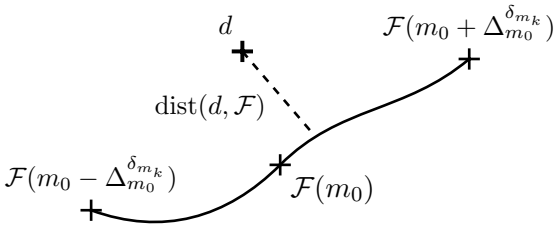


Figure 5.2: Change of variable for the forward operator in one dimension. Our estimates are designed to give the value of Δ_{m_0} such that the limited deflection condition is respected in the direction δ_{m_k}.

One can then straightforwardly use $m(t) = (m_0 + (2t-1)\Delta_{m_0}^{\delta_{m_k}})$ and the chain rule to

define the velocity and acceleration in terms of the derivative with respect to m. Following the notation of Cartan (1971), we write

$$\begin{cases} V(t) = 2\Delta_{m_0}^{\delta_{m_k}} D\mathcal{F}(m_0)\delta_{m_k}, \\ A(t) = 4(\Delta_{m_0}^{\delta_{m_k}})^2 D^2\mathcal{F}(m_0)(\delta_{m_k}, \delta_{m_k}), \end{cases}$$

where δ_{m_k} acts as the direction of the derivation. Finally, the deflection for the interval $[m_0 - \Delta_{m_0}^{\delta_{m_k}}, m_0 + \Delta_{m_0}^{\delta_{m_k}}]$ in the direction δ_{m_k} is approximated with a rectangle approximation of the integral (which is consistent when the interval is sufficiently small) and we write

$$\Theta \leq \int_0^1 \frac{\|A(t)\|}{\|V(t)\|} dt \sim \frac{\|A(1/2)\|}{\|V(1/2)\|} = 2\Delta_{m_0}^{\delta_{m_k}} \frac{\|D^2\mathcal{F}(m_0)(\delta_{m_k}, \delta_{m_k})\|}{\|D\mathcal{F}(m_0)(\delta_{m_k})\|}.$$

We can further identify an upper bound for the formulation,

$$\frac{\|D^2\mathcal{F}(m_0)(\delta_{m_k}, \delta_{m_k})\|}{\|D\mathcal{F}(m_0)(\delta_{m_k})\|} \leq \frac{\|D^2\mathcal{F}(m_0)\|\|\delta_{m_k}\|^2}{\lambda_{\min}^{D\mathcal{F}}\|\delta_{m_k}\|} = \frac{\|D^2\mathcal{F}(m_0)\|}{\lambda_{\min}^{D\mathcal{F}}},$$

where $\lambda_{\min}^{D\mathcal{F}}$ denotes the lowest singular value of $D\mathcal{F}(m_0)$.

We impose the quantity Θ to be lower than or equal to $\pi/2$ in order to obtain a condition on Δ_{m_0} that ensures that the deflection property is satisfied in the direction δ_{m_k}

$$\Delta_{m_0}^{\delta_{m_k}} = \frac{\pi}{4} \frac{\|D\mathcal{F}(m_0)(\delta_{m_k})\|}{\|D^2\mathcal{F}(m_0)(\delta_{m_k}, \delta_{m_k})\|}. \tag{5.2}$$

Similarly we define a lower bound valid for all directions δ_{m_k} passing by m_0,

$$\Delta_{m_0} = \frac{\pi}{4} \frac{\|\lambda_{\min}^{D\mathcal{F}}\|}{\|D^2\mathcal{F}(m_0)\|}. \tag{5.3}$$

The size of the model space indicates how large is the basin of attraction when minimizing the functional starting from m_0. When the search of the solution is restricted to the ball $\mathcal{M}_\theta^{m_0}$, one can expect that the LD property still holds approximately.

5.1.2.2 Distance to the attainable set, radius of curvature

The neighborhood $\mathcal{V}$ introduced in Proposition 5.1 is necessary because the observed data, d, do not belong in the attainable set in general. This is explained by the limited accuracy of the recording devices, model error and noise. Following the FC condition, the size of the radius of curvature gives an indication of the possibility of convergence. Because the data are not in the attainable set, they must be 'not too far' so that they can still be processed. This maximum distance is materialized by the radius of curvature.

For the numerical estimates, we follow the Definition 5.2. We consider the same interval as for the deflection: $[m_0 - \Delta_{m_0}^{\delta_{m_k}}\delta_{m_k}, m_0 + \Delta_{m_0}^{\delta_{m_k}}\delta_{m_k}]$ so that the estimation of the maximal distance in the direction δ_{m_k} (with $\|\delta_{m_k}\| = 1$) is obtained analogously to what has been done for the deflection. Our estimate is denoted $\widehat{R}_{m_0}^{\delta_{m_k}}$ and is extracted from Definition 5.2 with $a(t)$. It is defined by

$$\widehat{R}_{m_0}^{\delta_{m_k}} = \frac{\|D\mathcal{F}(m_0)(\delta_{m_k})\|^2}{\|D^2\mathcal{F}(m_0)(\delta_{m_k},\delta_{m_k}) - \xi(m_0,\delta_{m_k})\|}, \tag{5.4}$$

where

$$\xi(m_0,\delta_{m_k}) = \big\langle D^2\mathcal{F}(m_0)(\delta_{m_k},\delta_{m_k}), \frac{D\mathcal{F}(m_0)(\delta_{m_k})}{\|D\mathcal{F}(m_0)(\delta_{m_k})\|}\big\rangle \frac{D\mathcal{F}(m_0)(\delta_{m_k})}{\|D\mathcal{F}(m_0)(\delta_{m_k})\|}.$$

A lower bound for the radius of curvature in all directions passing by m_0 follows

$$R \geq \frac{(\lambda_{\min}^{D\mathcal{F}})^2}{\|D^2\mathcal{F}(m_0) - \xi(m_0)\|} = \widehat{R}_{m_0}, \tag{5.5}$$

with $\lambda_{\min}^{D\mathcal{F}}$ the lowest singular value of the Fréchet derivative and

$$\xi(m_0) = \big\langle D^2\mathcal{F}(m_0), \frac{D\mathcal{F}(m_0)}{\|D\mathcal{F}(m_0)\|}\big\rangle \frac{D\mathcal{F}(m_0)}{\|D\mathcal{F}(m_0)\|}.$$

5.2 Model space size estimates via limited deflection

In this section, we select a model m_0 and estimate the quantity $\Delta_{m_0}^{\delta_{m_k}}$ for different directions. The direction, δ_{m_k}, is normalized and can be seen as the geometry of the parameter to be reconstructed. The larger the estimate is, the larger is the interval size and the better is the chance to converge. Here we consider a two-dimensional geophysical setup with data (observations from sources and receivers) located only near the surface. The upper boundary (Γ_1) consists in free surface boundary condition meanwhile the lateral and bottom ones (Γ_2) use artificial boundaries (PML or absorbing boundary conditions) to avoid wave reflections. We sketch the computational domain in Figure 5.3.

5.2.1 Geometry and sequential frequency

Let us first consider an acoustic medium with constant density. The model m_0 is taken to represent a smooth background velocity, analogous to the starting model for the iterative reconstruction. It is pictured in Figure 5.4 and is of size 9.2×3 km (it is the starting medium used for the reconstruction of the acoustic Marmousi medium using FWI).

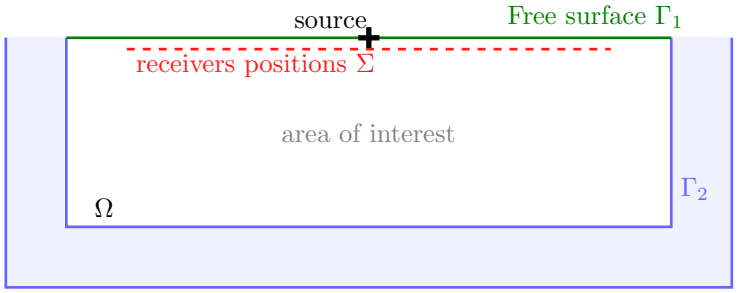


Figure 5.3: Illustration of the computational domain for the quantitative estimates of the quantity $\Delta_{m_0}^{\delta_{m_k}}$ such that it validates the limited deflection property. It characterizes a geophysical setup where the blue area stands for the location of the PMLs, positioned to avoid the reflection of the waves attaining the lateral and bottom boundaries.

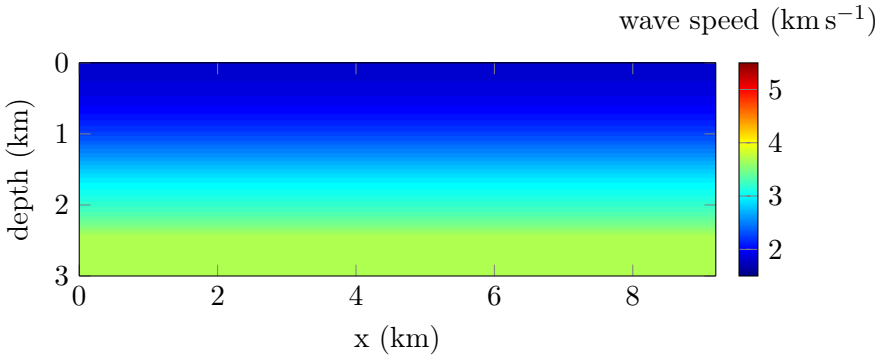
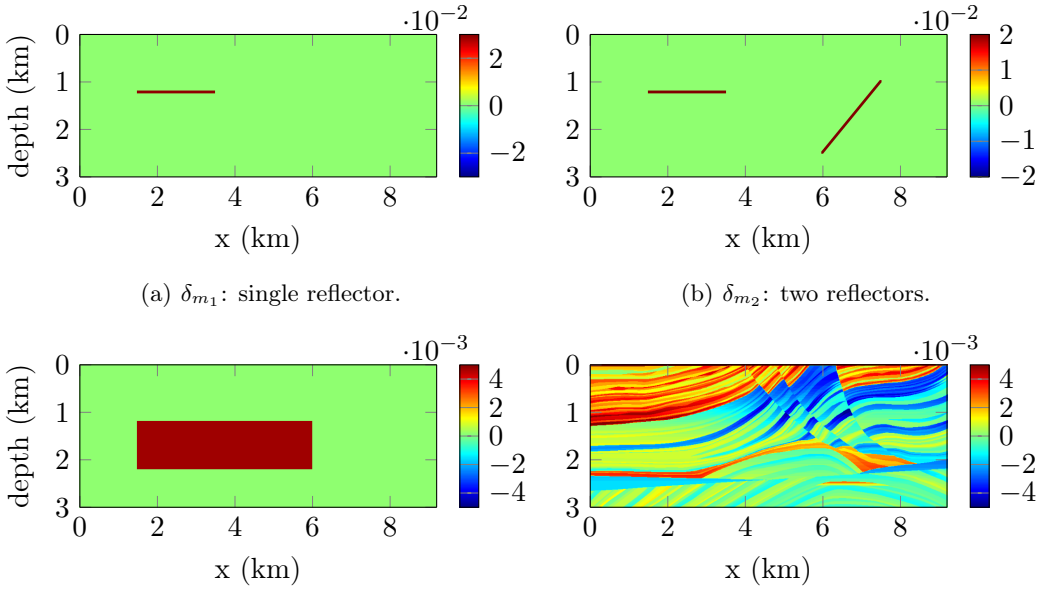


Figure 5.4: Acoustic wave speed c_0 used to deduce $m_0 = c_0^{-2}$. It is of size 9.2×3 km.

In this acoustic situation, the propagation of waves follows the traditional Helmholtz equation (1.27). We consider the parameter of interest to be $m_0 = c_0^{-2}$. It is expressed in $\mathrm{s}^2\,\mathrm{m}^{-2}$, and similarly for the selected directions δ_{m_k}. This choice is motivated by the wave equation where the velocity c appears as c^{-2}, see Equation (1.27).

We study the evolution of the $\Delta_{m_0}^{\delta_{m_k}}$ for different directions δ_{m_k}, which are chosen to be based on straight reflectors, or extracted from the Marmousi model. They are shown in Figure 5.5. Our purpose is to analyze the size of the interval depending on the geometry of the search direction, as well as the frequency. We have designed four directions, δ_{m_1}, δ_{m_2}, δ_{m_3} and δ_{m_m}. δ_{m_1} has a single, slim reflector (Figure 5.5(a)), δ_{m_2} has two slim reflectors (Figure 5.5(b)), δ_{m_3} has a single large reflector (Figure 5.5(c)) and δ_{m_m} represents the reflectors extracted from the Marmousi velocity medium (Figure 5.5(d)). The amplitude for the directions are selected so that $\|\delta_{m_k}\| = 1$, $\forall k$.

We follow Equation (5.2) to compute the quantity of interest, which requires the first and second order Fréchet derivatives in the direction δ_{m_k}. In this acoustic situation, the forward problem gives the pressure at the receivers position for each source s and frequency



(a) δ_{m_1}: single reflector.

(b) δ_{m_2}: two reflectors.

(c) δ_{m_3}: single large reflector.

(d) δ_{m_m}: reflectors following Marmousi structure.

Figure 5.5: Four different directions for the estimation of the quantities of interest ($\widehat{R}_{m_0}^{\delta_{m_k}}$ and $\Delta_{m_0}^{\delta_{m_k}}$). The amplitude is determined so that $\|\delta_{m_k}\| = 1$. The values of δ_{m_k} are given in $(\mathrm{m\,s^{-1}})^{-2} = \mathrm{s^2\,m^{-2}}$.

ω such that

$$\mathcal{F}_\omega^{(s)} : m \quad \rightarrow \quad p^{(s)}(\mathbf{x}) \mid_\Sigma \quad = \quad \{p^{(s)}(\mathbf{x}_1), \ldots, p^{(s)}(\mathbf{x}_{n_{rcv}})\},$$

where $p^{(s)}$ stands for the pressure field associated with source s, and $\mathbf{x}_i$ denotes the position of the i^{th} receiver. The forward problem associated with a selected frequency is a vector of $\mathbb{C}^{(n_{rcv} \times n_{src})}$. n_{rcv} is the number of receivers per source and n_{src} the number of sources. In this experiment, we take 19 sources and 183 receivers for each source. Both are located near the surface (see Figure 5.3). The computation of $\Delta_{m_0}^{\delta_{m_k}}$ is processed for sequential frequencies from 1 to 15 Hz, every 0.5 Hz, and for each of the four directions of Figure 5.5. We also incorporate an additional direction, δ_{m_c}, which corresponds to a constant background perturbation.

We first point out that whatever direction is taken, the estimation of $\Delta_{m_0}^{\delta_{m_k}}$ decreases with increasing frequency. This is the expected behavior that indicates that lower frequencies give a better chance for the iterative minimization to converge when no prior information is known for the initial model. Namely, the lower frequencies increase the radius of convergence, which is the natural intuition that has widely been observed in applications, as in Bunks et al. (1995); Pratt et al. (1996); Sirgue and Pratt (2004). The latest also propose how to select consecutive frequencies, based on wavenumber coverage quantities. Comparing the geometry of the directions, at low frequencies, perturbations δ_{m_1} and δ_{m_2}

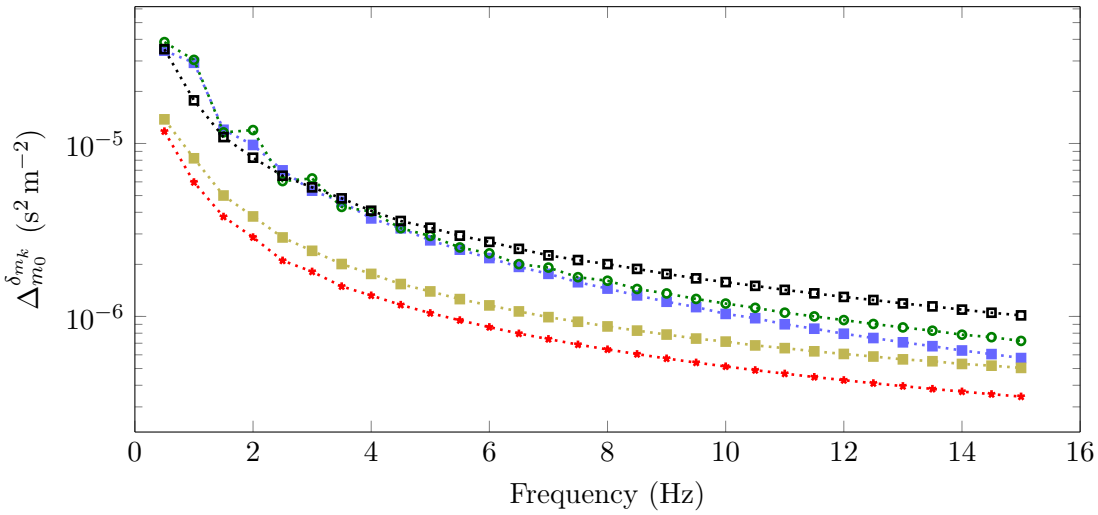


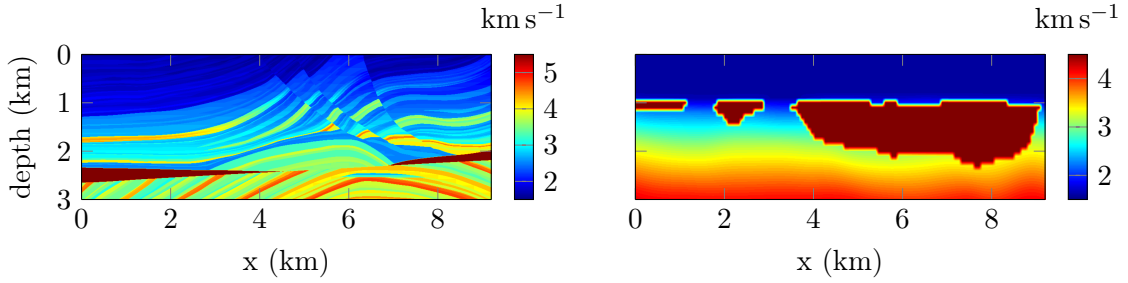
Figure 5.6: Evolution of the interval size verifying the deflection condition, quantity $\Delta_{m_0}^{\delta_{m_k}}$, with frequency using m_0 as a smooth velocity background and different directions δ_{m_k}. The blue squares (··■··) employ one reflector (δ_{m_1}, Figure 5.5(a)), the green circles (··●··) two reflectors (δ_{m_2}, Figure 5.5(b)) the red stars (··★··) one large reflector (δ_{m_3}, Figure 5.5(c)) and the black squares (··■··) the reflectors extracted from the Marmousi model (δ_{m_m}, Figure 5.5(d)). The yellow squares (··■··) represent the estimates using constant direction.

give the larger interval but they are surprisingly rapidly overcome by the many reflectors of the Marmousi medium. Yet, all give very similar size. On the other hand, having one large reflector produces a marked reduction of the size of the interval, which indicates that having a large objects is complicated to recover, even more than the many reflectors of Marmousi-like models. This is actually observed in numerical applications for geophysics where salt domes are more difficult to recover, we illustrate this difficulty with the numerical applications, see Chapter 6. Note that if low frequencies are required to counter the lack of prior information, the increase of frequency is motivated by the stability result which shows that higher frequencies are necessary to improve the resolution, see Chapter 3.

We now investigate the influence of the initial model m_0 onto the estimates. We consider two models: the acoustic Marmousi medium and a model encompassing objects of high contrasts, given in Figure 5.7. The respective estimates for $\Delta_{m_j}^{\delta_{m_k}}$ are plotted in Figure 5.8.

The interval size follows the same pattern as for the smooth background: we have a decrease of the estimate with increasing frequencies. In terms of magnitude, it appears that the estimates are more affected by the perturbation (direction) selected than the initial model.

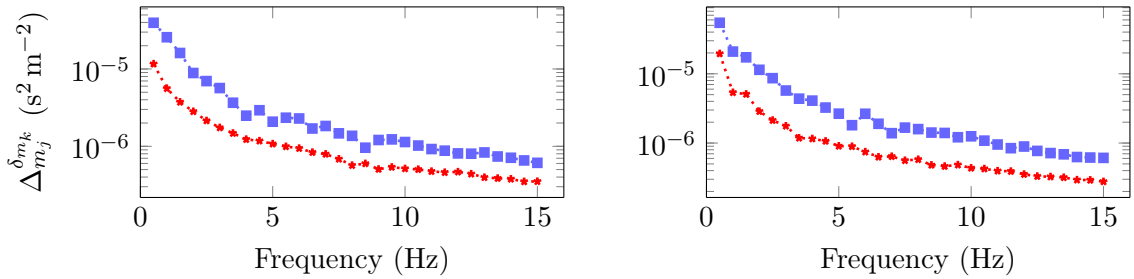
As a short conclusion for the iterative minimization algorithm, the frequency should evolve from low to high in order facilitate the convergence. When no prior information



(a) c_m: Marmousi velocity model.

(b) c_s: velocity including high contrast objects.

Figure 5.7: Acoustic models of size 9.2×3 km used as initial models for the estimation of $\Delta_{m_j}^{\delta_{m_k}}$, we take $c_j = m_j^{-1/2}$, $j = \{m, s\}$. The estimation are plotted in Figure 5.8.



(a) Using m_m from the Marmousi medium of Figure 5.7(a).

(b) Using m_s from model incorporating salt domes, see Figure 5.7(b).

Figure 5.8: Evolution of the interval size verifying the deflection condition, quantity $\Delta_{m_j}^{\delta_{m_k}}$, with frequency where m_j is the Marmousi model or a model containing salt domes, see Figure 5.7. The blue squares (··■··) employ one reflector (δ_{m_1}, Figure 5.5(a)) and the red stars (··★··) one large reflector (δ_{m_3}, Figure 5.5(c)).

is known for the initial model, one should take advantage of the larger interval size $\Delta_{m_j}^{\delta_{m_k}}$ given by low frequencies. Here we motivate the well-known frequency progression from a quantitative estimates of the size of the basin of attraction. However, it is not that simple because of the lack of low frequency information and their possible dependence on noise.

5.2.2 Frequency bandwidth data

We have seen that the sequential progression in frequency must be conducted from low to high regime. We now investigate the case of frequency bandwidth: when the forward problem encompasses not only one, but several frequencies. This approach is particularly common for time domain inverse problem where interval of frequencies can be successively considered. The forward problem is a vector of $\mathbb{C}^{(n_{rcv} \times n_{src} \times n_\omega)}$, where n_ω is the number of frequency taken in the subgroup.

We consider the smooth background of Figure 5.4 for m_0, motivated by its representation of an initial guess for reconstruction. We proceed to similar computations but using groups of frequency. Every group consists of ten frequencies. The initial group is the frequencies from 0.1 to 1 Hz using 0.1 Hz increment. Similarly, we design fifteen groups so that the last one is from 14.1 to 15 Hz, employing identical increment. The results are shown Figure 5.9.

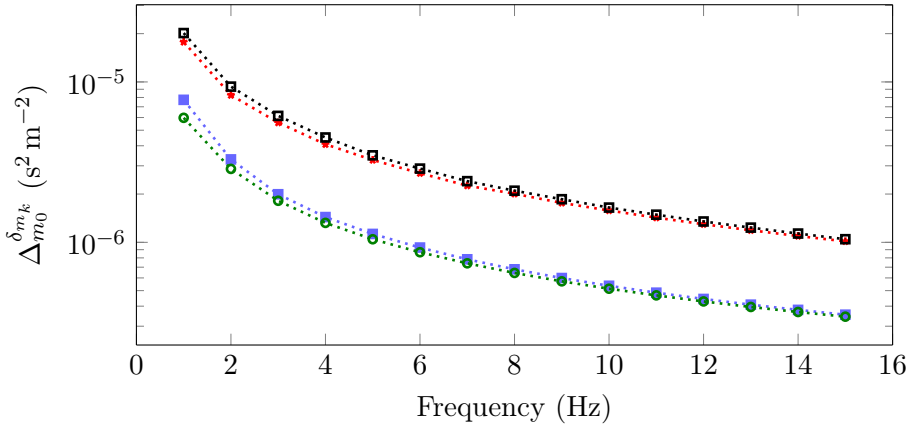


Figure 5.9: Evolution of the interval size verifying the deflection condition, quantity $\Delta_{m_0}^{\delta_{m_k}}$, with sequential or group of frequency using m_0 as a smooth velocity background and different directions δ_{m_k}. The blue squares (··■··) employ one large reflector (δ_{m_3}, Figure 5.5(c)) and subgroup of ten frequencies. The green circles (··o··) employ one large reflector (δ_{m_3}, Figure 5.5(c)) and sequential frequency. The red stars (··*··) employ reflectors extracted from the Marmousi model (δ_{m_m}, Figure 5.5(d)) and subgroup of ten frequencies. The black squares (··□··) employ reflectors extracted from the Marmousi model (δ_{m_m}, Figure 5.5(d)) and sequential frequency. For the group of frequency, the x axis indicates the largest frequency in the subgroup.

We basically see very few differences between the estimates using group of frequencies and the largest frequency in the group taken individually. In Table 5.1, we detail the numerical estimates and compare sequential and group of frequencies.

single frequency	$\Delta_{m_0}^{\delta_{m_k}}$	**frequency group**	$\Delta_{m_0}^{\delta_{m_k}}$
0.1 Hz	1.2×10^{-1}	$\{0.1, 0.2, \ldots, 1\}$ Hz	2.3×10^{-3}
1.0 Hz	2.3×10^{-3}	$\{1.1, 1.2, \ldots, 2\}$ Hz	1.1×10^{-3}
1.1 Hz	2.1×10^{-3}	$\{2.1, 2.2, \ldots, 3\}$ Hz	7.1×10^{-4}
2.0 Hz	1.1×10^{-3}	$\{3.1, 3.2, \ldots, 4\}$ Hz	5.2×10^{-4}
2.1 Hz	1.0×10^{-3}	$\{4.1, 4.2, \ldots, 5\}$ Hz	4.2×10^{-4}

Table 5.1: Evolution of $\Delta_{m_0}^{\delta_{m_k}}$ with frequency depending on the use of sequential or group of frequencies. Here δ_{m_k} is chosen to be the Marmousi directions, $\delta_{m_k} = \delta_{m_m}$, according to the Figure 5.5(d). $\Delta_{m_0}^{\delta_{m_k}}$ is given in $s^2\,m^{-2}$.

We observe that the estimates for the frequency group exactly match the largest frequency taken independently. Because the estimates decrease with frequency it means that using group of frequencies produces an unwanted decrease of the interval size compared with

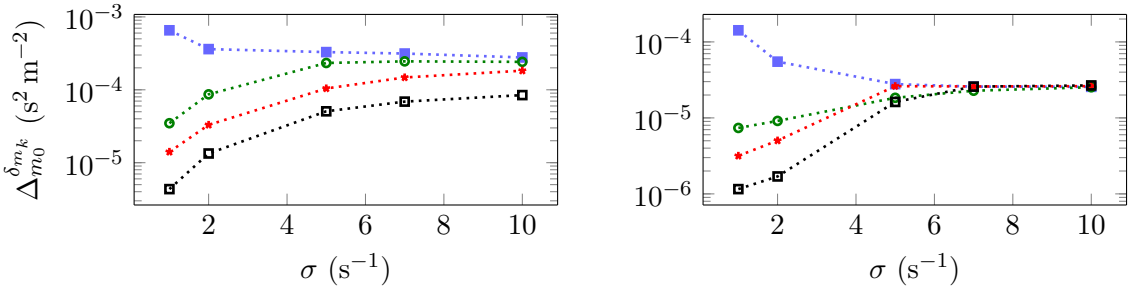
taking the lowest frequency in the group individually. In particular, for initial iterations, without prior information, it is important to consider the lowest frequency available alone instead of group of adjacent frequencies. This strategy is actually advocated in Brenders et al. (2012), motivated, as the frequency progression in Bunks et al. (1995), to reduce the cycle-skipping effect. Here we justify this choice by a quantitative estimation of the size of the radius of convergence. It has to be mitigated by the fact that certainly more information are contained in the group of frequencies, which for example can be more robust to noise. It only indicates that the basin of attraction is larger for individual frequency, which is an useful property without prior knowledge on the subsurface.

5.2.3 Complex frequencies

The use of complex frequencies (also referred to as the Laplace-Fourier domain) has shown advantageous behavior when used during reconstruction algorithm. In this case the frequency is expressed with Equation (1.25), which gives $-\omega^2 = (\sigma + 2\mathrm{i}\pi f)^2$. It incorporates a damping factor σ and the traditional component f (in Hz). Here we are interested in deciding how to select the progression of those coefficients to obtain the best convergence properties. Several applications have been carried out with progression that appears mostly intuitive. In the case of zero-frequency ($f = 0$), we mention Shin and Cha (2008), who conduct the iterative inversion from low to high damping coefficients. However, for a complex frequency progression ($f \neq 0$ and $\sigma \neq 0$), the evolution of the damping parameter at fixed f is proposed to be from high to low damping, as in Shin et al. (2010); Petrov and Newman (2014).

We compute the estimation of $\Delta_{m_0}^{\delta_{m_k}}$ with complex frequencies to see how it affects the basin of attraction. We consider m_0 the smooth background velocity and similar setup as employed in Subsection 5.2.1. In Figure 5.10, we compare the evolution with damping factor at fixed f, in particular we depict the choice $f = 0$, which is when only a damping is present, case which can be assimilated to the Laplace domain (Shin and Cha, 2008). We compare two selected directions: the Marmousi structures and the large reflector, illustrated in Figures 5.5(c) and Figures 5.5(d) respectively.

Two situations arise from the estimates: when the frequency only consists in the damping term ($f = 0$), then the interval decreases with increasing damping. However, when $f \neq 0$ the interval grows with increasing damping coefficients. Moreover, the largest size is obtained when $f = 0$. It means that when complex frequencies are available, one should start with the damping coefficient only, $f = 0$ and select increasing damping factor. Then, one can select $f \neq 0$ as low as possible and select the damping factors from high to low. Finally, the standard frequency progression when $\sigma = 0$ is conducted from low to high.



(a) Using direction δ_{m_m} (Marmousi reflectors).

(b) Using direction δ_{m_3} (single large reflector).

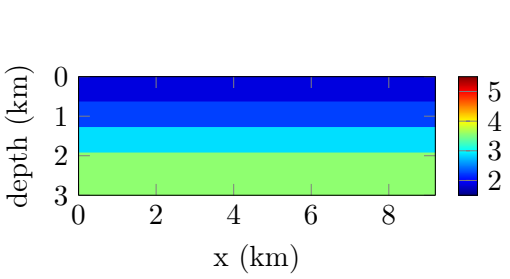
Figure 5.10: Evolution of the interval size verifying the deflection condition, quantity $\Delta_{m_0}^{\delta_{m_k}}$, with complex frequency. m_0 is taken to be the smooth background velocity of Figure 5.4. The angular frequency is defined by $-\omega^2 = (\sigma + 2i\pi f)^2$. The blue squares (··■··) use $f = 0$ Hz, the green circles (··●··) $f = 1$ Hz, the red stars (··★··) $f = 2$ Hz and the black squares (··■··) $f = 5$ Hz.

More generally we observe that the use of complex frequency (in particular when $f = 0$) increases notably the size of the interval, hence, it provides a fundamental benefit for the reconstruction when no information is initially known. It basically acts as a very low frequency (note that the magnitude of the estimates for complex frequencies eventually match the low frequency ones). This will be further illustrated in the numerical experiments of Chapter 6.

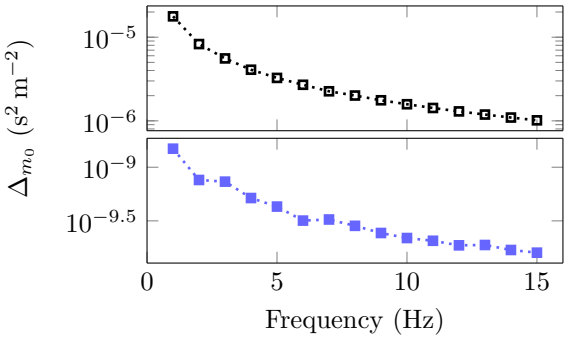
5.2.4 Lower bound estimates

The lower bound for the maximal size of the interval was obtained in the Equation (5.3). It involved the lowest singular value of the Fréchet derivative and the computation of the full second order derivative matrix. In order to perform the computation, we need to reduce the model space to limit the number of coefficients. This reduction is performed through compression where adjacent values are averaged to generate a coarse representation of the wave speed, similarly to the techniques introduced in Chapter 3, Section 3.4. We apply this to the smooth velocity background and show the resulting wave speed in Figure 5.11(a) where the number of coefficients has been reduced to 56. Using the benefit of this limited size, we can compute the lower bound Δ_{m_0}. We present the result in Figure 5.11(b), where we compare with the directional estimate $\Delta_{m_0}^{\delta_{mm}}$, i.e., when the reflectors extracted from the Marmousi model, δ_{m_m}, are used for the direction, see Figure 5.5(d).

The lower bound gives an important reduction compared to the directional estimates (four orders of magnitude). Yet it confirms the main property which is the decrease of the size of interval verifying the limited deflection with increasing frequencies. The lower frequencies are clearly beneficial for the iterative minimization.



(a) Compression of the acoustic starting model of size 9.2×3 km, where only 56 coefficients are kept to represent the velocity, indicated in $\mathrm{km\,s^{-1}}$.



(b) The lower bound of the interval size verifying the deflection condition is marked by the blue squares (··■··), the black squares (··□··) use the direction of the reflectors extracted from the Marmousi model (δ_{m_m}, Figure 5.5(d)).

Figure 5.11: Lower bound of the interval size computed from a wave speed in a reduced model space dimensions.

5.2.5 Influence of the data

The quantity of data to conduct the reconstruction necessarily has an influence on the size of the basin of attraction, and we expect its increase when increasing the amount of data available. In the original example of Subsection 5.2.1, we employed 19 sources and 183 receivers. Here we increase the quantity of data by incorporating sources and receivers on the side and bottom of the domain of interest. This configuration is sketched Figure 5.12.

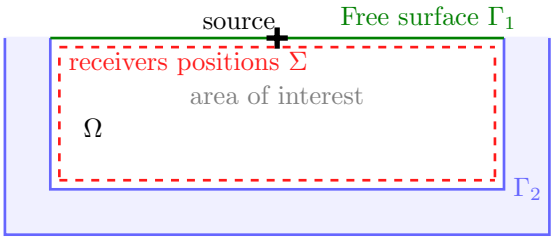


Figure 5.12: Illustration of the computational domain for the quantitative estimates of the quantity $\Delta_{m_0}^{\delta_{m_k}}$ assuming the data can be recovered on the four sides of the domain, here sources can potentially be placed on the sides as well.

More precisely, we consider a total of 50 sources positioned all around the domain and 484 receivers also positioned on the four sides. In Figure 5.13, we compare the estimate with frequency for the direction δ_{m_3} of Figure 5.5(c), which consists in a large reflecting object.

Increasing the amount of data available and incorporating bottom and sides illumination information provide an increase of the basin of attraction size. However, the difference

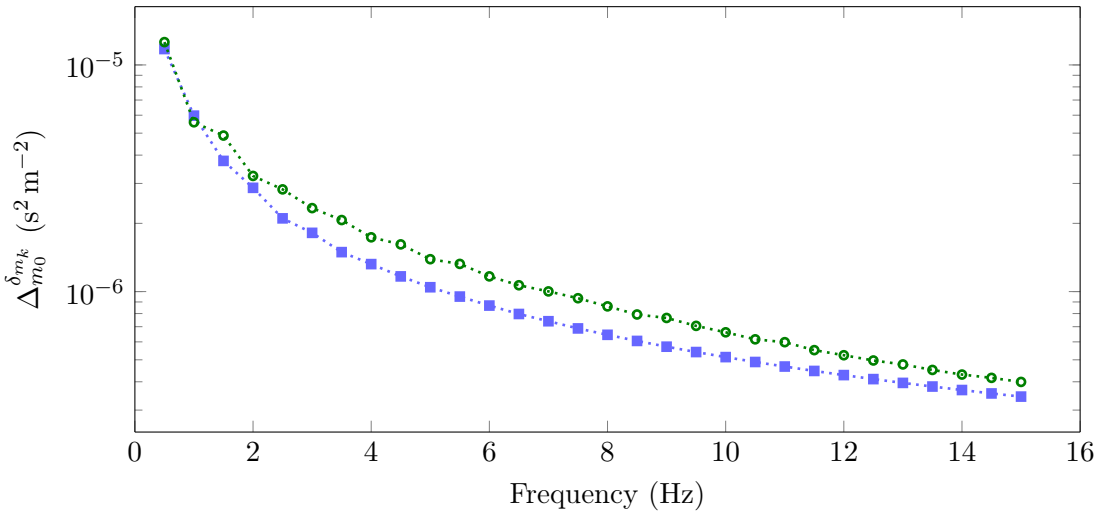


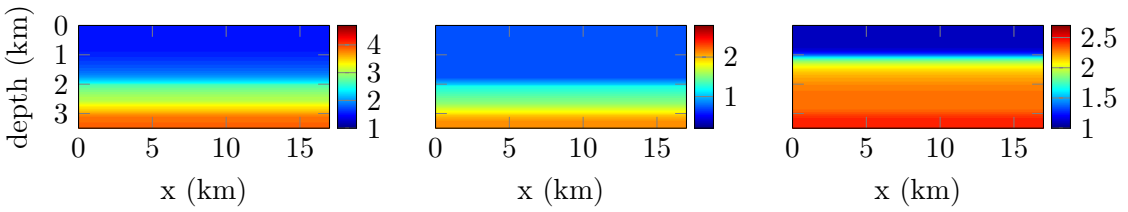
Figure 5.13: Evolution of the interval size verifying the deflection condition, quantity $\Delta_{m_0}^{\delta_{m_k}}$, with frequency using m_0 as smooth velocity background for direction δ_{m_3}. The blue squares (···■···) employ an acquisition of 19 sources and 183 receivers, which are only located at the top surface. the green circles (···o···) employ 50 sources and 484 receivers, which are located all around the domain of interest.

is not very important and one would expect a larger contrast between the situations. We might need to increase the size of the domain to see a larger contrast.

5.2.6 Elastic model reconstruction

It is natural to expect a similar behavior when stepping from acoustic to elastic and incorporating the multi-parameters reconstruction. This certainly increases the complexity of the reconstruction but the pattern of low frequencies improving the basin of attraction should be verified. Here we consider isotropic elastic media, characterized by three models: the Lamé parameters λ and μ, and the density ρ. The propagation of waves in such media follows Equation (1.29). In order to estimate the quantity $\Delta_{m_0}^{\delta_{m_k}}$, that validates the limited deflection property, we consider the models m_0 to be a smooth background, similarly to the previous acoustic case. They are illustrated in Figure 5.14.

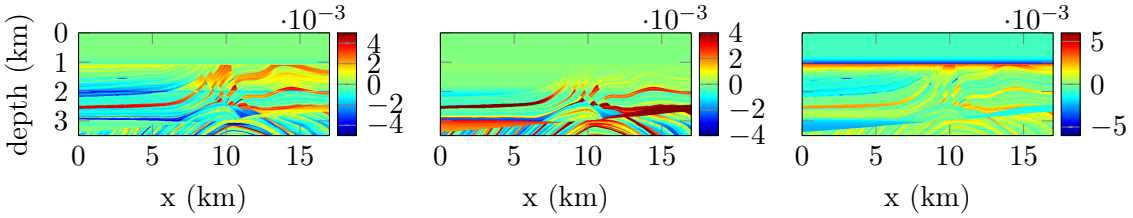
We follow a geophysical acquisition setup where we consider 19 sources and 168 receivers for each source, both being located at the surface. In this case, the derivatives have to be computed with respect to the three parameters. For the directions, we consider the elastic Marmousi medium and extract the different reflectors. The directions for the Lamé parameters and the density are presented in Figure 5.15 where the amplitude has been chosen so that $\|\delta_{m_m}\| = 1$. Because of the difference of magnitude between the different quantities, we actually employ a scaling parameter to obtain dimensionless coefficients



(a) P-velocity given in $\mathrm{km\,s^{-1}}$. (b) S-velocity given in $\mathrm{km\,s^{-1}}$. (c) Density given in $10^3\ \mathrm{kg\,m^{-2}}$.

Figure 5.14: Elastic smooth model of size 17×3.5 km used to define m_0.

such that the parameters of interest are $(\lambda/\lambda_0, \mu/\mu_0, \rho/\rho_0)$, where λ_0, μ_0 and ρ_0 are scalar coefficients given by the maximal values of the respective quantities, similarly to what we use to study the parametrization (Remark 5.1 below). For simplicity, we further omit the scaling coefficients λ_0, μ_0 and ρ_0 in the notation. In Figure 5.16, we show the evolution of the estimates $\Delta_{m_0}^{\delta_{m_m}}$ with frequency. We have considered that the forward problem is given by the measurements of displacement in the two spatial directions.



(a) Direction for λ (Pa). (b) Direction for μ (Pa). (c) Direction for ρ (10^3 $\mathrm{kg\,m^{-2}}$).

Figure 5.15: The elastic directions extracted from the Marmousi model. The amplitude is accordingly selected to give $\|\delta_{m_m}\| = 1$, where the parameters of interest are the Lamé coefficients and the density, such that $m = \{\lambda, \mu, \rho\}$.

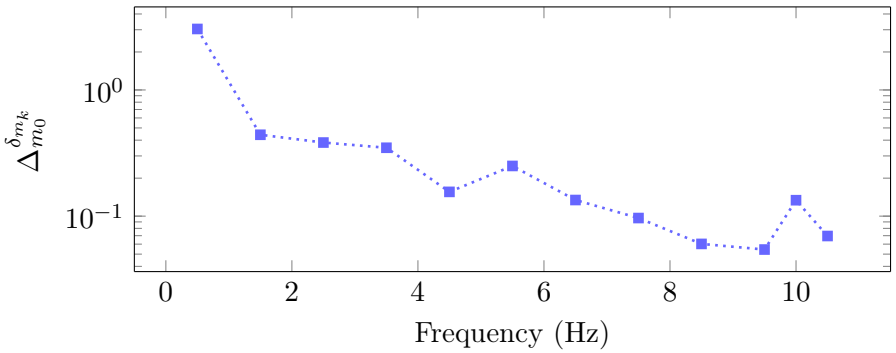
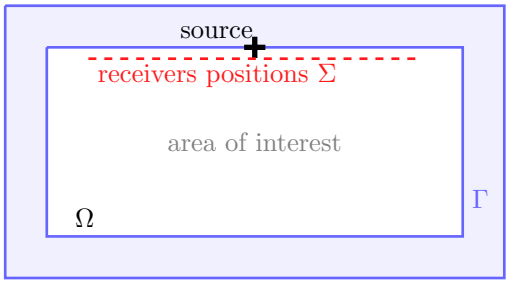


Figure 5.16: Evolution of the interval size verifying the deflection condition, quantity $\Delta_{m_0}^{\delta_{m_m}}$, with frequency using m_0 as a smooth elastic background and using the Marmousi structures of Figure 5.15 for the directions. The derivation uses scaled quantities: $(\lambda/\lambda_0, \mu/\mu_0, \rho/\rho_0)$.

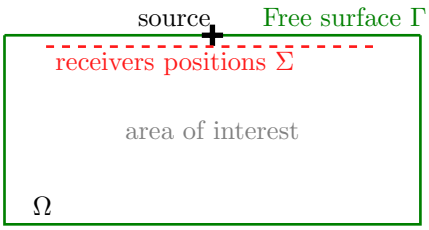
We see that the estimates of $\Delta_{m_0}^{\delta_{m_k}}$ for the least squares minimization algorithm associated with the elastic wave equation behave similarly to the acoustic case: the size of the interval so that the limited deflection property is obtained decreases with frequency.

5.2.7 Influence of the boundary conditions

The estimates we have derived are related to the least squares optimization problem and can be employed for other situations than the inverse wave problem affiliated with a geophysical setup. Let us first study the consideration of alternative boundary conditions for the domain of interest. We acknowledge two situations with a completely absorbing medium or Dirichlet condition for all boundaries ($p = 0$). They are sketched in Figure 5.17. The latest one is certainly closer to the medical imaging situation where free surface can be considered for all boundaries of the domain of interest.



(a) Computational domain using PML for all boundaries.

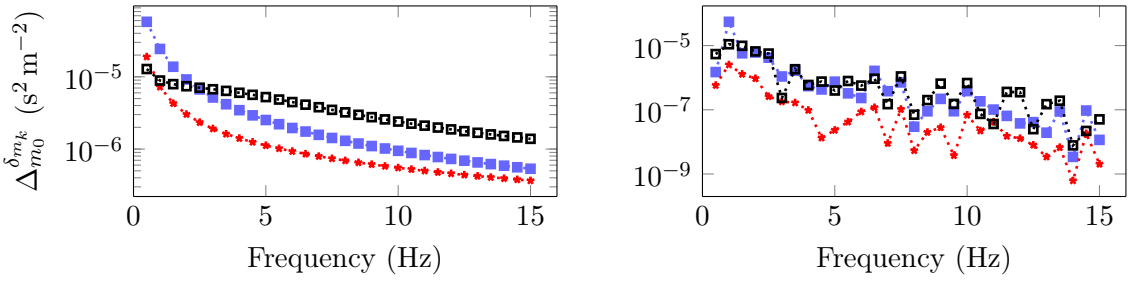


(b) Computational domain using Dirichlet boundary conditions only.

Figure 5.17: Illustration of different boundary conditions for the computational domain used to estimates the quantity $\Delta_{m_0}^{\delta_{m_k}}$.

We compute the estimates according to the new situations and show the new evolution of $\Delta_{m_0}^{\delta_{m_k}}$ with sequential frequency progression. We consider the smooth velocity background of Figure 5.4 to extract m_0 and the directions introduced in Figure 5.5 so we match the same context of Subsection 5.2.1, where only the boundary conditions of the medium are changed. In Figures 5.18, we show the respective evolution of $\Delta_{m_0}^{\delta_{m_k}}$ for Dirichlet and absorbing media.

The maximal size of the interval that verifies the deflection condition decreases with increasing frequencies independently of the situation, as expected. However, in the case of full Dirichlet boundary the behavior is much more chaotic and reveals a much higher magnitude in the evolution, namely it has four order of magnitude versus only two for the other situation. It indicates that the requirement for low frequency may even be more important in this case. Moreover we intuit that this behavior may be due to the multiple reflections introduced by the Dirichlet condition on all boundaries.



(a) Absorbing medium where all boundaries are treated as PML.

(b) Medium where all boundaries have a Dirichlet condition set to 0.

Figure 5.18: Evolution of the interval size verifying the deflection condition, quantity $\Delta_{m_0}^{\delta_{m_k}}$, with frequency where m_0 is the smooth background model. The blue squares (··■··) employ one reflector (δ_{m_1}, Figure 5.5(a)), the red stars (··★··) one large reflector (δ_{m_3}, Figure 5.5(c)) and the black squares (··□··) use reflectors extracted from the Marmousi model (δ_{m_m}, Figure 5.5(d)).

5.2.8 Perspectives

The quantitative estimates we have presented are general and can be established analogously to any optimization problem involving leas-squares minimization. We have developed the estimation of the size of the basin of attraction as an indication on how to control the convergence of our problem. In our case, it is realized via the choice of frequency, namely by initiating the iterations at the lowest available frequency to counterpart the lack of prior information. The quantitative estimates could easily be extended to many other situations in order to analyze the effect of a method or parametrization onto the problem, for example we can naturally think of:

- how does the choice of different norms of cost functions affect the estimates?
- Other inverse problems, based on different equations (e.g., Maxwell's equations) may present different patterns,
- does the type data (reflectivity in seismic) also affect the estimates?

5.3 Distance to the attainable set

The FC property requires the data not to be 'too far' from the attainable set. The maximal distance is materialized by the radius of curvature for which we have provided the numerical estimates in the direction δ_{m_k}: $\widehat{R}_{m_0}^{\delta_{m_k}}$, see Equation (5.4) and the illustration Figure 5.2. We follow the similar situations as in Section 5.2, where we estimate the size of the domain verifying the LD property, and estimate the maximum distance according to several directions. We take m_0 to be represented from the smooth background of Figure 5.4 and

the directions given in Figure 5.5. The structure of the successive estimations is similar to Section 5.2.

5.3.1 Geometry and sequential frequency

The maximal distance between the data and the attainable set according to different directions is plotted with respect to the frequency in Figure 5.19.

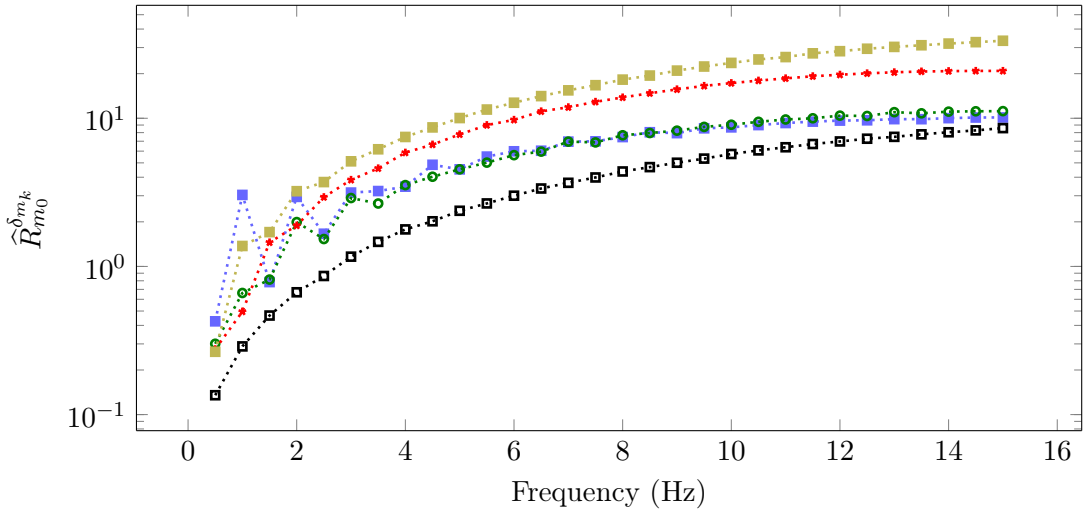


Figure 5.19: Evolution of the maximal distance between the data and the attainable set, quantity $\widehat{R}_{m_0}^{\delta_{m_k}}$, with frequency using m_0 as a smooth velocity background and different directions δ_{m_k}. The blue squares (··■··) employ one reflector (δ_{m_1}, Figure 5.5(a)), the green circles (··●··) two reflectors (δ_{m_2}, Figure 5.5(b)) the red stars (··★··) one large reflector (δ_{m_3}, Figure 5.5(c)) and the black squares (··■··) reflectors extracted from the Marmousi model (δ_{m_m}, Figure 5.5(d)). The yellow squares (··■··) represent the estimates using constant background direction.

We observe that the maximal distance allowed for the data increases with frequency. The different directions also affects the relative magnitude; in particular, the directions extracted form the Marmousi model reduce the maximal distance allowed between the data and the attainable set compared to having one large reflector. It is the opposite compared to the behavior observed to verify the deflection property, where the Marmousi directions are advantageous. We note that the behavior is opposed to the interval size verifying the deflection. Here increasing the frequency increases the acceptable distance to the attainable set. It however follows the stability analysis carried in Chapter 3, where increasing the frequency is shown to improves the stability.

5.3.2 Frequency bandwidth data

We reproduce the experiment where frequencies are summed so that not only one but several frequencies are contained simultaneously in the data. We follow the similar approach as for the estimates relative to the deflection and choose a group of frequencies with 0.1 Hz step. We create fifteen groups, each of ten frequencies such that the first one is $\{0.1, 0.2, \dots, 1\}$ Hz and the last one is $\{14.1, 14.2, \dots, 15\}$ Hz. In Figure 5.20 we compare the evolution between sequential and group of frequencies for the large reflector direction δ_{m_3}, and the reflectors extracted from the Marmousi medium δ_{m_m}. The estimates for the group of frequencies are plotted at the horizontal position corresponding to the largest frequency in the group (i.e. integer frequency).

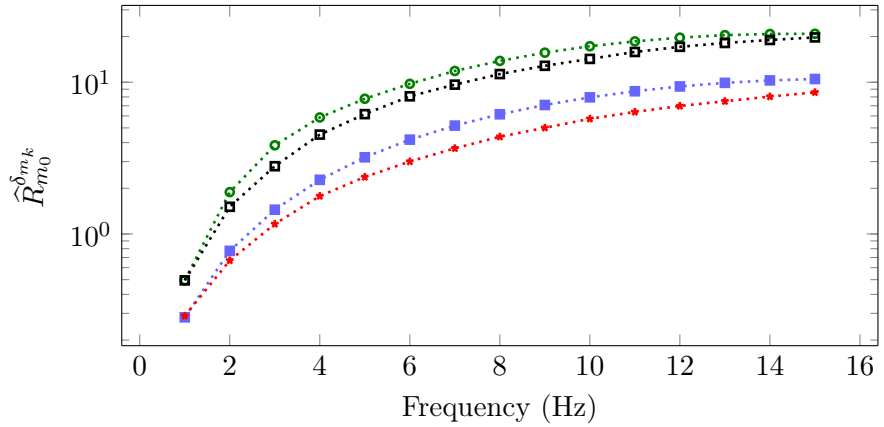


Figure 5.20: Evolution of the maximal distance between the data and the attainable set, quantity $\widehat{R}_{m_0}^{\delta_{m_k}}$, with sequential or group of frequency using m_0 as a smooth velocity background and different directions δ_{m_k}. The blue squares (··■··) employ one large reflector (δ_{m_3}, Figure 5.5(c)) and subgroup of frequency. The green circles (··○··) employ one large reflector (δ_{m_3}, Figure 5.5(c)) and sequential frequency. The red stars (··★··) employ reflectors extracted from the Marmousi model (δ_{m_m}, Figure 5.5(d)) and subgroup of frequency. The black squares (··□··) employ reflectors extracted from the Marmousi model (δ_{m_m}, Figure 5.5(d)) and sequential frequency. For the groups of frequencies, the x axis indicates the largest frequency in the subgroup.

We observe some differences between the type of reflectors, as in Figure 5.19. Contrary to the basin of attraction, we clearly see the benefits of using several frequencies in the data. It provides a noticeable improvement in the maximal distance between the observation and the attainable set which gives the finite curvature property. One can imagine that in this case, the frequency bandwidth acts as some regularization of the problem by adding information. Therefore, we see two different situations: the low frequency improves the convergence by increasing the basin of attraction (observed from the deflection estimates), but high frequencies allow an increase of the radius of curvature, thus an increase of the distance between the data and the attainable set. Furthermore, using sequential frequency provides benefits to the radius of convergence, but using several frequencies increase the radius of curvature.

5.3.3 Lower bound estimates

When reducing the number of coefficients to represent the model, we can compute the lowest singular value of the Fréchet derivative and following lower bound $\widehat{R}_{m_0}$ with Equation (5.5). We follow the realization of Subsection 5.2.4 and generate an initial model m_0 represented with 56 coefficients, see Figure 5.11(a). In Figure 5.21, we compare the lower bound with directional estimates using the reflectors extracted from the Marmousi model.

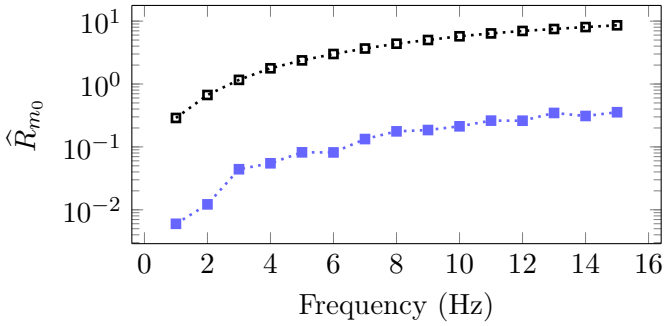


Figure 5.21: The lower bound of the maximal distance between the data and the attainable set is marked by the blue squares (··■··), the black squares (··□··) use the direction of the reflectors extracted from the Marmousi model (δ_{m_m}, Figure 5.5(d)). We have reduced the representation of the model according to Figure 5.11(a).

The lower bound behaves as expected but still shows an important reduction compared to the directional estimates. There is approximately one or two orders of magnitude difference between the lower bound and directional estimates. It confirms the general behavior that the distance between the data and the attainable set increases with frequency.

5.3.4 Influence of the data

We increase the quantity of data available by incorporating sources and receivers on the sides and bottom of the domain, following Figure 5.12. We now take 50 sources positioned all around the domain and 484 receivers also positioned on the four sides. In Figure 5.22, we compare the estimates of the curvature with frequency for the large reflection δ_{m_3} of Figure 5.5(c).

The incorporation of data from all directions gives an important improvement of the possible distance with the attainable set, particularly for low frequencies. Higher frequencies are less affected but we still observe the benefits of adding information.

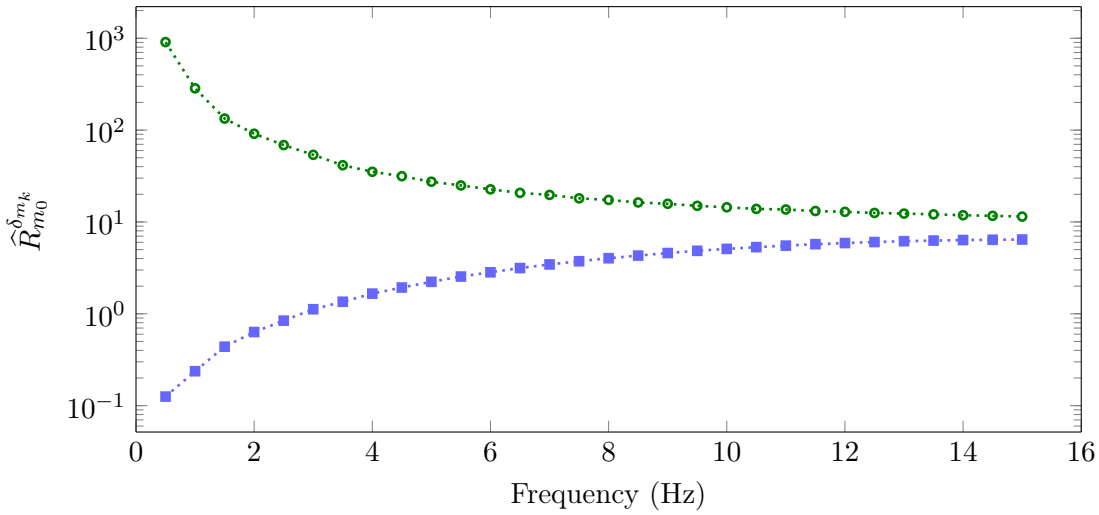


Figure 5.22: Evolution of the maximal distance between the data and the attainable set, quantity $\widehat{R}_{m_0}^{\delta_{m_k}}$, with frequency using m_0 as smooth velocity background for direction δ_{m3}. The blue squares (···■···) employ an acquisition of 19 sources and 183 receivers, which are only located at the top surface. the green circles (···o···) employ 50 sources and 484 receivers, which are located all around the domain of interest.

5.3.5 Elastic model reconstruction

We review the estimates in the context of elastic reconstruction, following the situation of Subsection 5.2.6. It is a multi-parameters reconstruction where the quantities of interest are the density ρ and the Lamé parameters λ and μ. m_0 is the smooth background depicted in Figure 5.14; the direction, extracted from the Marmousi model, has been introduced in Figure 5.15. The seismic acquisition consists in 19 sources and 168 receivers for each sources, located at the surface. In Figure 5.23, we plot the estimates of $\widehat{R}_{m_0}^{\delta_{m_k}}$ for this elastic medium inversion.

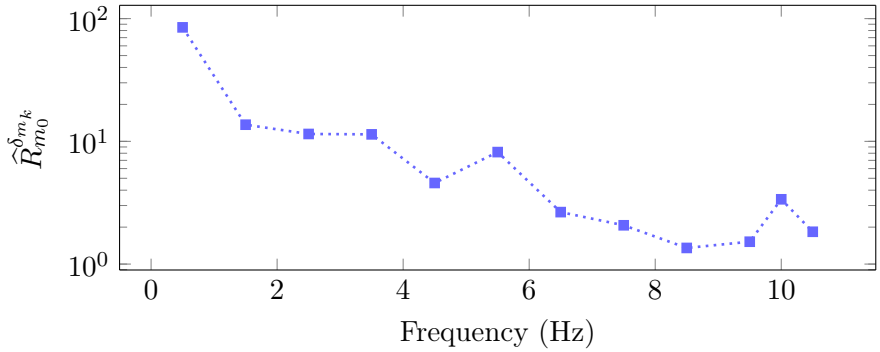
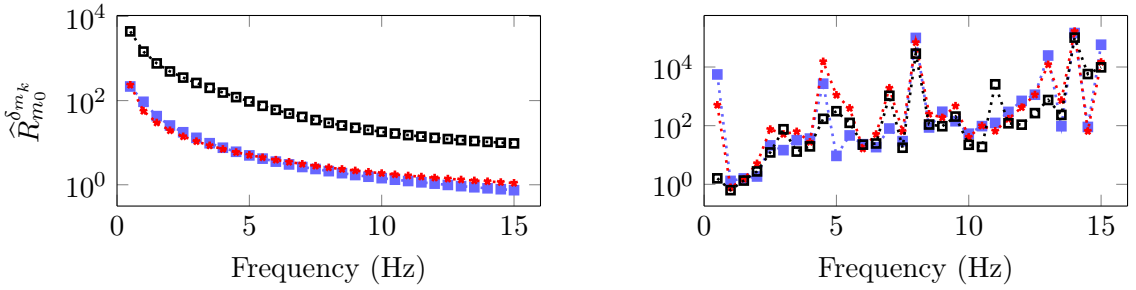


Figure 5.23: Evolution of the maximal distance between the data and the attainable set, quantity $\widehat{R}_{m_0}^{\delta_{m_k}}$, with frequency using m_0 as a smooth elastic background and using the Marmousi structures of Figure 5.15 for the directions. The derivation uses scaled quantities: $(\lambda/\lambda_0, \mu/\mu_0, \rho/\rho_0)$.

The maximal distance between the data and the attainable set appears now to be decreasing with frequency, contrary to the acoustic situation. However, the final magnitude is overall comparable to the acoustic case.

5.3.6 Influence of the boundary conditions

Many other configurations could be probed, similarly to what we discussed for the estimation of the model size verifying the LD property. We illustrate the case of alternative boundary conditions, introduced in Subsection 5.2.7. It consists in representing the domain with absorbing or Dirchlet boundary conditions on all sides, see Figure 5.17. The associated estimates of the maximal distance between the data and the attainable set is plotted in Figure 5.24.



(a) Medium where all boundaries are treated as PML.

(b) Medium where all boundaries have a Dirichlet condition set to 0.

Figure 5.24: Evolution of the maximal distance between the data and the attainable set, quantity $\widehat{R}_{m_0}^{\delta_{m_k}}$, with frequency where m_0 is the smooth background model. The blue squares (··■··) employ one reflector (δ_{m_1}, Figure 5.5(a)), the red stars (··★··) one large reflector (δ_{m_3}, Figure 5.5(c)) and the black squares (··□··) use reflectors extracted from the Marmousi model (δ_{m_m}, Figure 5.5(d)).

The medium where all boundaries are absorbing presents a decrease of the distance with increasing frequencies, contrary to the other setups. As for the deflection estimates, the medium with Dirichlet boundary condition renders a more perturbed evolution, probably induced by the multiple reflections.

5.4 Analysis of parametrization for FWI

5.4.1 Parametrization

We have briefly mentioned the concept of parametrization in Subsection 4.4.6. The choice of a parametrization is crucial because different parameters can cost abnormal influences on

the data, directly affecting the behavior of the iterative reconstruction. For instance, the elastic isotropic version gives three unknown models and $m = \{\lambda, \mu, \rho\}$. This may appear as the 'natural' selection following the wave equation. However, many possibilities are available, following some quantities introduced in Subsection 1.4.3. For multi-parameters reconstructions, we have already mentioned the selection advanced by Tarantola (1986), suggesting that the parameters should be as uncorrelated as possible.

The choice of a parameter affects the computation of the gradient and Hessian (Pratt et al., 1998), and consequently the iterative reconstruction. Based on numerical experiments, Tarantola (1986) further suggests the use of the density, P-wave impedance and S-wave impedance for elastic inversion. The use of the impedance is also promoted in Lavaud et al. (1999) in the case of Amplitude Versus Offset (AVO). More recently, FWI test cases illustrating this idea have been provided by Brossier (2011); Köhn et al. (2012). In general, the density has been recognized for being more complicated to retrieve (Virieux and Operto, 2009). Jeong et al. (2012) propose a two-step strategy to overcome the issue, inverting the density after the other parameters. For acoustic reconstruction, Prieux et al. (2013a) compare the combination of two parameters between wave speed, density and impedance with diffraction patterns and decide that the combination of the wave speed and density is the best. The study is pursued in Prieux et al. (2013b) for elastic media where P- and S-wave speeds are selected. As far as an analytical study of the inverse problem is considered, Beretta et al. (2014) show that the density and the Lamé parameters must be considered separately in elasticity. Extension can be carried out for anisotropy, where the stiffness tensor coefficient C_{ij} can be of interest. In the case of VTI media, a sensitivity analysis was conducted by Kamath and Tsvankin (2016) using the time domain wave equation. Guitton and Alkhalifah (2017) also compares the reconstruction for VTI media. We also mention the work of Tromp et al. (2005) in seismic tomography where Fréchet kernels based on the bulk and shear moduli, as well as the P- and S-velocities, are computed.

Another consideration is the application of a function of the parameter. For instance, we can consider the acoustic situation with constant density. Following the Helmholtz equation, there is only one parameter to be retrieved, which is the velocity. However, one can choose to invert the slowness (c^{-1}) or the slowness squared (c^{-2}) instead of the velocity (c). The logarithmic of the parameter is also advanced, see Tarantola (2005); de Hoop et al. (1999).

This change of parametrization may appear as a tool at first glance but we will illustrate how it becomes critical. It first implies a modification of the gradient of the cost function (given Equation (4.16)). We assume that $m = \{m_i\}$ is composed of several parameters, and apply the chain rule to consider an alternative for one of the parameter, $\tilde{m}_k$. The

derivation follows to be

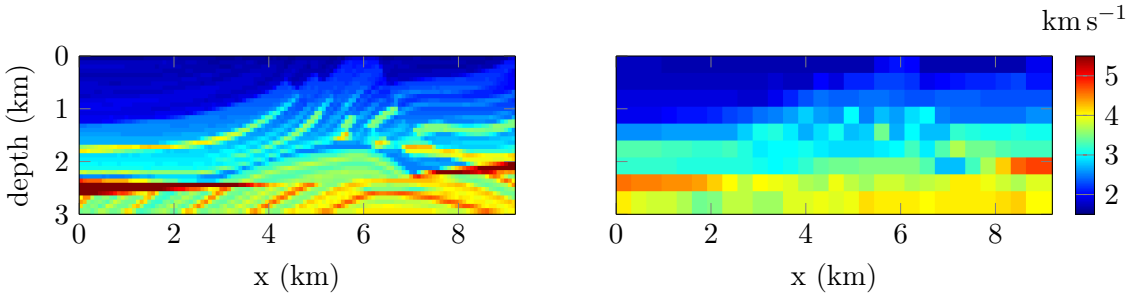
$$\frac{\partial}{\partial \tilde{m}_k}\Big(\mathcal{F}(m)\Big) = \sum_i \frac{\partial \mathcal{F}(m)}{\partial m_i}\frac{\partial m_i}{\partial \tilde{m}_k}.$$

Traditionally we compute the gradient with respect to a reference parametrization, $\partial_m \mathcal{F}$ and then apply the change by computing $\partial_{\tilde{m}} m$, usually consisting in simple numerical operations.

5.4.2 Model partition

We propose to analyze the effect of the parametrization onto the Fréchet derivative, in terms of its condition number and singular values. The computation of the full matrix Fréchet derivative may be expensive, and consequently is the singular values decomposition. That is why we consider the reduced model representation to achieve efficiently the numerical computations. The model representation has been introduced in Chapter 3 for the purpose of the stability, and illustrated in Section 3.4.

We restrict ourselves to the piecewise constant model representation. Let us recall the model partition with Marmousi wave speed of size 9.2×3 km. The number of coefficients to represent the model is denoted by N where reducing N implies a coarser velocity representation, as illustrated in Figure 5.25.



(a) Representation with 4 255 coefficients.

(b) Representation with 252 coefficients.

Figure 5.25: The Marmousi model of size 9.2×3 km represented with two different partitions. This process can be assimilated with compression in order to reduce the size of the model space.

5.5 Acoustic parametrization

5.5.1 Acoustic models

In the case of acoustic media, the subsurface is characterized by the density and the bulk modulus. Alternative quantities can be extracted, the most commons are

$$\begin{aligned} c(\mathbf{x}) &= \sqrt{\frac{\kappa(\mathbf{x})}{\rho(\mathbf{x})}} && \text{wave speed,} \\ I(\mathbf{x}) &= \sqrt{\rho(\mathbf{x})\kappa(\mathbf{x})} && \text{impedance.} \end{aligned} \tag{5.6}$$

In the case where the density is constant, the propagation of waves follows the Helmholtz equation (1.27), and only the velocity characterizes the medium so that $m := c$. In this case the possible parametrizations are only given by functions of the velocity. The standard ones have been mentioned above, they consist in the slowness (c^{-1}) and the slowness squared (c^{-2}). It is straightforward to obtain the appropriate Fréchet derivative

$$\begin{cases} \dfrac{\partial \mathcal{F}(c)}{\partial c^{-1}} = -c^2 D\mathcal{F}(c), & \text{slowness parametrization,} \\ \dfrac{\partial \mathcal{F}(c)}{\partial c^{-2}} = -\dfrac{c^3}{2} D\mathcal{F}(c), & \text{slowness squared parametrization,} \end{cases} \tag{5.7}$$

where $D\mathcal{F}(c)$ stands for the derivative with respect to c.

Let us now consider the acoustic case where the density is not constant, see Equation (1.28). We define the model to reconstruct to be $m = \{\kappa, \rho\}$, motivated by the structure of the equation. The other quantities that have been defined from those parameters are the velocity and the impedance, see System (5.6). In the context of multi-parameters inversion, it is important to notice that changing one of the parametrization may affect the other one as well. Assuming the gradient has been computed with respect to κ and ρ we introduce the notation

$$\mathcal{F}(\kappa, \rho) = \tilde{\mathcal{F}}(m_1, m_2),$$

where m_1 and m_2 represent the alternative parameters (velocity, impedance, etc) but may also be unchanged (here we only require $m_1 \neq m_2$). The choice of parametrization defines which quantities are assumed to be independent in the wave equation. The appropriate gradients with respect to these chosen parameters are given by

$$\begin{cases} \dfrac{\partial}{\partial m_1}\left(\tilde{\mathcal{F}}(m_1, m_2)\right) = \dfrac{\partial \mathcal{F}(\kappa,\rho)}{\partial \kappa}\dfrac{\partial \kappa}{\partial m_1} + \dfrac{\partial \mathcal{F}(\kappa,\rho)}{\partial \rho}\dfrac{\partial \rho}{\partial m_1}, \\ \dfrac{\partial}{\partial m_2}\left(\tilde{\mathcal{F}}(m_1, m_2)\right) = \dfrac{\partial \mathcal{F}(\kappa,\rho)}{\partial \kappa}\dfrac{\partial \kappa}{\partial m_2} + \dfrac{\partial \mathcal{F}(\kappa,\rho)}{\partial \rho}\dfrac{\partial \rho}{\partial m_2}. \end{cases}$$

Let us assume we select $m_1 = \kappa$. It is important to remark that the term $\dfrac{\partial \rho}{\partial m_1}$ is zero only if $m_2 = \rho$, otherwise κ and ρ are not considered independent anymore and the gradient with respect to κ has to be updated accordingly. Based on original gradients computed with respect to κ and ρ, we give some alternatives in Table 5.2.

parametrization (m_1, m_2)	$\dfrac{\partial \kappa}{\partial m_1}$	$\dfrac{\partial \rho}{\partial m_1}$	$\dfrac{\partial \kappa}{\partial m_2}$	$\dfrac{\partial \rho}{\partial m_2}$
(κ, c)	1	c^{-2}	0	$-2\kappa c^{-3}$
(κ, I)	1	$-I^2\kappa^{-2}$	0	$2I\kappa^{-1}$
(c, ρ)	$2c\rho$	0	c^2	1
(I, ρ)	$2I\rho^{-1}$	0	$-I^2\rho^{-2}$	1
(I, c)	$2I\rho^{-1}$	$2I\kappa^{-1}$	$2c\rho$	$-2\kappa c^{-3}$

Table 5.2: Derivative formula for different acoustic parametrization assuming the original derivations were carried out with respect to κ and ρ.

Following the choice of parameters, one can then apply a general function such as the inverse, the inverse squared, or the logarithm following some simple calculations, for example,

$$\begin{cases} \dfrac{\partial \mathcal{F}(m)}{\partial m^{-1}} = -m^2 D\mathcal{F}(m), & \text{inverse parametrization,} \\ \dfrac{\partial \mathcal{F}(m)}{\partial m^{-2}} = -\dfrac{m^3}{2} D\mathcal{F}(m), & \text{inverse squared parametrization,} \\ \dfrac{\partial \mathcal{F}(m)}{\partial \log(m)} = m D\mathcal{F}(m), & \text{logarithmic parametrization,} \end{cases}$$

5.5.2 Acoustic Fréchet derivative, constant density

Let us first consider the inverse reconstruction of an acoustic isotropic medium that is defined with a single model: the wave speed c, and where the waves are prescribed by the Helmholtz equation (1.27). It follows that the Fréchet derivative is represented with a matrix of size $[(n_{rcv} n_{src} n_\omega) \times N]$, where n_{src} is the number of sources, n_{rcv} is the number of receivers per source and n_ω is the number of simultaneous frequencies. For this experiment, the acquisition set consists of 19 sources and 183 receivers for each sources, following the situation we already employed in Sections 5.2 and 5.3, and we refer to the illustration of Figure 5.3.

The parametrizations that we consider here are the velocity (c), the slowness ($1/c$) and

the slowness squared ($1/c^2$). The change of variable has been given in the Equation (5.7). In order to study the influence of the geometry onto the parametrization, we introduce three different acoustic wave speeds:

- Marmousi velocity, presented in Figures 5.7 and 5.25, where the latest one shows the effect of partitioning,
- smooth velocity background, presented in Figure 5.4,
- smooth velocity background with strong contrasting objects, which is given in Figure 5.7.

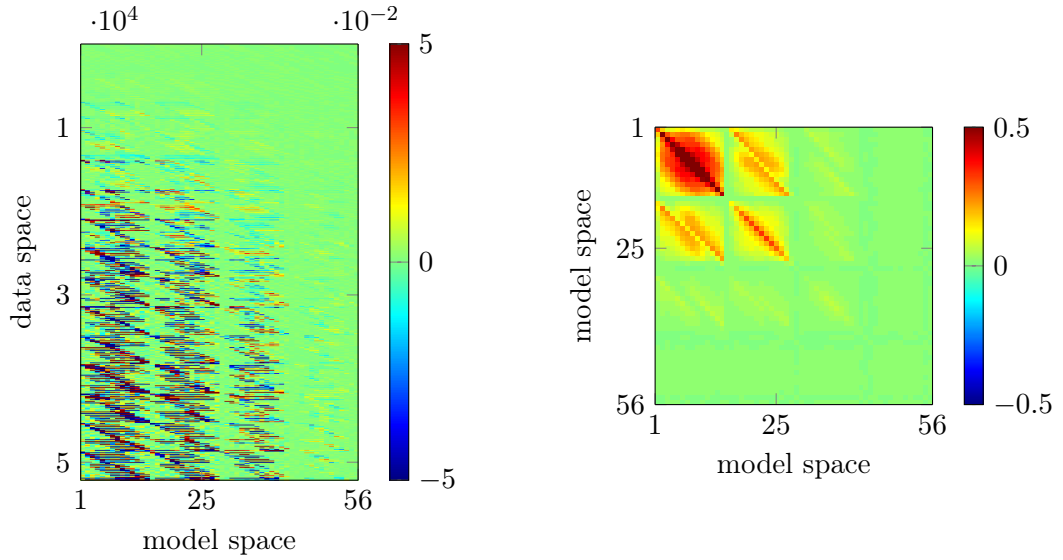
5.5.2.1 Matrix representation using the Gauss-Newton Hessian

We compute the Fréchet derivative for different partitionings and parametrizations. We consider the concatenation of all frequencies of interest, from 1 to 15 Hz, with 1 Hz increment. Eventually, the Fréchet derivative is a complex matrix of size $[(n_{rcv}n_{src}n_{\omega})\times N] = [52\,155 \times N]$. As an alternative, one can represent the Gauss-Newton Hessian, which is given by $H_{GN} = D\mathcal{F}^* D\mathcal{F}$, see Section 4.5 for further information. The advantage of the Gauss-Newton Hessian is that it is a symmetric, squared matrix, of size $N \times N$. In Figure 5.26, we compare the Fréchet derivative and Gauss-Newton Hessian associated with the Marmousi wave speed. In Figure 5.27, we compare the Gauss-Newton Hessian matrices depending on the parametrization for the acoustic Marmousi model. The original matrices are scaled with their respective maximum values so that they can all fit on the same scale.

We observe the effect of the parametrization onto the Gauss-Newton Hessian. The slowness squared parametrization, $1/c^2$, provides much more insight into the values meanwhile the velocity parametrization c restricts the resolution to the upper left of the Hessian. The slowness parametrization, $1/c$, gives an intermediate result. We further refer to Pratt et al. (1998) where an interpretation of the Gauss-Newton Hessian and its underlying pattern can be found. From this observation we expect the slowness squared parametrization to allow a better reconstruction as the magnitude of information is better shared.

In Figure 5.28, we compare the effect of the velocity model, comparing the Marmousi wave speed and the medium encompassing salt domes, see Figure 5.7, using the slowness squared parametrization.

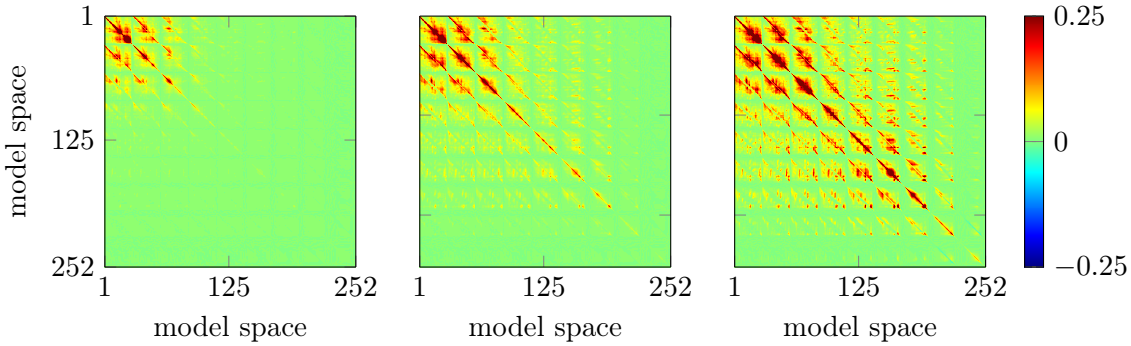
There is an important difference between the two media. the Gauss-Newton Hessian resulting from the use of the model with few high contrast reflectors shows a pattern that is more 'straight'. Clearly, the reflection induced by the many structures of the Marmousi



(a) Fréchet derivative $D\mathcal{F}$.

(b) Gauss-Newton Hessian $D\mathcal{F}^* D\mathcal{F}$.

Figure 5.26: Comparison of the Fréchet derivative and Gauss-Newton Hessian for the acoustic Marmousi wave speed. The data cover a set of frequency from 1 to 15 Hz, using 1 Hz increment. The partitioning is selected to generate $N = 56$ coefficients, this corresponds to the size of the model space.



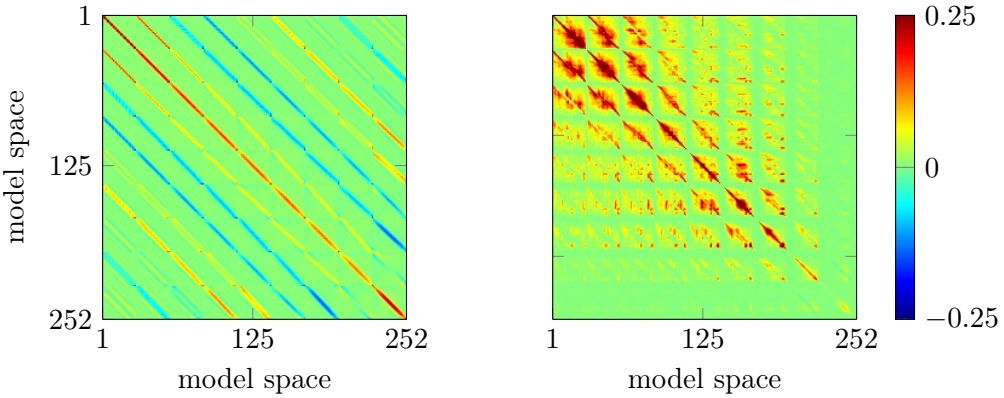
(a) Parameterization c.

(b) Parameterization $1/c$.

(c) Parameterization $1/c^2$.

Figure 5.27: Comparison of the Gauss-Newton Hessian, $D\mathcal{F}^* D\mathcal{F}$, for the acoustic Marmousi wave speed. The data cover a set of frequency from 1 to 15 Hz, using 1 Hz increment. The partitioning is selected to generate $N = 252$ coefficients, this corresponds to the size of the matrix. The matrices are scaled with their respective maximum value to fit on the same scale.

model are revealed somehow in the matrix while the only few reflections of the salt model produce a very different pattern.



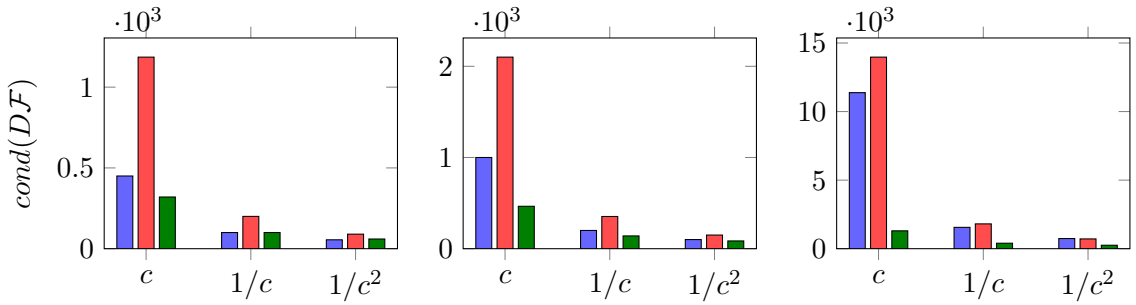
(a) Using the velocity model with high contrast objects, see Figure 5.7(b).

(b) Using the Marmousi velocity model, see Figure 5.7(a).

Figure 5.28: Comparison of the Gauss-Newton Hessian, $D\mathcal{F}^* D\mathcal{F}$, for the models of Figure 5.7, using the slowness squared parametrization. The data cover a set of frequency from 1 to 15 Hz, using 1 Hz increment. The partitioning is selected to generate $N = 252$ coefficients, this corresponds to the model space and size of the matrix. The matrices are scaled with their respective maximum value to fit on the same scale.

5.5.2.2 Condition number of the Fréchet derivatives

In Figure 5.29, we compare the condition number of the Fréchet derivative associated with the velocity c, the slowness $1/c$, and the slowness squared $1/c^2$ for different model partitionings. We also compare the effect of different velocity models: Marmousi, the salt model and the smooth model.



(a) $N = 56$.

(b) $N = 252$.

(c) $N = 1\,026$.

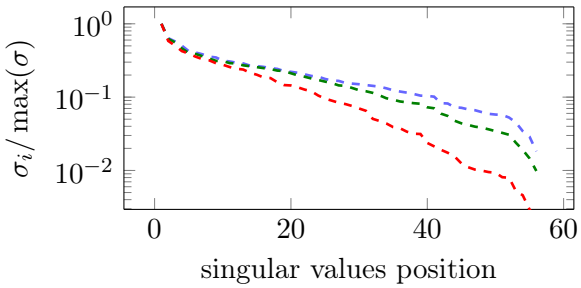
Figure 5.29: Comparison of the Fréchet derivative condition number for different parametrizations and different partitionings. The medium assume a constant density $\rho = 100\ \mathrm{kg\,m^{-2}}$. The blue bars ( ) indicate the use of the Marmousi velocity (of Figure 5.7(a)), the red bars ( ) the salt velocity (of Figure 5.7(b)), and the green bars ( ) the smooth acoustic velocity (of Figure 5.4).

There is a clear benefit in using the slowness or slowness squared compared to the velocity. Finally, the use of slowness squared seems to give the smallest condition number. When we increase the partition, the number of unknowns increases, which impacts on the

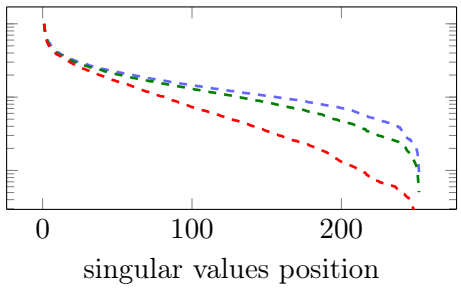
condition number as well. We remark that the salt model generates a higher condition number, reflecting its higher reconstruction complexity, result that we also observed with the deflection analysis of Section 5.2.

5.5.2.3 Singular values of the Fréchet derivatives

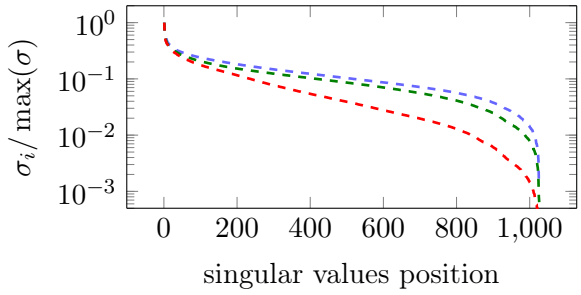
In Figure 5.30, we compare the evolution of the singular values of the global Fréchet derivative (encompassing all frequencies) for different partitionings and depending on the parametrization. The Marmousi velocity has been employed for this experiment. The singular values are scaled with their respective largest value. The slowness squared parametrization gives much better results as the singular values rate of decrease is much less than the velocity and slowness parametrization. It indicates that this parametrization is more robust to noise.



(a) $N = 52$.



(b) $N = 252$.

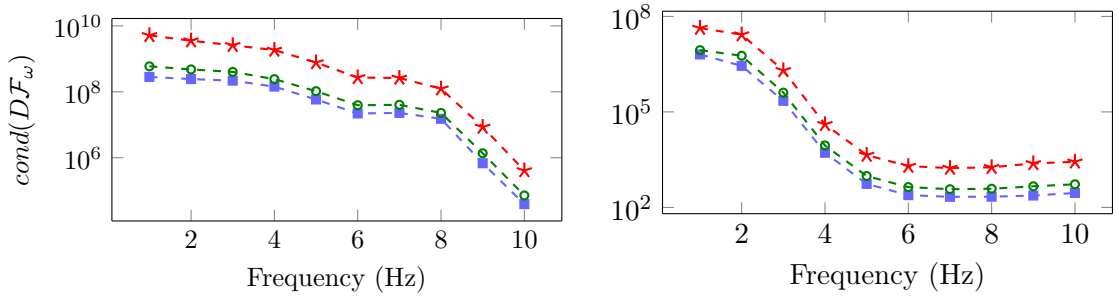


(c) $N = 1026$.

Figure 5.30: Evolution of the Fréchet derivative singular values for different partitioning. The blue dashed line (- - -) represents the slowness squared parametrization, the green dashed line (- - -) the slowness parametrization and the red dashed line (- - -) the velocity parametrization. This experiment uses the Marmousi velocity (of Figure 5.7(a)).

5.5.2.4 Sequential frequency progression

We now separate every frequency and compute the Fréchet derivative associated with individual frequency. In Figure 5.31, we compare the evolution of the Fréchet derivative condition number with frequency for different partitioning and parametrization, using the Marmousi wave speed. Once again, the slowness squared parametrization gives the lowest condition number, which indicates the benefit. We observe that the condition number decreases with frequency, which is consistent with the stability analysis. Indeed, we have shown in Chapter 3 that the stability constant estimates decrease with frequency, and we expect the condition number to decrease with frequencies. Then increasing the partition naturally increases the condition number, because of the addition of unknowns in the problem.



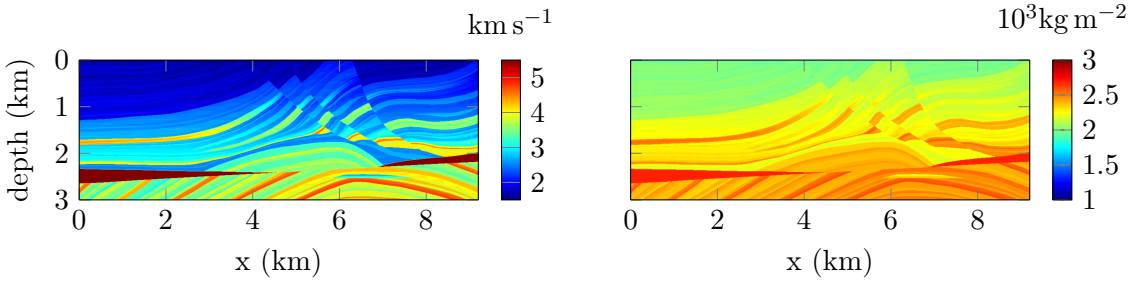
(a) $N = 1026$. (b) $N = 252$.

Figure 5.31: Evolution of the Fréchet derivative condition number with frequency for different partitioning. The blue squares (- ■ -) represent the slowness squared parametrization, the green circles (- ○ -) the slowness parametrization and the red crosses (- ✶ -) the velocity parametrization. This experiment uses the Marmousi velocity (of Figure 5.7(a)).

5.5.3 Acoustic Fréchet derivative, variable density

Incorporating a variable density provides many possibilities for the choice of parameters to be inverted, and we have selected the ones presented in Table 5.2, that are: the velocity, bulk modulus, density and impedance. For the experiments, we consider the acoustic Marmousi model with a variable density, given in Figure 5.32.

In Figure 5.33, we compare the Gauss-Newton Hessians associated with different parametrizations. The choice of comparing the inverse function of the bulk modulus and density with the inverse squared function of the velocity and impedance is motivated by the definition of the quantity and the square root appearing in Equation (5.6), e.g., $\kappa = c^2\rho$.



(a) Marmousi velocity.

(b) Marmousi density.

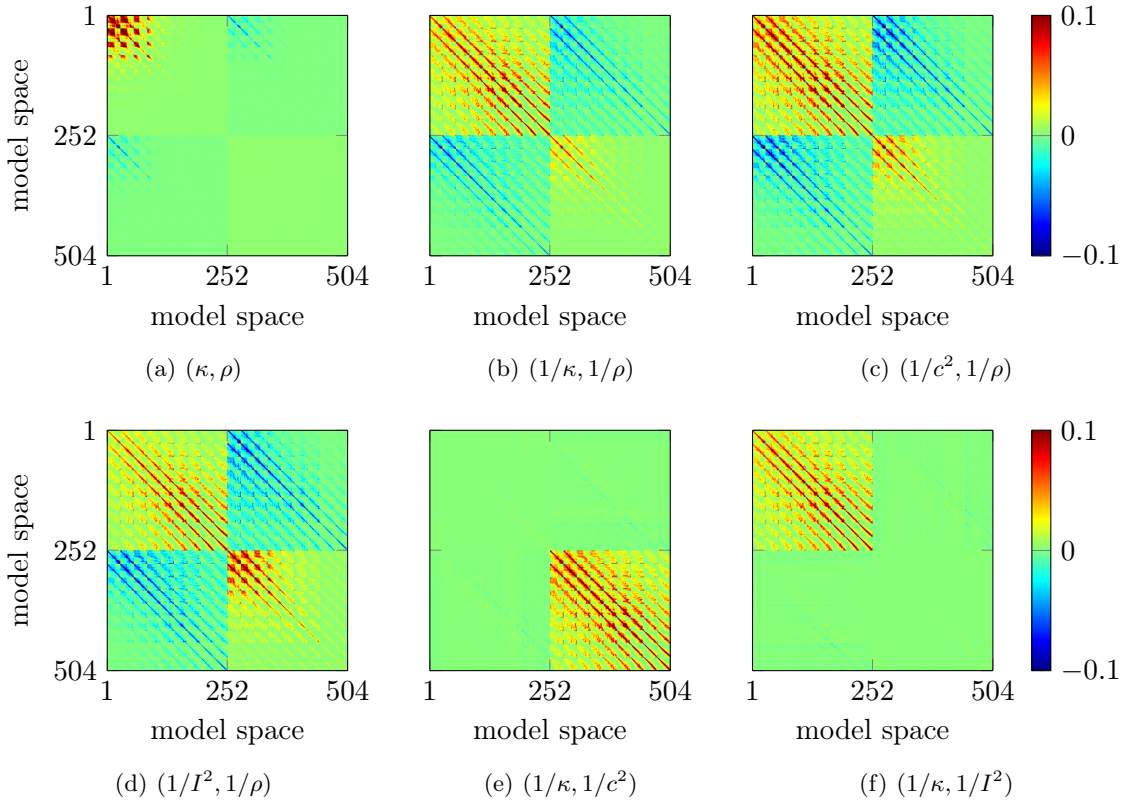
Figure 5.32: Representation of the acoustic Marmousi model with variable density, the medium is of size 9.2×3 km.

Remark 5.1. *In the context of multi-parameters inversion, the coefficients may have a very different magnitude. For example, the density is of around one thousand kilogram per meters cubed (or squared), while the bulk modulus is of several giga Pascal, i.e.* 10^9 *magnitude. In order to counter this effect, we usually introduce a scaling coefficient with the selected parameter to make the quantity dimensionless and of the same magnitude. Namely, instead of choosing* (κ, ρ) *we take* $(\kappa/\kappa_0, \rho/\rho_0)$ *such that* $\kappa_0 = \max(\kappa)$ *and* $\rho_0 = \max(\rho)$*. We proceed similarly for any other parametrization. For clarity we later omit this scaling parameter in the notation.*

The Gauss-Newton Hessians have distinctively four sub-matrices: the diagonal blocks correspond with the derivative of the selected parameters squared and the extra diagonal ones (which are identical as the matrix is symmetric) represent the multiplication of the different derivatives (see Equation (4.40)). Similarly to the acoustic case with constant density, we observe a clear benefit in using the inverse parameterization, which gives a much better insight regarding the matrix, see Figures 5.33(a) and 5.33(b). The use of the velocity, impedance or bulk modulus in addition to the density provides very similar pattern and scale, with a slight improvement in contrast for the impedance (see Figures 5.33(b), 5.33(c) and 5.33(d)). It also appears to be a wise choice to have the density separated from the other parameters, as the coupling of two parameters between velocity, impedance and bulk modulus gives a poor resolution, Figures 5.33(e) and 5.33(f).

Let us fix the second parameter to be $1/\rho$. In Figure 5.34, we compare the condition number of the Fréchet derivative for the other parameters, depending on the function applied, in particular, logarithm and square root. We also notice very few differences between the three remaining parameters (c, κ and I).

The comparison of the condition numbers confirms what is observed from the Gauss-Newton Hessian matrices, the inverse parametrization gives the best results for the bulk



(a) (κ, ρ) (b) $(1/\kappa, 1/\rho)$ (c) $(1/c^2, 1/\rho)$

(d) $(1/I^2, 1/\rho)$ (e) $(1/\kappa, 1/c^2)$ (f) $(1/\kappa, 1/I^2)$

Figure 5.33: Comparison of the Gauss-Newton Hessian, $D\mathcal{F}^* D\mathcal{F}$, for the acoustic Marmousi wave speed with variable density. The data cover a set of frequency from 1 to 10 Hz, using 1 Hz increment. The partitioning is selected to generate $N = 252$ coefficients, so that the size of the matrix is 504. The matrices are scaled with their respective maximum value to fit on the same scale.

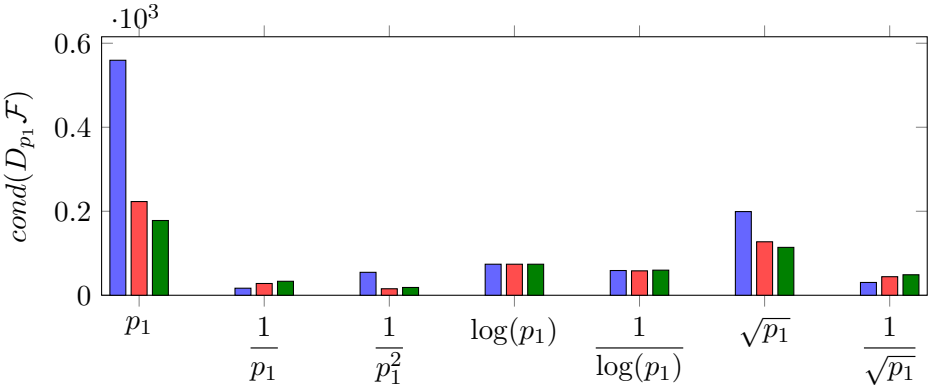
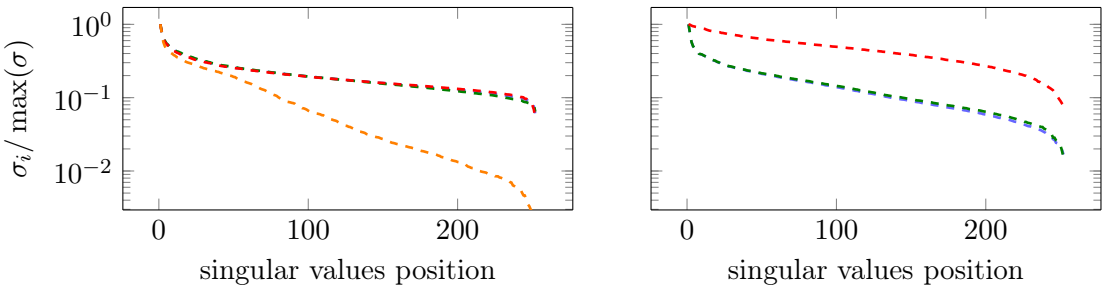


Figure 5.34: Comparison of the Fréchet derivative condition number for different parametrization associated with the Marmousi medium with variable density imposing model partition $N = 252$. The blue bars ( ) employ $p_1 = \kappa$, the red bars ( ) $p_1 = I$ and the green bars ( ) $p_1 = c$. The second parameter is fixed to be $p_2 = 1/\rho$.

modulus κ, and it is the inverse squared for the impedance and velocity, I and c respectively.

The results are confirmed in Figure 5.35 where we compare the singular values evolu-

tion for the different parametrizations. In Figure 5.35(a) we see that using $1/\kappa$, $1/I^2$ or $1/c^2$ in addition to the density provides very similar rates for the singular values. However, the use of κ creates a disadvantage as the singular values decrease much faster. In Figure 5.35(b), we compare the evolution for the singular values associated with the Fréchet derivative computed with respect to the density ($1/\rho$ actually). It is affected by the selected combination and we identify $1/c^2$ to be the best candidate, meanwhile $1/\kappa$ and $1/I^2$ give very similar results.



(a) Singular values of the Fréchet derivative with respect to the first parameter.

(b) Singular values of the Fréchet derivative with respect to $1/\rho$ using different choice for the first parameter.

Figure 5.35: Evolution of the Fréchet derivative singular values for partitioning $N = 252$. (a) The blue dashed line (- - -) represents the singular values for $1/\kappa$ using a parametrization $(1/\kappa, 1/\rho)$, the green dashed line (- - -) is the singular values for $1/I^2$ using parametrization $(1/I^2, 1/\rho)$, the red dashed line (- - -) is the singular values for $1/c^2$ using parametrization $(1/c^2, 1/\rho)$ and the orange dashed line (- - -) is the singular values for κ using parametrization (κ, ρ).
(b) The blue dashed line (- - -) represents the singular values for $1/\rho$ using a parametrization $(1/\kappa, 1/\rho)$, the green dashed line (- - -) is the singular values for $1/\rho$ using parametrization $(1/I^2, 1/\rho)$, the red dashed line (- - -) is the singular values for $1/\rho$ using parametrization $(1/c^2, 1/\rho)$.

There are many possibilities and combinations available for the selection of two parameters. We distinctively see that this choice must be made wisely as it influences the Fréchet derivative. The inverse function of the parameter improves the behavior, particularly as the singular values rate of decrease is much slower. Also in our estimates, the density is kept separated and the impedance, velocity or bulk modulus are selected in supplement. It appears that they are globally very close but the velocity, with $1/c^2$, seems the best candidate as it also influences positively the singular values of the Fréchet derivative associated with the density. We shall compare the parametrization in the reconstruction algorithm in the following experiments.

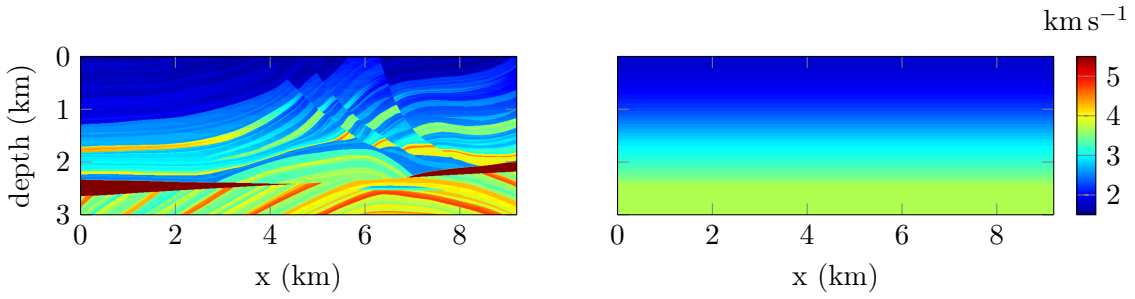
5.5.4 Acoustic FWI experiments

After we have analyzed the acoustic parametrizations in terms of effect on the condition number and singular values of their respective Fréchet derivative, we now illustrate their

influence during the iterative reconstruction procedure. We provide two acoustic synthetic experiments based on the Marmousi model, with constant and variable density.

5.5.4.1 Marmousi experiment, constant density

Regarding the acoustic situation with constant density, it seems that the slowness squared parametrization should behave much better than the other parametrizations we have investigated. We want to verify this observation by employing the FWI on a Marmousi test case, and perform the reconstruction for the three parametrizations of interest. We consider the acquisition made of 91 sources and 183 receivers for each. The true and starting models are presented in Figure 5.36, where we just have side to side the models already presented in Figures 5.25 and 5.4. Here we use synthetic data that are generated in the frequency domain with a discontinuous Galerkin discretization. The reconstruction is performed using a finite difference discretization, to avoid inverse crime.

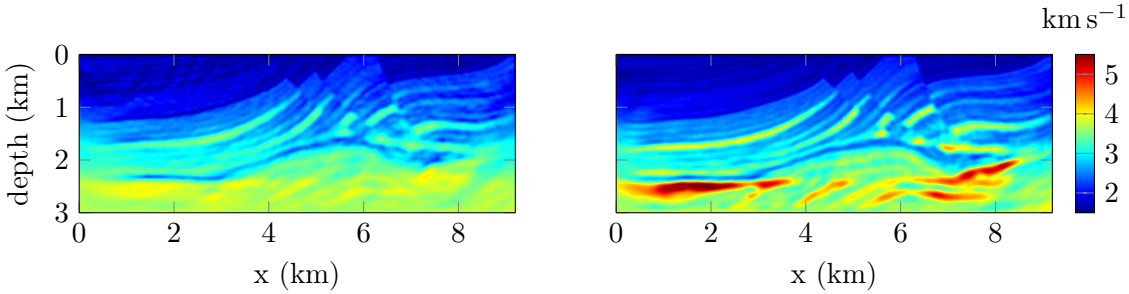


(a) Marmousi velocity.

(b) Starting model for the reconstruction of the Marmousi velocity.

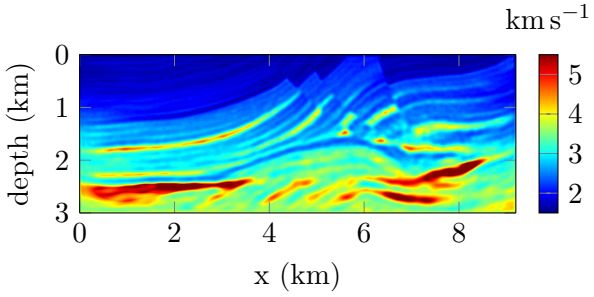
Figure 5.36: Experimental FWI setup for the Marmousi medium of size 9.2×3 km, the true and starting wave speed are given.

We perform an iterative minimization using frequencies from 2 to 10 Hz, with 1 Hz step. In Figure 5.37, we compare the results using exactly the same algorithm but different parametrization: c, $1/c$ and $1/c^2$. Clearly the velocity inversion provides limited accuracy for the deepest structures, while the others detect the contrasts. Also the amplitude of the velocity is much lower than expected for parametrization using c, meanwhile slowness and slowness squared parametrizations are accurate. Here the inversion of $1/c^2$ gives the overall best reconstruction. This result was predicted by the analysis of the Fréchet derivatives of Subsection 5.5.2.



(a) Velocity parametrization.

(b) Slowness parametrization.



(c) Slowness squared parametrization.

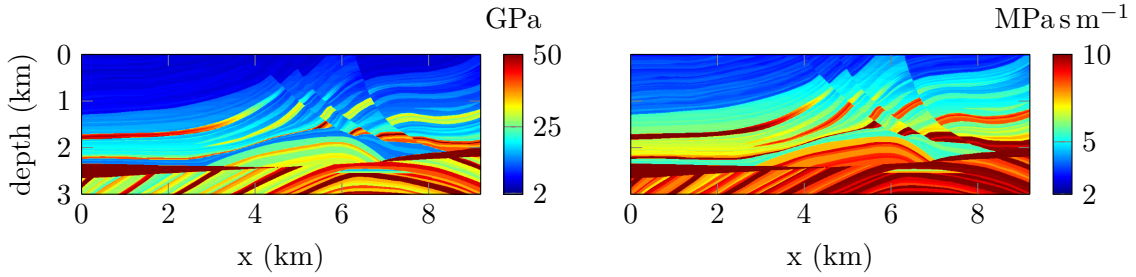
Figure 5.37: Final reconstruction after 10 Hz iterations of the Marmousi medium using different parametrization, the initial and true velocities are presented in Figure 5.36.

5.5.4.2 Marmousi experiment, variable density

Let us pursue the reconstruction depending on the parametrization for acoustic media by incorporating the density. The velocity and density models are presented in Figure 5.32. Let us first extract the other quantities, following the relation (5.6), and illustrate the resulting models in Figure 5.38 with the underlying bulk modulus and impedance. We obviously obtain the similar structures as seen for the density and velocity. For the reconstruction, we assume no knowledge of the structures and start with a one-dimensional variation, with depth only, illustrated in Figure 5.39.

For the reconstruction, we iterate from 2 to 10 Hz with 20 iterations per sequential frequencies and 1 Hz step. We first fix one of the two parameters to be $1/\rho$ and compare the recovery of the other parameter, depending of its choice (bulk modulus, velocity or impedance) and function applied. This comparison is presented Figure 5.40.

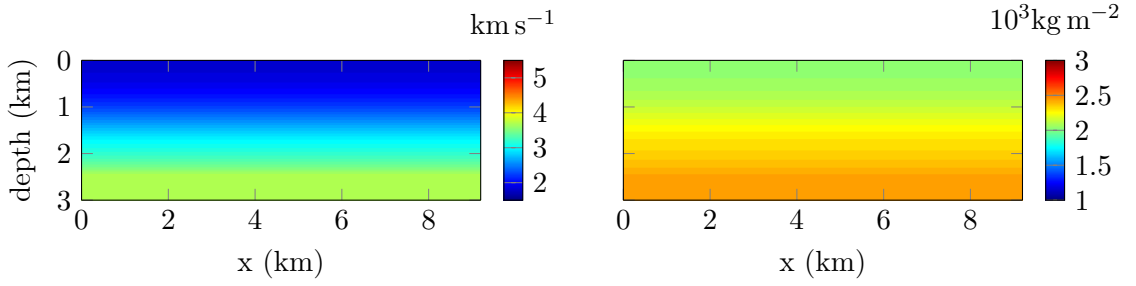
We see that the choice between the impedance, bulk modulus or velocity does not affect the recovery much. However, the choice of function to apply is revealed to be important, in particular the inverse of the parameters gives better recovery than when using the parameters themselves. The main structures are captured with accurate variations. It seems



(a) Marmousi bulk modulus.

(b) Marmousi impedance.

Figure 5.38: Quantities associated with the Marmousi model, extracted from the wave speed and density given Figure 5.32.



(a) initial velocity

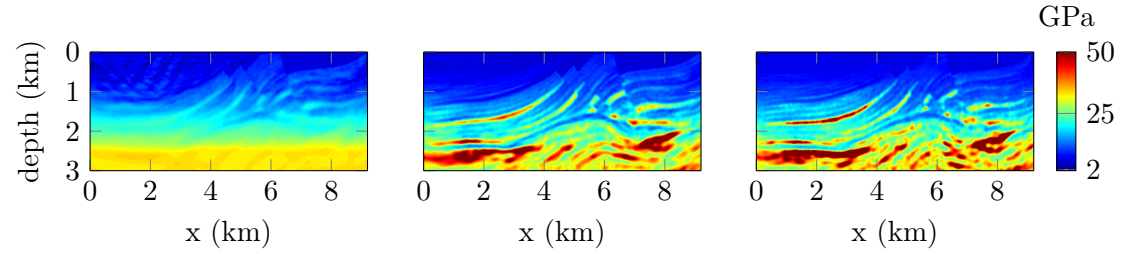
(b) initial density

Figure 5.39: Acoustic models used as initial guess for the recovery of the Marmousi medium.

that the velocity gives slightly less information on the deepest structures but the overall accuracy is very similar for the three parameters. In Figure 5.41, we compare the recovered density that is obtained in parallel.

The iterative minimization fails to reconstruct the density for all parametrizations. It is interesting to notice that it does not affect the recovery of the other, which is very accurate in comparison, see Figure 5.40. The couple density and velocity probably gives the better results, as anticipated by the singular value decomposition, but the overall resolution remains poor in comparison to Figure 5.40. Finally, we try to mix the parameters, thus not taking the density specifically. In Figure 5.42, we compare the reconstruction of the bulk modulus when associated to the impedance and wave speed, namely we compare the choice of parametrizations (κ^{-1}, I^{-1}) and (κ^{-1}, c^{-1}).

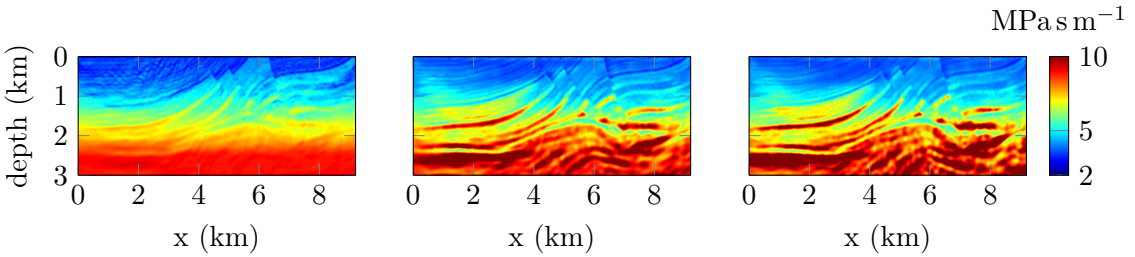
Clearly the reconstruction is the best when the density is associated with the bulk modulus. Taking the combination of the bulk modulus with the impedance or the velocity deteriorates the reconstruction. It confirms that the density should be kept alone, and it is probably the less correlated parameter.



(a) Reconstructed bulk modulus using (κ, ρ^{-1}).

(b) Reconstructed bulk modulus using $(\kappa^{-1/2}, \rho^{-1})$.

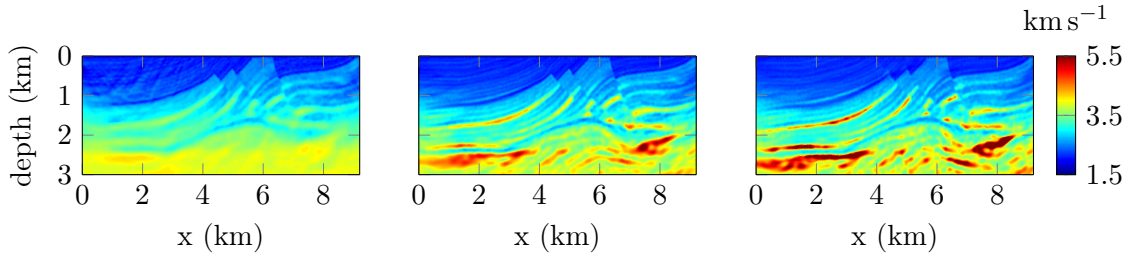
(c) Reconstructed bulk modulus using (κ^{-1}, ρ^{-1}).



(d) Reconstructed impedance using (I, ρ^{-1}).

(e) Reconstructed impedance using (I^{-1}, ρ^{-1}).

(f) Reconstructed impedance using (I^{-2}, ρ^{-1}).



(g) Reconstructed wave speed using (c, ρ^{-1}).

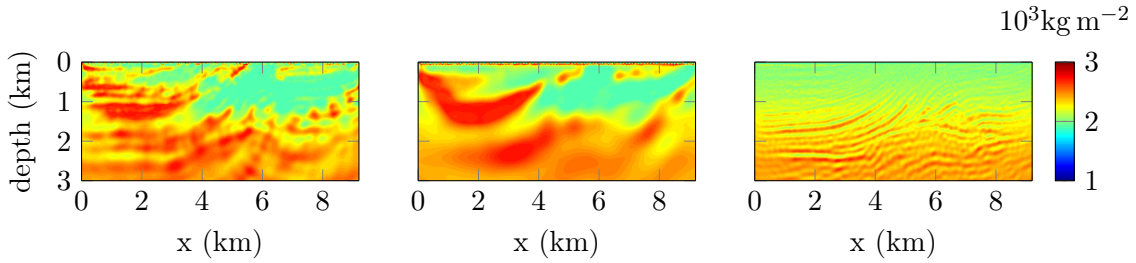
(h) Reconstructed wave speed using (c^{-1}, ρ^{-1}).

(i) Reconstructed wave speed using (c^{-2}, ρ^{-1}).

Figure 5.40: Final reconstruction after 10 Hz iterations of the Marmousi model using different parametrizations and functions. The recovered parameter is the bulk modulus for (a), (b) and (c); the impedance in (d), (e) and (f) and the wave speed in (g), (h) and (i). The second inverted parameter is fixed to be $1/\rho$.

Following the complexity of retrieving the density (Virieux and Operto, 2009), we decide to perform the iterative minimization on one single parameter, and the density is kept as its initial guess illustrated Figure 5.39(b), all along the iterative process. In Figure 5.43, we compare the recovery of the parameter.

As expected, the single parameter inversion keeping the initial density gives accurate reconstruction of the parameters. It is important to note that not knowing the density or failing in approximating it does not appear to prevent the recovery of the other parameters. After the reconstruction of one selected parameter between impedance, bulk modulus and velocity, one can then try to focus on the density only, as proposed in Jeong et al. (2012); or even take a first approximation from the Gardner's relation (Gardner et al., 1974).

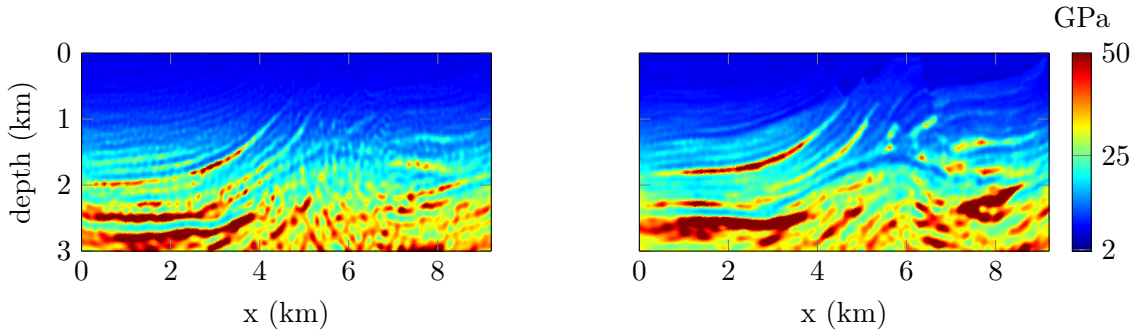


(a) Reconstructed density using (κ^{-1}, ρ^{-1}).

(b) Reconstructed density using (I^{-2}, ρ^{-1}).

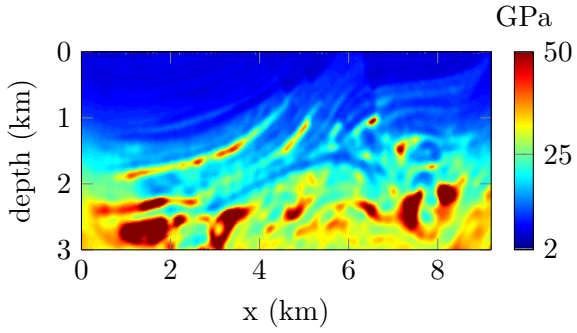
(c) Reconstructed density using (cp^{-2}, ρ^{-1}).

Figure 5.41: Final reconstruction after 10 Hz iterations of the Marmousi density from different parametrizations.



(a) Reconstructed bulk modulus using (κ^{-1}, I^{-1}).

(b) Reconstructed bulk modulus using (κ^{-1}, ρ^{-1}).

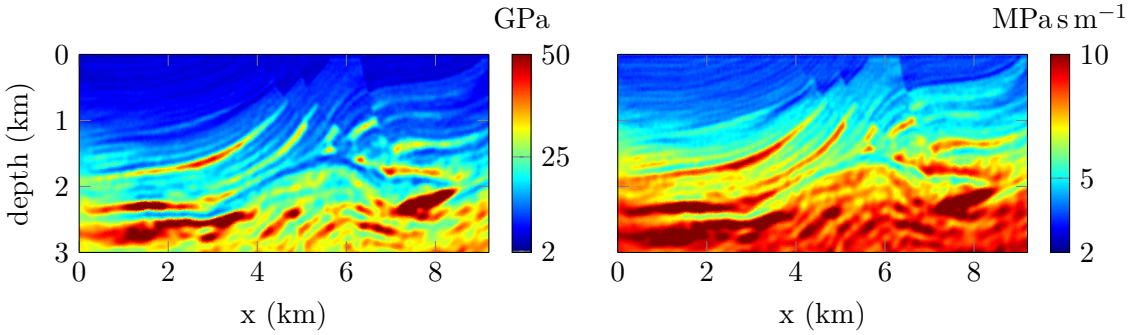


(c) Reconstructed bulk modulus using (κ^{-1}, c^{-1}).

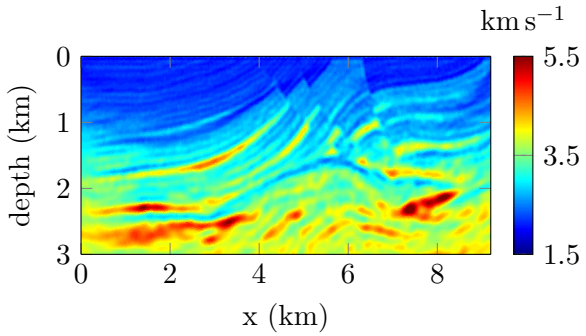
Figure 5.42: Final reconstruction after 10 Hz iterations of the Marmousi bulk modulus using different parametrizations.

5.6 Elastic parametrization

We can proceed similarly in the context of elastic medium where the model is now characterized by three parameters, related to the density (ρ) and the Lamé parameters (λ and μ).



(a) reconstructed bulk modulus using ($\kappa^{-1/2}$) inversion (b) reconstructed impedance using (I^{-1}) inversion



(c) reconstructed wave speed using (c^{-1}) inversion

Figure 5.43: Final reconstruction after 10 Hz iterations of the Marmousi model using different parametrization and keeping the density fixed as its initial value pictured Figure 5.39(b).

5.6.1 Elastic models

Let us first review some of the standard physical quantities that can be introduced depending on the density and the Lamé parameters:

$$\nu(\mathbf{x}) = \frac{\lambda(\mathbf{x})}{2(\lambda(\mathbf{x}) + \mu(\mathbf{x}))} \qquad \text{Poisson ratio,}$$

$$\kappa(\mathbf{x}) = \lambda(\mathbf{x}) + \frac{2}{3}\mu(\mathbf{x}) \qquad \text{bulk modulus,}$$

$$c_p(\mathbf{x}) = \sqrt{\frac{\lambda(\mathbf{x}) + 2\mu(\mathbf{x})}{\rho(\mathbf{x})}} \qquad \text{P-wave speed,}$$

$$c_s(\mathbf{x}) = \sqrt{\frac{\mu(\mathbf{x})}{\rho(\mathbf{x})}} \qquad \text{S-wave speed,}$$

$$I_p(\mathbf{x}) = \rho(\mathbf{x})c_p(\mathbf{x}) = \sqrt{\rho(\mathbf{x})\big(\lambda(\mathbf{x}) + 2\mu(\mathbf{x})\big)} \qquad \text{P-impedance,}$$

$$I_s(\mathbf{x}) = \rho(\mathbf{x})c_s(\mathbf{x}) = \sqrt{\rho(\mathbf{x})\mu(\mathbf{x})} \qquad \text{S-impedance.}$$

In the case of an isotropic elastic medium, the propagation of wave follows Equation (1.29) and we take the basis subsurface parametrization to be $m = \{\lambda, \mu, \rho\}$, in accordance with

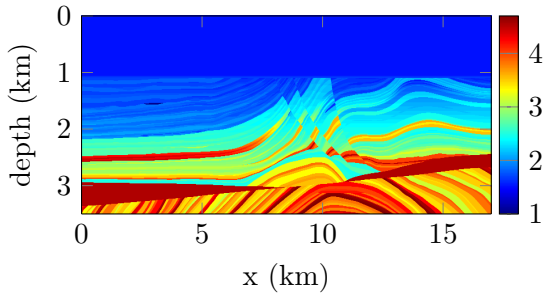
the wave equation. Following what has been observed in the acoustic situation, we leave aside the density ρ in the parametrization and limit our study to the four following combinations:

- Lamé parametrization (λ, μ, ρ),
- impedance parametrization (I_p, I_s, ρ),
- velocity parametrization (c_p, c_s, ρ),
- and the parametrization (ν, κ, ρ).

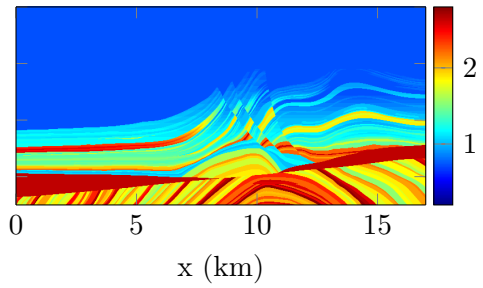
As before, we use coefficients to scale the quantities and guarantee similar magnitude, see Remark 5.1.

5.6.2 Elastic Fréchet derivative parametrization

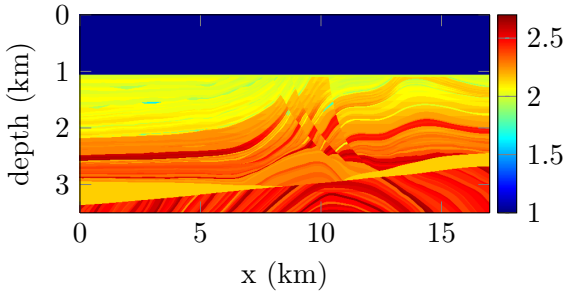
For the analysis of parametrizations, we consider the elastic Marmousi2 model, which we picture in Figure 5.44 with the P-wave speed, S-wave speed and density.



(a) P-velocity ($\mathrm{km\,s^{-1}}$).



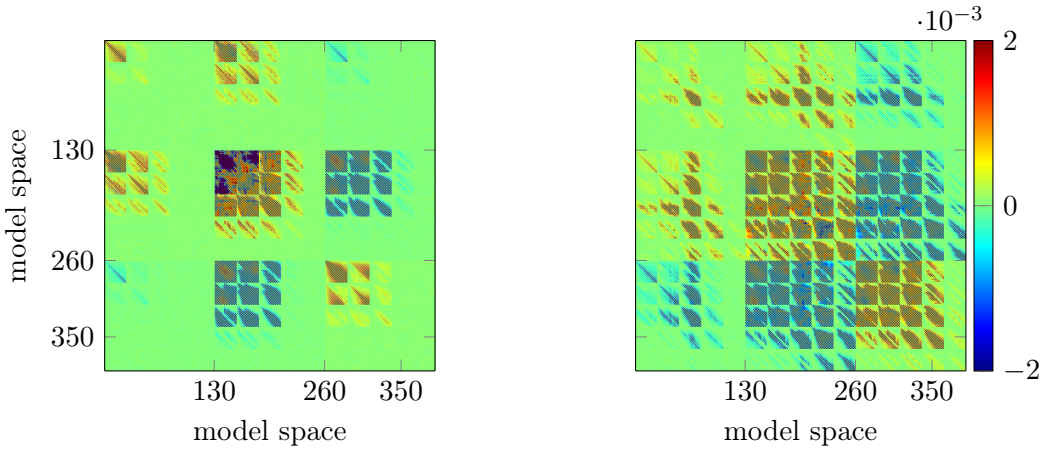
(b) S-velocity ($\mathrm{km\,s^{-1}}$).



(c) Density (10^3 $\mathrm{kg\,m^{-2}}$).

Figure 5.44: Elastic isotropic Marmousi2 medium of size 17×3.5 km, it is represented with the P- and S-wave speeds and the density.

The acquisition is composed of 19 sources and 168 receivers record the displacement information for every source, for a total of $168 \times 19 = 3192$ measurements per frequency. The frequency bandwidth covers from 1 to 10 Hz using 1 Hz step. Similarly to the acoustic case, we apply model partition to reduce the number of coefficients in the representation and allow the computation of the Fréchet derivative. In Figure 5.45, we compare the resulting Gauss-Newton Hessian from two parametrizations inherited from the Lamé parameters. In Figure 5.46 are compared the condition numbers of the Fréchet derivatives depending on the parametrization and the function applied on the parameters.



(a) Parametrization (λ, μ, ρ).

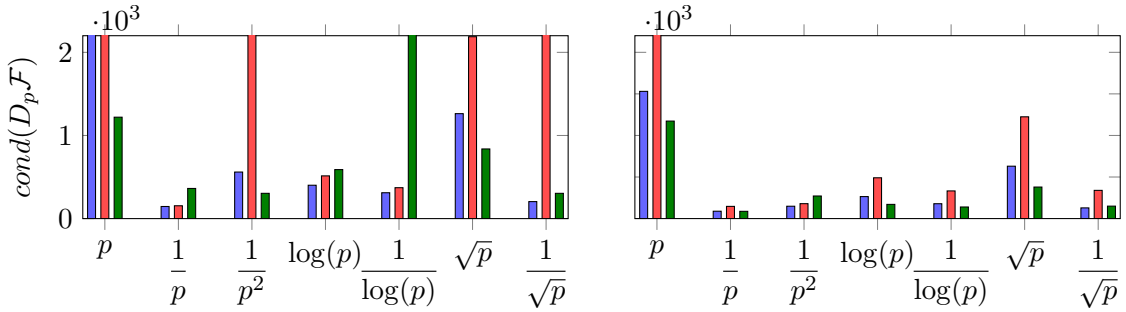
(b) Parametrization $(\lambda^{-1}, \mu^{-1}, \rho^{-1})$.

Figure 5.45: Comparison of the Gauss-Newton Hessian, $D\mathcal{F}^* D\mathcal{F}$, for the elastic Marmousi2 wave speed. The data cover a set of frequency from 1 to 10 Hz, using 1 Hz increment. The partitioning is selected to generate $N = 130$ coefficients per parameter, so that the size of the matrix is 3×130. The matrices are scaled with their respective maximum value to stand on the same scale.

We see a large variation of the condition number depending on the situation. Using the inverse function provides the best result for all parametrization. Note that despite the square root relation between some of the quantities (e.g., between the Lamé parameters and the velocities as, for instance, $c_s^2 \rho = \mu$) and contrary to the acoustic situation, the inverse squared parametrization is not the best choice and we should restrict to use of the inverse one.

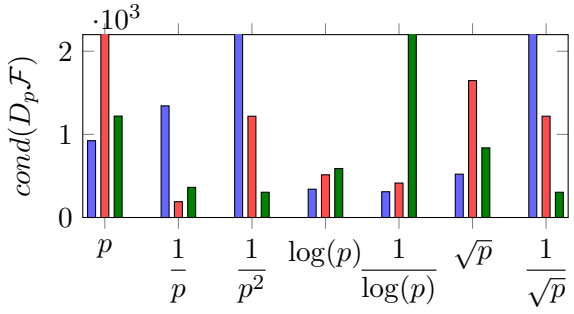
5.6.3 FWI experiments

We now compare the parametrization selection during the FWI reconstruction. We remain with the Marmousi2 elastic media presented Figure 5.44. The acquisition is now composed of 141 sources and 281 receivers per source. The starting model for the reconstruction does not assume any knowledge of the subsurface structures, they are pictured in Figure 5.14. We use a sequential frequency progression from 1 to 10 Hz. In this experiment, following

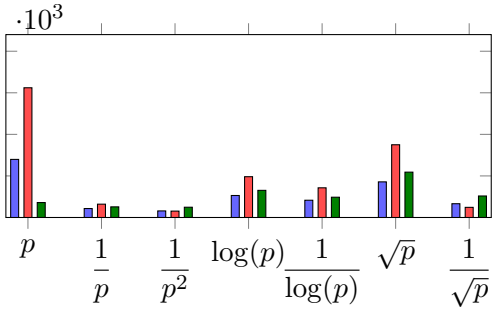


(a) Parametrization (λ, μ, ρ).

(b) Parametrization (ν, κ, ρ).



(c) Parametrization (I_p, I_s, ρ).



(d) Parametrization (c_p, c_s, ρ).

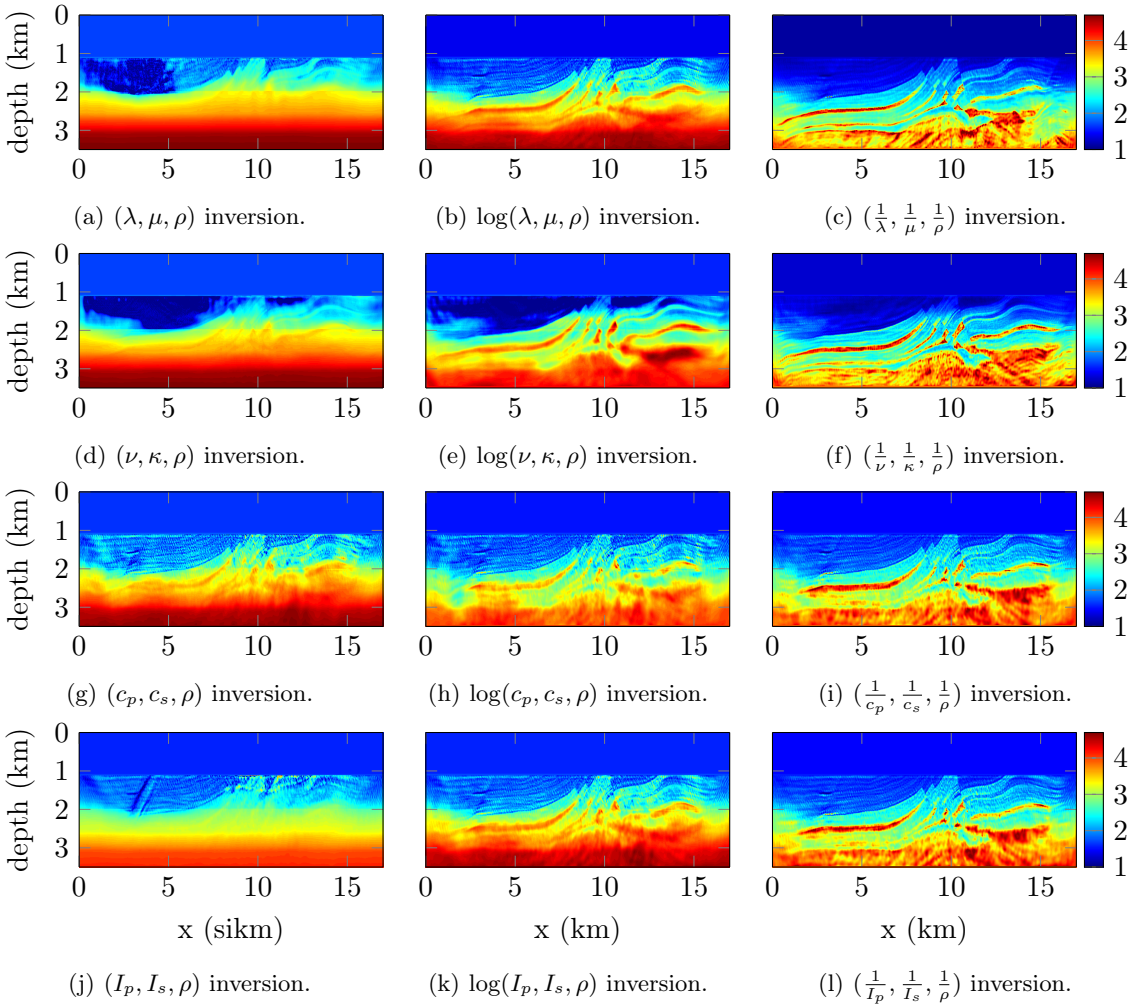
Figure 5.46: Comparison of the Fréchet derivative condition number for different parametrizations associated with the elastic Marmousi2 medium imposing the model partition $N = 530$ per parameter. The blue bars ( ) represent the condition number of the Fréchet derivative of the first parameter of interest, the red bars of the second parameter and the green bars ( ) of the last parameters. The parameters are indicated in the parenthesis.

the results of the acoustic situation, we disregard the density reconstruction and only focus on the remaining two parameters (which depend on the selected parametrization). Namely, the density is kept in its initial and false representation of Figure 5.14(c).

5.6.3.1 Ignoring surface waves

In Figure 5.47, the final reconstructions for the P-velocity are compared depending on the parametrization. In this experiment, we do not consider surface waves by imposing an absorbing condition for the upper surface. This simplification noticeably improves the reconstruction as we will see below.

The elastic reconstruction follows the analysis provided by the Fréchet derivative condition number: the inverse parametrization gives clearly the best recovery. The logarithmic function generates a smooth representation but the standard one (when no function is applied) lacks major information on the deepest structures. Concerning the elastic coefficients,



(a) (λ, μ, ρ) inversion. (b) $\log(\lambda, \mu, \rho)$ inversion. (c) $(\frac{1}{\lambda}, \frac{1}{\mu}, \frac{1}{\rho})$ inversion.

(d) (ν, κ, ρ) inversion. (e) $\log(\nu, \kappa, \rho)$ inversion. (f) $(\frac{1}{\nu}, \frac{1}{\kappa}, \frac{1}{\rho})$ inversion.

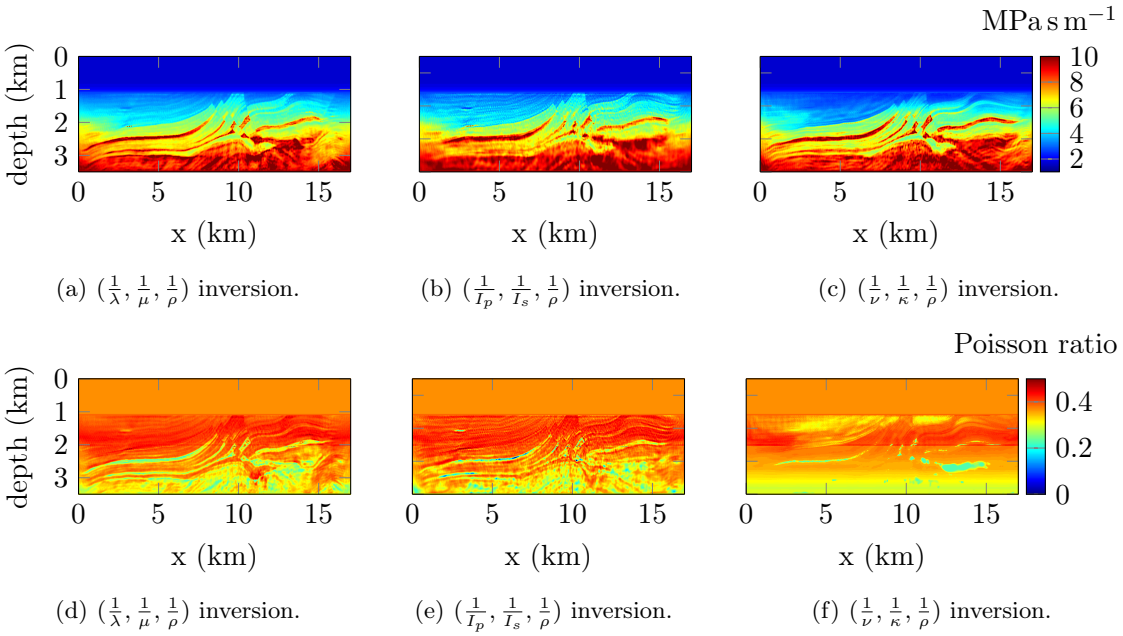
(g) (c_p, c_s, ρ) inversion. (h) $\log(c_p, c_s, \rho)$ inversion. (i) $(\frac{1}{c_p}, \frac{1}{c_s}, \frac{1}{\rho})$ inversion.

(j) (I_p, I_s, ρ) inversion. (k) $\log(I_p, I_s, \rho)$ inversion. (l) $(\frac{1}{I_p}, \frac{1}{I_s}, \frac{1}{\rho})$ inversion.

Figure 5.47: Final reconstruction after 10 Hz iterations of the Marmousi2 P-velocity model using different parametrization and functions. The density is kept fixed with the iterations, to its initial representation. In this experiment, surface waves are not considered.

the choice of the Lamé parameters gives more accurate resolution for the deepest contrasts but this intuition should be confirmed with non synthetic and more complicated examples. However, because the density is kept unknown, it seems natural that the impedance and velocity parametrizations are more affected while the other parametrization should suffer less. We see that the inaccuracy of the density does not prevent us from recovering the models. It is not really fair to compare the recovery of the wave speed while some parametrizations do not aim at recovering it directly. In Figure 5.48, we compare alternative quantities: the P-impedance and Poisson ratio.

The different quantities do not provide the same accuracy in the subsurface approximation. It is interesting to notice that the choice of parametrization may lead to accuracy improvement on quantities that were not considered. For instance the inversion using



(a) $(\frac{1}{\lambda}, \frac{1}{\mu}, \frac{1}{\rho})$ inversion. (b) $(\frac{1}{I_p}, \frac{1}{I_s}, \frac{1}{\rho})$ inversion. (c) $(\frac{1}{\nu}, \frac{1}{\kappa}, \frac{1}{\rho})$ inversion.

(d) $(\frac{1}{\lambda}, \frac{1}{\mu}, \frac{1}{\rho})$ inversion. (e) $(\frac{1}{I_p}, \frac{1}{I_s}, \frac{1}{\rho})$ inversion. (f) $(\frac{1}{\nu}, \frac{1}{\kappa}, \frac{1}{\rho})$ inversion.

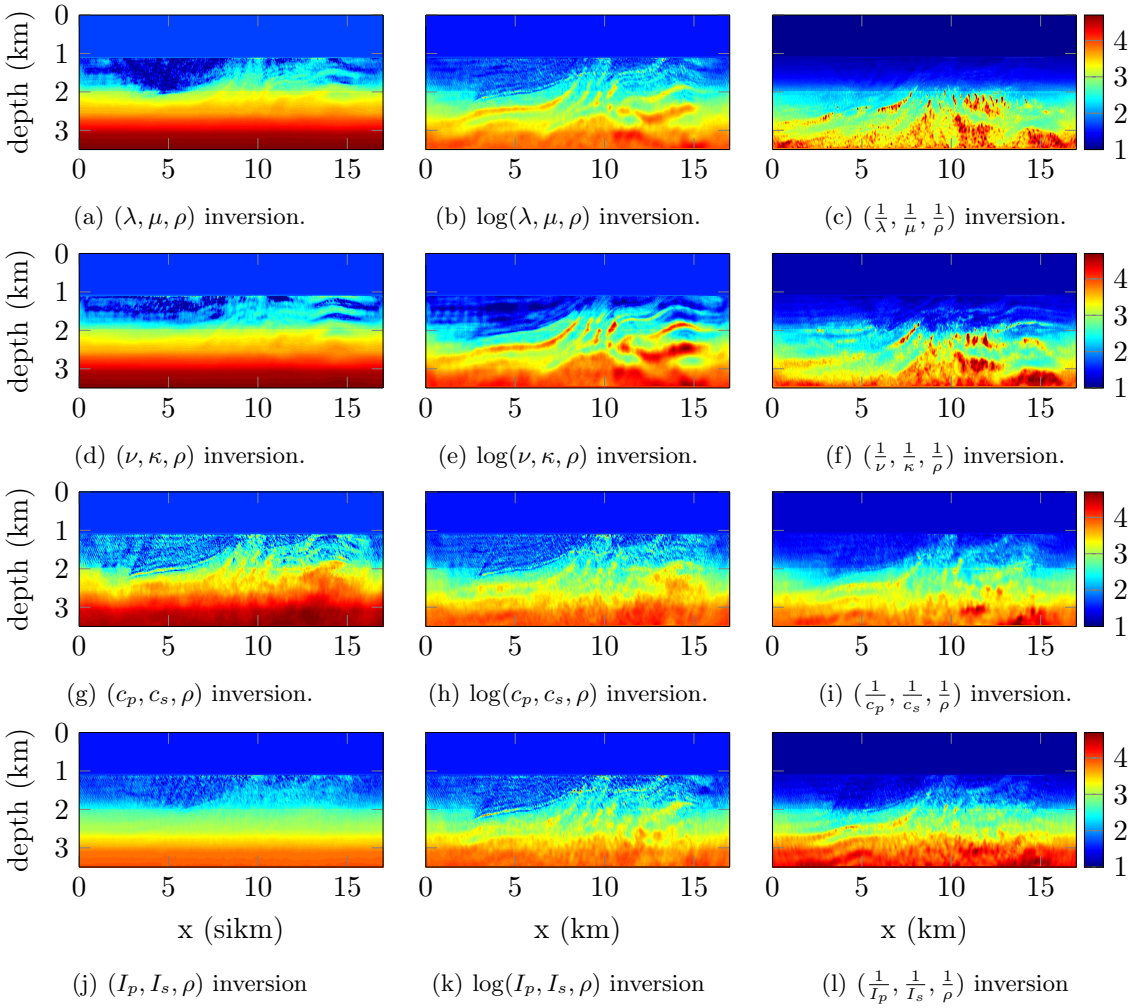
Figure 5.48: Final reconstruction after 10 Hz iterations of the Marmousi P-impedance (top) and Poisson ratio (bottom) using different parametrizations. The density is kept fixed with the iterations, to its initial representation.

the Lamé parameters (Figures 5.48(a) and 5.48(d)) gives insight into the impedance and Poisson ratio with more accuracy than inverting directly those quantities, (Figures 5.48(b) and 5.48(f) respectively). In this experiment, we can observe that the Lamé parameters inversion (with the inverse function applied: $1/\lambda$ and $1/\mu$) gives the best reconstruction, providing more information on the deepest structures, and the appropriate contrast.

5.6.3.2 Free Surface boundary condition

We reproduce this elastic experiment incorporating the surface waves, that is, imposing a free surface boundary condition on the top surface. It is known to increase the complexity of the recovery (due to the introduction of multiple surface reflections), as shown in Gélis et al. (2007). Indeed, we observe a reduction in the accuracy compared to the original example, see Figures 5.49.

In this case, the logarithm function appears to provide the best reconstructions. It gives a smooth velocity with appropriate accuracy for the upper part of the models, yet the deepest parts are missing. The inverse functions fail to recover the structures, opposite to the case where surface waves are deleted. The combination of parameters has also a clear impact on the reconstructions, while it was relatively insensitive previously (the use



(a) (λ, μ, ρ) inversion. (b) $\log(\lambda, \mu, \rho)$ inversion. (c) $(\frac{1}{\lambda}, \frac{1}{\mu}, \frac{1}{\rho})$ inversion.

(d) (ν, κ, ρ) inversion. (e) $\log(\nu, \kappa, \rho)$ inversion. (f) $(\frac{1}{\nu}, \frac{1}{\kappa}, \frac{1}{\rho})$ inversion.

(g) (c_p, c_s, ρ) inversion. (h) $\log(c_p, c_s, \rho)$ inversion. (i) $(\frac{1}{c_p}, \frac{1}{c_s}, \frac{1}{\rho})$ inversion.

(j) (I_p, I_s, ρ) inversion (k) $\log(I_p, I_s, \rho)$ inversion (l) $(\frac{1}{I_p}, \frac{1}{I_s}, \frac{1}{\rho})$ inversion

Figure 5.49: Final reconstructions after 10 Hz iterations of the Marmousi2 P-velocity model using different parametrizations and functions. The density is kept fixed along the iterations as its initial representation. Free surface boundary condition is incorporated for the upper boundary.

of inverse function had the major effect). The choice of Lamé parameters and the couple made of the Poisson ratio and bulk modulus give a much better recovery than the use of impedance or velocity, which fails. However, this is a simple experimental comparison and it should be confirmed by a thorougher study on various subsurface geometries.

Remark 5.2. *The difficulties of the reconstruction provoked by the surface waves have also been noted by Gélis et al. (2007); Brossier (2009); Yuan et al. (2015). In our experiment, the iterative reconstruction is conducted using a gradient descent algorithm and we can expect at least some improvement by employing a more advanced iterative scheme, possibly involving Hessian information for example. Another alternative can be a preprocessing of the data to suppress the surface waves. The sensitivity of the inverse scheme to the surface waves is also related to the subsurface geometry. We will illustrate in Chapter 6, Section 6.10, that subsurface parameters including salt domes suffer less from the surface waves.*

5.7 Conclusion

In this chapter, we have characterized the framework of FC/LD optimization problem. Such problems provide important guarantees regarding the convergence of the iterative scheme towards the proper minimizer (in particular by assuring the absence of local minimum). We have studied the two quantities involved: the deflection and the curvature. From the deflection, we extract an estimation of the model space size for which the LD property is satisfied. We wish to have the largest size possible. The FC property indicates the maximal distance allowed between the data and the attainable set. The successive estimates with frequency have been proven to be very useful to give a quantitative understanding of the behavior of the problem. Here, we have try to supplement the general procedure with some additional quantification regarding the frequency progression, the geometry, etc.

1. The frequency progresses from low to high values, and we have provided quantitative understanding on the basin of attraction size behavior for this well-known progression.
2. Using multi frequencies data is not useful to increase the basin of attraction, meaning that initial iterations, when no prior is known, should use the lowest frequency sequentially.
3. However, the maximal distance between the data and the attainable set is affected positively by the use of frequency bandwidth.
4. The use of complex frequencies, especially when only the damping is considered, provides an increase in the size of the basin of attraction.
5. The progression using complex frequencies depends on the damping factor. Namely, when only the damping is taken, it should vary from low to high values. When both imaginary and real part are non-zero, the damping should go from high to low at fixed Fourier frequency. This is precisely what has been observed in Shin and Cha (2008); Shin et al. (2010); Petrov and Newman (2014), here we have given a quantitative behavior of this progression.
6. The geometry has an impact on the convergence, having one large reflector provokes a reduction of the basin of attraction, meaning that it requires more prior information (or alternatively lower or complex frequencies).

It would be interesting to pursue the study to find the progression in the selection of frequency based on the basin of attraction size. This could be compared to what is proposed by Sirgue and Pratt (2004).

Then, we have analyzed the effect of the parametrization on the problem, by studying the Fréchet derivative of the associated forward operator. Both condition number and

singular values of the derivative give information on how the parametrization affects the problem. Two aspects have been analyzed, first the choice of the quantities (density, bulk modulus, Poisson ratio, etc) and then the application of a function onto those coefficients. It is not easy to comprehend how the reconstruction can be positively or negatively affected but we have shown some fundamental differences. Many combinations are possible, and we have tried to review the major ones. In conclusion, applying the inverse function appears the best choice in the majority of cases. It also seems valuable to separate the density from the other parameters. We have demonstrated how a wrong choice of parametrization can easily lead to the failure of the method.

Chapter 6

Seismic reconstruction experiments using FWI method

Abstract

We perform seismic reconstruction of subsurface coefficients using Full Waveform Inversion (FWI), where observed data are only located at or near the surface. Motivated by the stability and convergence analysis, we design a multi-level algorithm where frequency progresses simultaneously with scale (model partition) in order to make full use of the analytical understanding of the inverse problem. We employ the computational toolbox we have developed during this project to process the underlying iterative minimization algorithm and illustrate the possibilities and efficiency in different configurations. The software we have created can handle marine and land acquisition, and we accordingly review geophysical examples for acoustic and elastic media, in two and three dimensions. We also illustrate the way towards anisotropy and attenuation. In addition to the type of medium, the specific characteristics of the subsurface are investigated (structures, layers, salt domes, etc). One of the key points in these experiments is to start from models with no prior information on the subsurface structures.

Contents

In this chapter, we investigate the efficiency of the FWI method for subsurface reconstruction in various situations. The inverse problem is solved using an iterative minimization algorithm, based on the techniques introduced in Chapter 4. The detail of the computational toolbox developed during this project, and the options implemented, are given in the Appendix B. Here we consider acoustic and elastic media and examine the reconstruction from observations located at or near the surface, for two and three-dimensional domains. Specificities of the geophysical inverse problem include the available data, only located at (or near) the surface, large scale area of interest (several kilometers cubed), and limited accuracy of the observed measurements. Following our choice to work in the frequency domain, we require the Fourier transform of the original time domain observations.

In Section 6.1, we present the multi-level multi-frequency algorithm that is deployed for the benchmark. The design of this algorithm is motivated by the stability analysis of the inverse problem of Chapter 3, and the convergence (more precisely the evolution of the size of the basin of attraction) realized in Chapter 5. It induces a mutual progression in frequency (from low to high) and model partition (from low scale to refined scale).

Several possibilities are available to conduct the optimization scheme, and some have been already reviewed. For example in Section 4.7 we have analyzed the effect of Hessian information. Due to the numerical cost involved, here we solely employ nonlinear conjugate gradient methods, without accounting for the Hessian. Similarly, the parametrization has been studied in Chapter 5, where we reviewed some of the possible combinations for acoustic and elastic inverse problems. Following this analysis, the parameter inverted are the following, when no mention is specified:

- acoustic with constant density case: the wave propagation is given by Equation (1.27) and we invert the slowness squared $(1/c^2)$,
- acoustic with variable density case: the wave propagation is given by Equation (1.28) and we invert the inverse of the bulk modulus and density $(1/\kappa, 1/\rho)$,
- elastic case: the wave propagation is given by Equation (1.29) and we invert the inverse of the Lamé parameters and density $(1/\lambda, 1/\mu, 1/\rho)$.

Regarding the cost function, different norms can be employed, and we focus here both on the traditional least squares and the logarithm one. The latest is in particular crucial in the case of complex frequency. Hence when no mention is specified we employ the logarithmic norm with complex frequencies and the least squares for real frequencies.

In our experiments, we also insist on using initial models without any prior information. None of the subsurface structures are assumed at the beginning to push the algorithm as far

as possible. Therefore, all starting models are represented with one-dimensional profiles, which only varies with the depth. Those points of departure are also motivated because we mainly work with synthetic data with noise. The feasibility of the reconstruction is then tightly intertwined with the (complex) frequencies available, in particular, the lowest ones. We also illustrate how the noise affects the data and can make the algorithm fails, in particular by making the lowest frequencies unavailable.

The geometry of the subsurface structures impacts on the efficiency of the reconstruction. In particular, we illustrate how high contrasted objects (salt domes) are more complicated to recover and rely massively on lower or complex frequencies. It confirms the observation made in Chapter 5, where the size of the basin of attraction is shown to shorten for high contrasted objects. In Section 6.7, we further demonstrate, with some test cases, how it is related to the scale of the model.

The detailed computational toolbox developed, and the features implemented, are revealed in the Appendix B. In the various experiments we propose here, many aspects are faced. We mention

- the effect of the noise,
- the effect of model partition (compression),
- the effect of the subsurface geometry and underlying difficulties to recover salt domes,
- the difficulties for density reconstruction,
- surface waves perturbing the elastic reconstruction.

In Table 6.1, we list the experiments that are presented in this chapter, with the detail of the size and quantity of data (number of receivers and sources). Despite the number of tests, we will try to learn from the specificities of each of them. The code has been written in parallel to be able to perform on clusters. Here the majority of the computations has been processed on the Conte clusters deployed at Purdue University[3]. The experiments have all been running on several cores to reduce the computational burden. The more demanding experiment, which is the three-dimensional elastic Epati model (see Section 6.11) has been running on 1024 cores, probing scalability performances. In Table 6.2, we further detail the lowest frequency employed, the total number of iterations and the computational time for the different experiments. Except for the three-dimensional elastic case, we use a relatively limited amount of parallel resources. Namely, we only use one node (which is composed of sixteen cores). The computational time could be reduced by increasing the number nodes.

[3] https://www.rcac.purdue.edu/compute/conte/, specifications: 16 cores per node with two 8-Core Intel Xeon-E5; 64 GB per node; Infiniband: 40 Gbps FDR10.

Model name	**Type**	**Size** (km)	n_{src}	n_{rcv}	$\sim$ **Ndof**
1. Marmousi	acoustic 2D	9.2×3	91	183	4×10^5
2. Overthrust	acoustic 2D	20×4.65	199	399	8×10^5
3. Sigsbee2A	acoustic 2D	24.4×9.1	135	232	1×10^6
4. Seam	acoustic 2D	35×15	174	349	1.1×10^6
5. Louro	acoustic 3D	$2.6 \times 1.6 \times 0.6$	96	1000	3×10^5
6. Statoil[4]	acoustic 3D	$2.6 \times 1.5 \times 1.2$	160	1376	7×10^5
7. Marmousi2	elastic 2D	17×3.5	141	281	1×10^6
8. Pluto	elastic 2D	31.16×7.08	150	310	1.2×10^6
9. Epati	elastic 3D	$1.8 \times 1.4 \times 1$	176	936	1×10^7
10. Toy model	TTI 2D	2.46×0.6	150	304	3×10^5

Table 6.1: List of media used for the FWI experiments. We consider the same number of receivers for all sources so that the number of data at fixed frequency is given by $n_{src} \times n_{rcv}$. Ndof indicates the approximate number of degrees of freedom for the numerical discretization (we try to keep a relative similar amount between finite differences and Galerkin methods when both are compared).

Model name	**Type**	**Starting frequency**	n_{iter}	n_{cpu}	**Total time**
1. Marmousi	acoustic 2D	(2Hz,0)	180	4	1h45min
2.a Overthrust	acoustic 2D	(1Hz,0)	200	4	4h
2.b Overthrust	acoustic 2D	(1Hz,10)	400	4	8h
3. Sigsbee2A	acoustic 2D	(0Hz,1)	420	4	9h30min
4. Seam	acoustic 2D	(0Hz,1)	420	4	12h
5. Louro	acoustic 3D	(2Hz,0)	180	16	2h
6. Statoil	acoustic 3D	see Chapter 7			
7. Marmousi2	elastic 2D	(1Hz,0)	200	16	2h
8. Pluto	elastic 2D	(0Hz,1)	440	4	14h
9. Epati	elastic 3D	(2Hz,0)	195	1024	100h
10. Toy model	TTI 2D	(1Hz,0)	200	8	4h

Table 6.2: Computational details regarding the FWI experiments. We use complex frequency such that $-\omega^2 = (\sigma + 2\mathrm{i}\pi f)^2$, the initial frequency is given as (f,σ). n_{iter} is the total number of iterations for the minimization, n_{cpu} indicates the number of cores used for the computation. Except for the 3D elastic case, few computational resources are used, meaning that the computational time can certainly be improved.

6.1 Multi-level, multi-frequency algorithm

We design a multi-level, multi-frequency algorithm based on the stability and convergence analysis provided in Chapters 3 and 5. In Chapter 3, we have established that the stability of the inverse problem improves (i.e., the stability constant decreases) when the number of

[4]The 3D acoustic Statoil model is actually used in Chapter 7 to illustrate the performance working with Cauchy data.

coefficients to represent the unknown decreases. However, the stability appears to worsen when the frequency decreases. Numerically, we can make use of these results with the following:

- at low frequency, one should use a coarse representation for the parameters to prevent the stability to diminish because of the frequency,
- at high frequency, one can use a refined partition to improve the resolution, benefiting from the improved stability given by higher frequencies.

Stability-wise, we could be tempted to remain in a high frequency regime in order to allow an improved resolution while maintaining a relatively low stability constant. However, from the evaluation of the size of the basin of attraction (cf Chapter 5), we have seen that low frequencies are required to expect any convergence when minimal prior information is assumed. That is why we have designed a natural mutli-level multi-frequency scheme where frequency and scale progress simultaneously. Denoting N the number of coefficients in the representation (low N implies coarse representation), the iterations are conducted pairing frequency and scale such that:

1. we start with the couple (ω_0, N_0) where ω_0 is the lowest frequency available (to increase the size of the basin of attraction) and where N_0 is relatively low (the coarse representation is required to reduce the stability constant).
2. After the iterative minimization has been conducted for this set, we naturally take the current recovery as the new initial guess. It allows the frequency to be increased (assuming this current reconstruction belongs to the new basin of attraction area). Following the stability, increasing the frequency means that we can increase the scale, i.e., refine to gain accuracy. Hence the next couple (ω_1, N_1) is chosen where $\omega_1 \geq \omega_0$ and $N_1 \geq N_0$.
3. We continue the iterative scheme successively updating the frequency and scale, to reach the desired resolution eventually.

In this project, we have been able to give foundations to this very natural algorithm with joint progression in frequency and scale, from the stability and convergence point of view. In the following, we are only using an empirical progression in frequency (usually 1 Hz step) and scale. The next task is to be able to quantify the progression and introduce automatic parameters deciding on the next frequency and scale. It represents an ongoing investigation, where the conjecture is to expect the wavelength to play a major role in the decision, associated with the stability constant bounds definition and convergence estimation.

6.2 Acoustic 2D Marmousi model

The acoustic Marmousi model has been presented to illustrate the effect of reconstruction using the gradient information only or with the Hessian in Section 4.7. In this example, we used synthetic data modeled in the frequency domain directly, here we use time domain data to start with, and incorporate noise. To make this section independent we first recall the main properties of the medium and show the velocity model.

6.2.1 Acquisition setup and data

The Marmousi model is a two-dimensional acoustic medium of size 9.2×3 km, very popular in seismic (extended for elastic by the work of Martin (2004) that we later study in Section 6.9). It is defined with a wave speed for which the velocity varies between 1450 and 5500 $\mathrm{m\,s^{-1}}$. This velocity is presented in Figure 6.1. This velocity model has several structures which can be assimilated to faults in the subsurface. In this acoustic example we further assume to have a constant density $\rho = 1000\ \mathrm{kg\,m^{-2}}$.

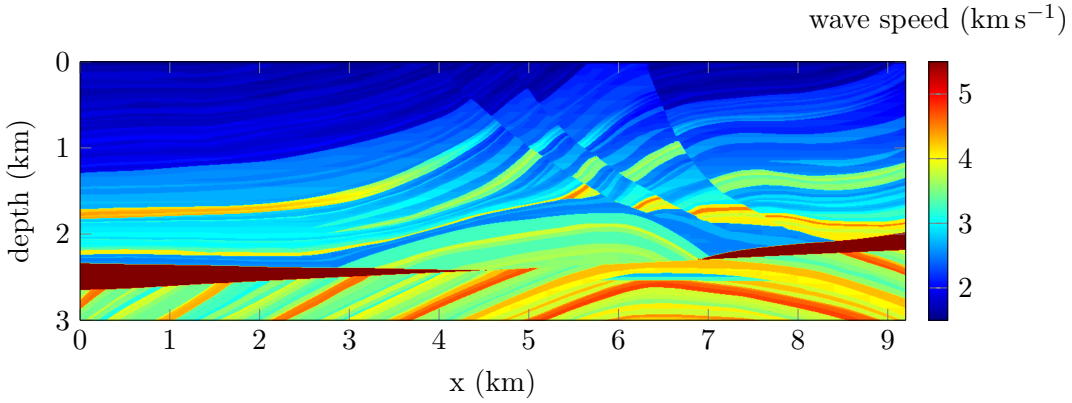


Figure 6.1: Wave speed for the acoustic Marmousi medium of size 9.2×3 km, the velocity varies between 1450 and 5500 $\mathrm{m\,s^{-1}}$.

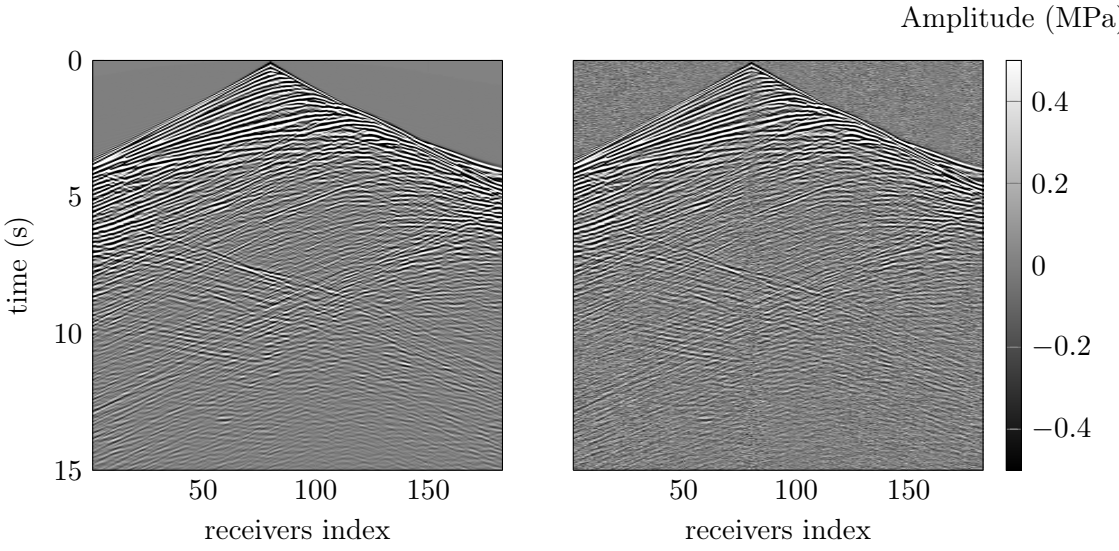
The seismic acquisition is conducted with 91 sources located near the surface, every 100 m at the exception of the corners. For each source, 183 receivers record the pressure information resulting from the wave propagation. In this experiment we consider an acoustic medium with constant density so that the propagation of waves is given by the Helmholtz equation. We generate data in the time domain using the research code Hou10ni[5].

[5]This code is developed at INRIA Bordeaux Sud-Ouest, in the Project-Team Magique 3D; it uses Interior Penalty Discontinuous Galerkin (IPDG) method for the discretization, see (https://team.inria.fr/magique3d/software/hou10ni/)

It results in time domain traces, obtained for every source. We further incorporate white Gaussian noise for the signal recorded by a receiver; here the noise is generated independently for each receiver and each source. More precisely by denoting $S_{k,j}$ the signal recorded by the receiver k for the source j, the signal to noise ratio is set to be of δ dB, we have

$$10 \log_{10} \left(\frac{S_{k,j}^T S_{k,j}}{N_{k,j}^T N_{k,j}} \right) = \delta, \tag{6.1}$$

where $N_{k,j}$ represents the noise vector applied to the original signal $S_{k,j}$ and where its values follow Gaussian distribution, ^{T} stands for the transposed. In this experiment, we consider $\delta = 5$ dB. The original trace and the one with noise are presented for a central shot in Figure 6.2



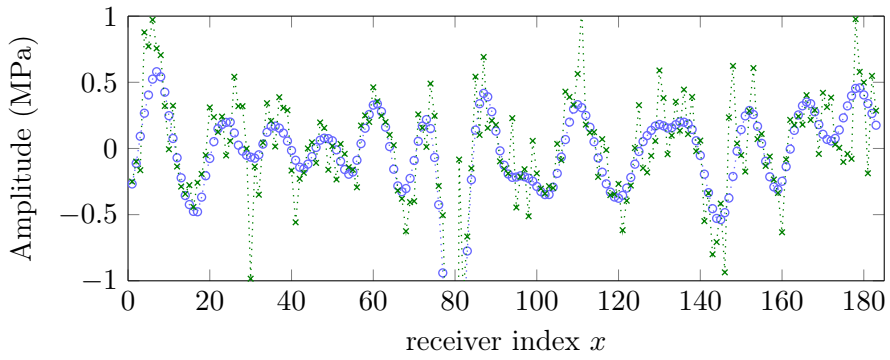
(a) Original trace

(b) Trace with added Gaussian white noise

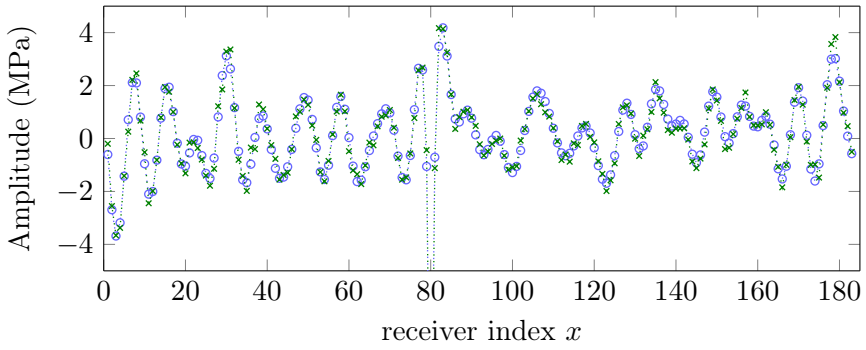
Figure 6.2: Time domain trace data for a central shot simulated using the Marmousi model. The horizontal axis represents the receivers index so that every column is the record of one receiver for this central source. The devices record the amplitude of the pressure field.

The Fourier transform is applied for the resulting time domain seismic traces in order to obtain the frequency domain data. Because our original source wavelet has a 10 Hz frequency peak for the source (Ricker source), the frequency data suffer from the noise at (relatively) low frequency meanwhile the higher frequencies remain slightly impacted. Figure 6.3 demonstrates this effect, where the 3 Hz content is extremely noised and barely usable while the 5 Hz content stays relatively similar to the original one without noise.

The initial model for the iterative reconstruction FWI algorithm is presented in Figure 6.4. It is a one-dimensional profile with increasing velocity with depth. None of the structures are known and the amplitude of the velocity is much lower than the true one.



(a) 3 Hz data component



(b) 5 Hz data component

Figure 6.3: Real part of the Fourier transform of the pressure measurements recorded by the receivers for a centrally localized shot illustrated Figure 6.2. The blue circles (⋯○⋯) represent the Fourier transform of the original data, the green crosses (⋯×⋯) depict the Fourier transform of the trace where Gaussian white noise has been incorporated.

In particular, we test our algorithm when no prior information is known on the models, so that the robustness and convergence of the scheme are emphasized.

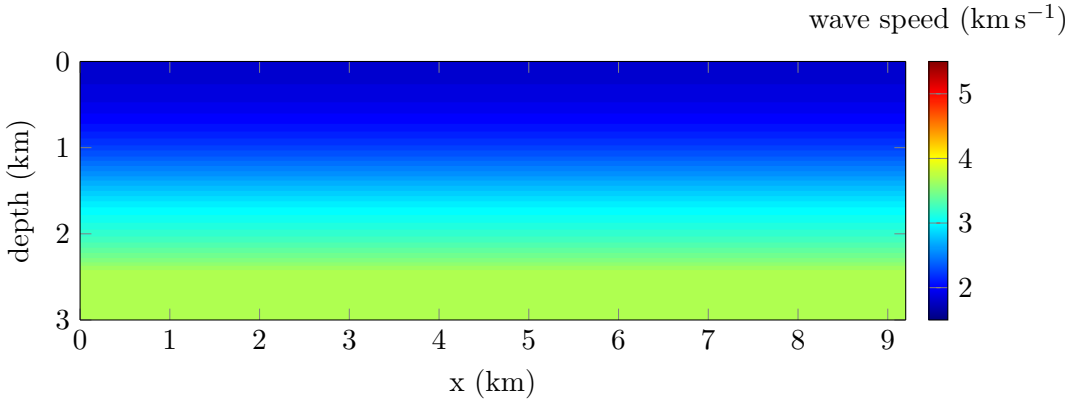


Figure 6.4: Initial wave speed for the iterative minimization algorithm and reconstruction of the Marmousi model. The model has no information on any of the true structures and consists in a one-dimensional variation of velocity with depth.

6.2.2 Reconstruction

For the reconstruction we compute the Fourier transform of the time domain data (with noise) for frequencies from 2 to 10 Hz. Furthermore, we only use the integer set of frequency. In Section 4.7, we have compared different methods to conduct the iterative minimization algorithm. We have seen that the use of second order information via the Hessian can help the reconstruction but can be computationally expensive. Here we limit the algorithm to the use of nonlinear conjugate gradient method which only necessitates the computation of the gradient of the cost function (see Subsection 4.6.1).

In addition to the model, the value of the source is also updated at each iteration. The initial value is given by the Fourier transform of the time domain source wavelet but this wavelet is not given with perfect accuracy. Thus we use the source reconstruction formula, depicted in Subsection 4.6.4.

The FWI algorithm has been implemented for the finite differences and continuous or discontinuous Galerkin discretizations. In the sequel we compare the reconstruction depending on the choice of discretization. A major difference is that Galerkin approach easily accounts for a change of order of approximation, contrary to the finite differences. In particular, our implementation is limited to first order finite differences so it is not completely fair to compare the method as is. However, we can already verify the proper convergence of the iterative scheme and the accuracy of the reconstruction (and reduce the step size of the discretization using finite differences to increase precision).

6.2.2.1 Discontinuous Galerkin discretization

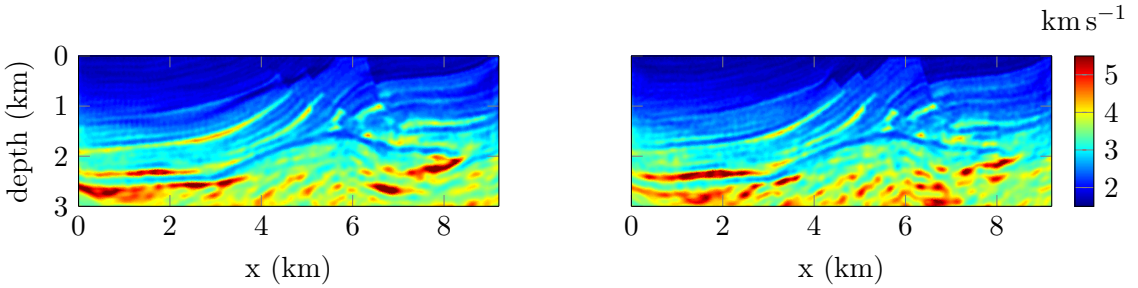
The Helmholtz equation is discretized using discontinuous Galerkin method. Because of the constant density, the only geophysical parameter to be reconstructed is the velocity. The order of approximation is set to 2 and we have a total of 734 022 degrees of freedom.

Cost function selection We have identified two cost functions for the minimization of the residuals: the traditional least squares functional and the logarithmic representation (see Subsection 4.3.2). We remind the two cost functions:

$$\mathcal{J}(m) = \frac{1}{2}\|\mathcal{F}(m) - d\|^2 \qquad \text{lest squares functional,}$$

$$\mathcal{J}(m) = \frac{1}{2}\|\log\big(\mathcal{F}(m)\big) - \log(d)\|^2 \qquad \text{logarithmic functional.}$$

We perform 20 iterations for every frequency between 2 and 10 Hz, with 1 Hz increment. We have a sequential progression in frequency. The reconstructions for the least squares and logarithmic cost function are shown in Figure 6.5.



(a) Wave speed reconstructed using the least squares misfit function.

(b) Wave speed reconstructed using the logarithmic misfit function.

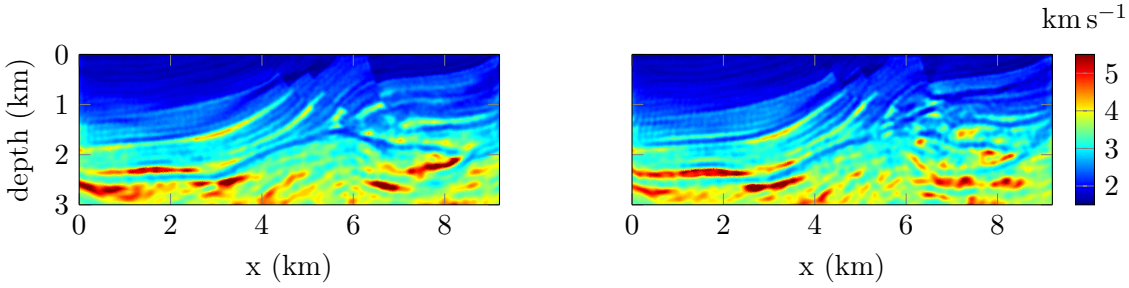
Figure 6.5: Reconstructions of the acoustic Marmousi wave speed assuming constant density and using frequencies from 2 to 10 Hz. 20 iterations are performed per frequency (for a total of 180 iterations).

Both reconstructions are accurate in the upper part of the medium (at low depth). Yet there is a lack of sharpness in the different layers compared to the true model. The use of the logarithmic cost function gives more perturbed recovery for the deepest structures, this can be seen in the deepest central, round structure. Despite the noise in the data, the overall reconstruction provides the appropriate subsurface interfaces and the values for the velocity are well attained.

Starting frequency In the first experiment the lowest frequency is set to 2 Hz. Here the lowest available frequency is set to 3 or 4 Hz. The reconstruction is presented in Figure 6.6. where the least squares cost function has been used.

The necessity of low frequency content in the data is illustrated by this example. Starting with 4 Hz clearly complicates the reconstruction and less accuracy is obtained in the different layers of the model. It is of importance for seismic applications because low frequency content is usually extremely affected by the noise, and even missing.

Order of approximation The order of approximation for the DG method can be raised to increase the accuracy of the numerical solution of the wave propagation. We investigate the usefulness of increasing the accuracy of the modeling in the context of the inverse problem. Because the observed data have noise and limited accuracy in general, it is not clear if improving drastically the accuracy of the forward problem will provide benefits.

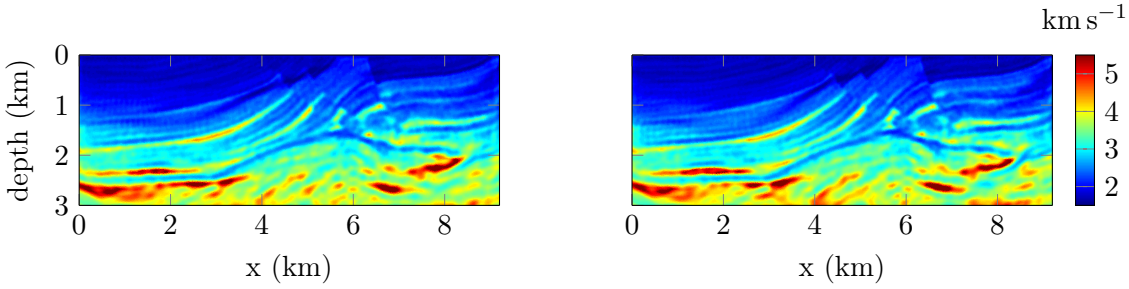


(a) Wave speed reconstructed using 3 Hz for the lowest frequency in the data.

(b) Wave speed reconstructed using 4 Hz for the lowest frequency in the data.

Figure 6.6: Reconstructions of the acoustic Marmousi wave speed assuming constant density and using different starting frequencies.

The reconstruction starting with 2 Hz frequency and using the least squares cost function is shown in Figure 6.7 where the DG method is used with order of approximation 3 and 4.



(a) Wave speed reconstructed using DG method with order of approximation 3.

(b) Wave speed reconstructed using DG method with order of approximation 4.

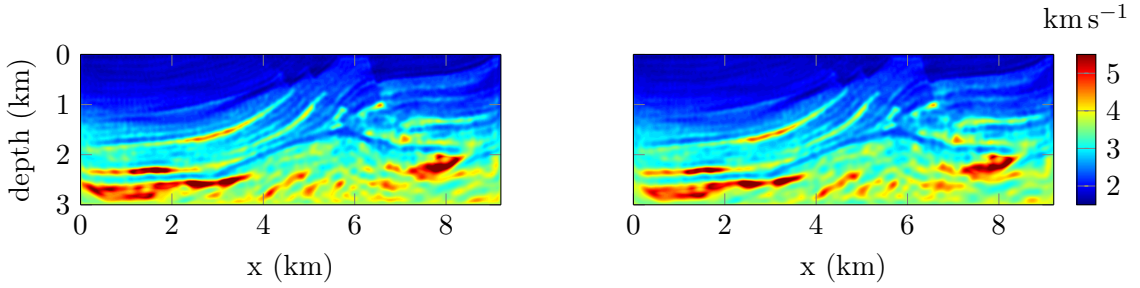
Figure 6.7: Reconstructions of the acoustic Marmousi wave speed assuming constant density using different order of approximation for the DG discretization method.

There are no benefit in increasing the order of approximation for the forward discretization in this test. It is clear that due to the limited accuracy of the observed measurements, there is no interest in simulating data with an over demanding accuracy.

6.2.2.2 Finite Differences discretization

Following the same idea, the experiment is conducted using a finite differences discretization and different step sizes. The results are shown in Figure 6.8.

We clearly obtain similar accuracy compared to DG. Here the decrease of the step size does not play any role, similarly to the increase of order of approximation in DG. This can be explained by the limited accuracy of the data which are the more important factor



(a) Wave speed reconstructed using FD method with discretization step of 10 m

(b) Wave speed reconstructed using FD method with discretization step of 5 m

Figure 6.8: Reconstructions of the acoustic Marmousi wave speed assuming constant density using different discretization step for the FD discretization method.

in the reconstruction. The precision of the forward problem is less important and gaining few digits of accuracy in the modeling does not improve the resolution when it is already beyond the noise level. However, some other aspect of the numerical resolution, such as the dispersion must be accounted for (e.g., Barucq et al. (2017) and the references therein).

This simple geophysical test case is the initial way to verify the FWI algorithm that we have implemented and control some of the options. The accuracy of the reconstruction is to be put in perspective with the relatively gentle geophysical structures. This also justifies the few differences obtained in the recovery with the different methods. For example, it is clear that the DG method is naturally more adapted to the topography, benefiting from the freedom in the mesh, but this is not useful in such simple experiment. Yet we have shown that the recovery of the deepest structures is already a complex procedure.

6.3 Acoustic 2D Overthrust model

The acoustic two-dimensional Overthrust (OT) model has been presented when comparing the reconstruction from gradient descent algorithms and the use of Hessian information, in Section 4.7. Here we will review the model to make the present section consistent. As for the Marmousi test case we use synthetic data generated in the time domain, with added noise.

6.3.1 Acquisition setup and data

The Overthrust model is originated from the SEG-EAGE 3D velocity model of Aminzadeh et al. (1994) where a two-dimensional section is extracted. It is of size 20×4.65 km. We assume a constant density with value $\rho = 1000\ \mathrm{kg\,m^{-3}}$ so that the medium is only defined with the wave speed, which is pictured in Figure 6.9. It varies between 2000 to 6000 $\mathrm{m\,s^{-1}}$.

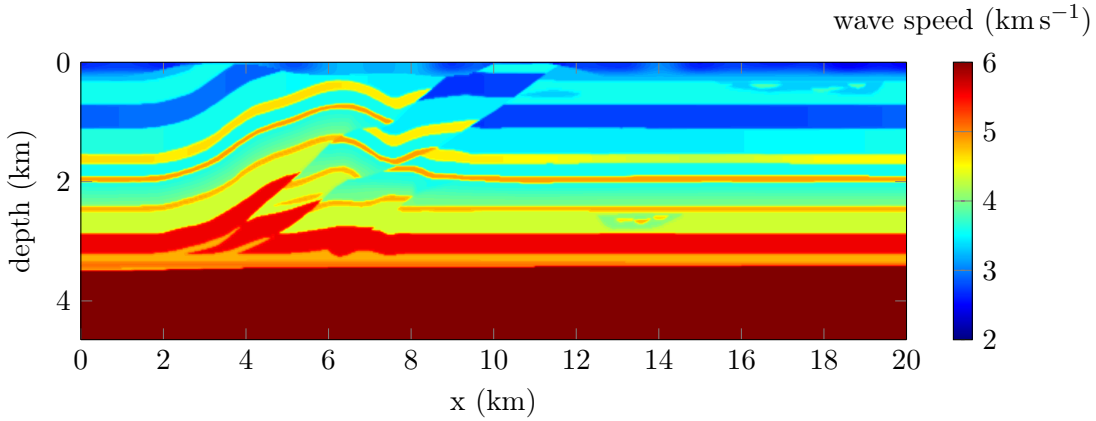


Figure 6.9: Wave speed for the acoustic Overthrust (OT) medium of size 20×5.65 km, the velocity varies between 2000 and 6000 $\mathrm{m\,s^{-1}}$.

The wave speed is composed of several layers with increasing velocity with depth. Most of the area consists in horizontal layers at the exception of a circular fault shape in the upper left part of the medium.

For the seismic acquisition, we define 199 sources equally spaced in the near surface zone. For every source we have 399 receivers that record the wave propagation, in terms of pressure field. Synthetic data are generated in the time domain with the already mentioned software Hou10ni. It is illustrated with a seismic trace resulting from a central shot in Figure 6.10.

Figure 6.11 represents the initial wave speed for the iterative reconstruction of the OT medium. We do not assume any knowledge of the structures and take a one-dimensional profile, to highlight the behavior of the algorithm when no initial assumption is made on the subsurface coefficients.

6.3.2 Reconstruction using compression

We first compute the Fourier transform of the data for integer frequencies between 1 and 10 Hz. The FWI algorithm performs 20 iterations per frequency. We employ the least

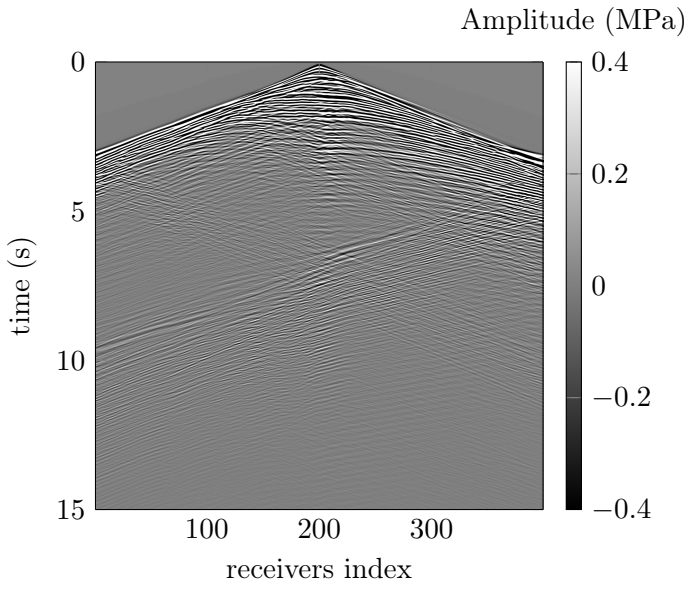


Figure 6.10: Time domain trace data for a central shot simulated using the OT model. The horizontal axis represents the receivers index so that every column is the record of one receiver for this centrally located source. The devices record the amplitude of the pressure field.

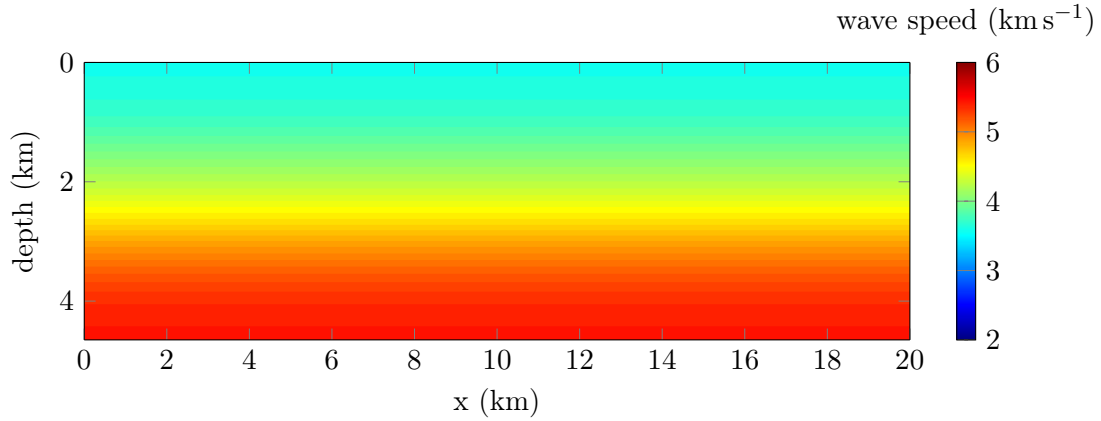


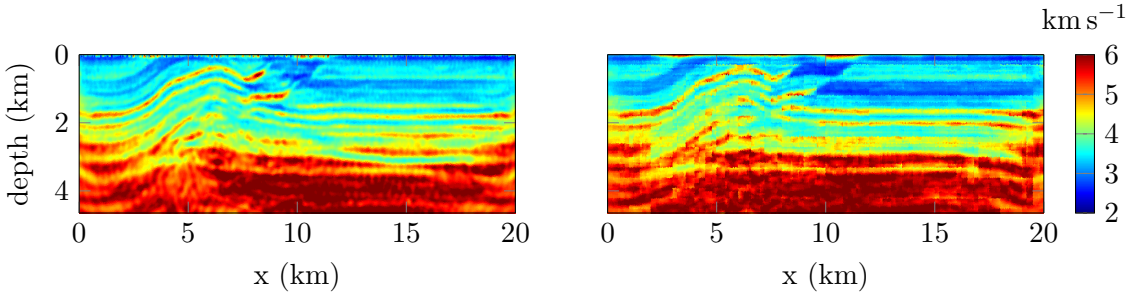
Figure 6.11: Initial wave speed for the iterative minimization algorithm and reconstruction of the OT model. The velocity is represented with a one-dimensional variation, increasing with the depth.

squares misfit function. In this experiment we compare the use (or not) of model representation based on compression during the algorithm, Table 6.3 indicates the relation we select between the number of unknowns in the representation and the frequency. Namely, following the multi-level scheme of Section 6.1, the low frequencies use small amount of unknowns and the scale is refined with increasing frequency. We further employ a structured decomposition to produce the subdomains, and piecewise constant representation. In this experiment, the original wave speed is represented with a mesh defined with 118 848 elements (without any consideration of geometry). We drastically reduce the number of possible coefficients for the velocity, so that the more refined scale employs 14 337 sub-

domains, and the initial (coarser) scale is of 330 coefficients, see Table 6.3. Figure 6.12 compares the final reconstruction after the 10 Hz iterations depending on the use, or not, of reduced model representation.

frequency iterations	**model representation (number of unknowns)**	**percentage compared to the original scale**
1	330	0.30%
2	504	0.42%
3	816	0.70%
4	1 008	0.85%
5	1 872	1.60%
6	3 264	2.75%
7	5 781	4.90%
8–10	14 337	12.10%

Table 6.3: Relation between frequency and scale selected for the reconstruction of the acoustic OT model.

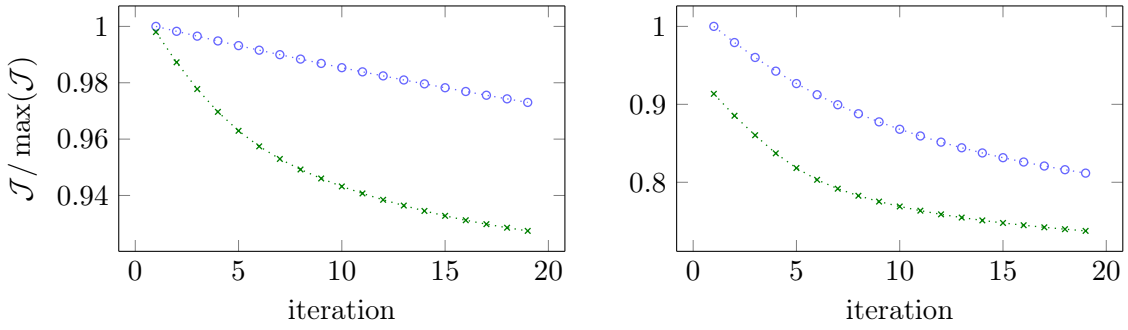


(a) Standard reconstruction with no specific model partition.

(b) Reconstruction using evolution of model partitions with frequency, prescribed in Table 6.3.

Figure 6.12: Reconstructions of the acoustic OT wave speed assuming constant density and using frequencies from 1 to 10 Hz. 20 iterations are performed per frequency (for a total of 200 iterations).

We observe a consequent improvement in the reconstruction when we employ piecewise constant model partitions to reduce the number of coefficients. In particular, in this experiment we benefit from the natural structures of the original medium with the different layers. We see that even the upper part of the reconstruction, near the surface, is much better recovered when a model partition is used. Also it clearly acts as some regularization by avoiding some artifact oscillatory behaviors that appear in the standard reconstruction. We remark that few information is retrieved for the deepest corners but once again the use of compression seems to increase the resolution. Despite the relatively low number of unknowns, 12% of the original scale, it is still able to capture the variations and the coarseness of the representation is not visible. The benefits are further illustrated in Figure 6.13 with the plot of the cost function evolution with the iterations for the starting frequencies (1 and 2 Hz iterations).



(a) 1 Hz iterations.

(b) 2 Hz iterations.

Figure 6.13: First iterations of the minimization algorithm for the reconstruction of the OT model using the starting wave speed of Figure 6.11 The blue dots (···o···) do not use any compression meanwhile the green crosses (···×···) reduce the model partition to 330 unknown coefficients at 1 Hz and 504 coefficients at 2 Hz, see Table 6.3.

The convergence is obtained in both cases and we remark a much faster decrease when model partition is used. The early iterations already provide great benefits and it is further confirmed when increasing the frequency. It illustrates how the reduction of unknowns in the model improves the convergence in this experiment.

6.3.3 Reconstruction using complex frequencies

We examine the effect of complex frequencies during the first iterations of the minimization algorithm. The new formula for ω is depicted in the Equation (1.25), with an imaginary and a real part, $-\omega^2 = (\sigma + 2\mathrm{i}\pi f)^2$. The transform of the time domain data is conducted appropriately using the seismic trace where no noise has been included. Here we do not take any 'zero' Fourier frequency part but only complex frequencies with both real and imaginary parts. Following the analysis provided in Section 5.2, the imaginary (related to the Fourier domain) part of the complex frequency is chosen to be increasing, while the real (Laplace domain) part of the frequency is chosen to be decreasing. The selected sets of complex frequencies are given in Table 6.4, with a total of eight complex frequency sets, for which the logarithmic norm is employed for the misfit function.

Starting from the initial model given in Figure 6.11, the reconstruction obtained after the last of the complex frequencies is given in Figure 6.14.

The use of complex frequencies provides an increase of the basin of attraction, which is especially important when no information is initially known on the model. It results in the recovery of a smooth wave speed with important information on the structures obtained for the near surface area. In particular, the horizontal layers of low velocity are

Set of iterations	**damping parameter (σ)**	**Fourier frequency (f)**
1	10	1 Hz
2	7	1 Hz
3	5	1 Hz
4	2	1 Hz
5	10	2 Hz
6	7	2 Hz
7	5	2 Hz
8	2	2 Hz

Table 6.4: Set of complex frequencies employed for the reconstruction of the acoustic OT model. Every set is performed with 20 iterations of the minimization algorithm. The complex frequency is defined by $-\omega^2 = (\sigma + 2i\pi f)^2$.

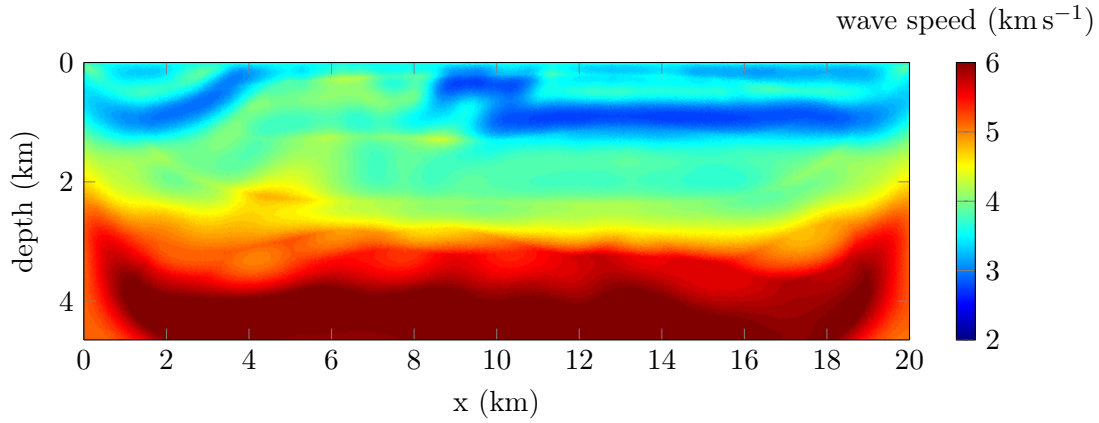


Figure 6.14: Reconstructions of the acoustic OT wave speed assuming constant density and using the complex frequencies set depicted in Table 6.4.

well approximated. After the initial set of complex frequencies, we continue the iterative reconstruction by only employing the standard Fourier frequency set that was originally taken for the reconstruction of Figure 6.12: from 2 to 10 Hz with 1 Hz increment and 20 iterations per frequency. In Figure 6.15, the final reconstruction is presented. For those iterations, we employ the traditional least squares functional.

The benefits of using complex frequencies for the initial iterations are clear. The structures of the wave speed are much better defined than when the set of frequencies is restricted to the Fourier domain. The horizontal layers of constant speed are identified with the appropriate values. It also improves the information retrieved for the deepest zone.

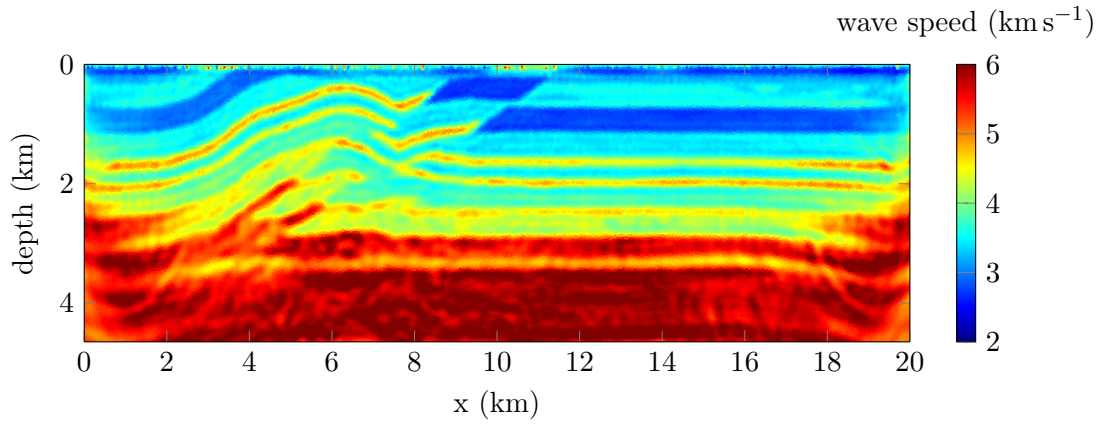
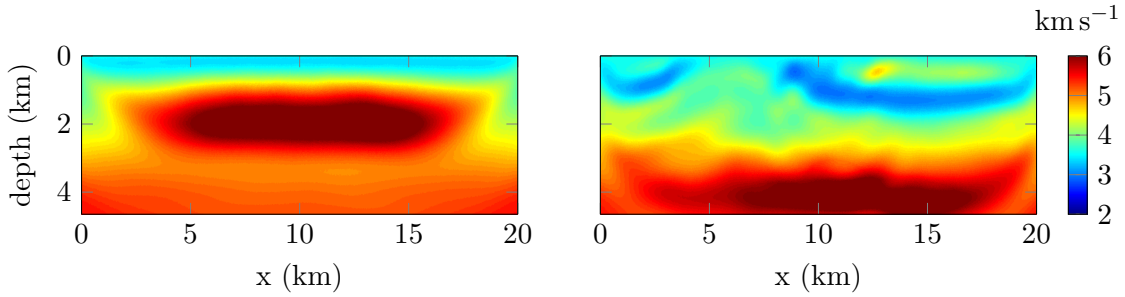


Figure 6.15: Reconstructions of the acoustic OT wave speed assuming constant density using frequencies from 2 to 10 Hz and starting with the reconstructed wave speed obtained from the complex frequencies iterations (Figure 6.14).

6.3.4 Incorporating noise in the traces

We now incorporate white noise in the time domain traces, following the same steps as for the Marmousi example, using Equation (6.1). We know that the low (and complex) frequencies are particularly affected by the noise, and we illustrate when deciding the level of noise to be 10 and 20 dB. We observe the impact of the noise on the data for the reconstruction using the complex frequencies sets (Table 6.4) in Figure 6.16.



(a) Reconstructed wave speed after complex frequencies from the data set where the level of added noise is 10 dB.

(b) Reconstructed wave speed after complex frequencies from the data set where the level of added noise is 20 dB.

Figure 6.16: Reconstructions of the acoustic OT wave speed assuming constant density and using the complex frequencies set depicted in Table 6.4. Different levels of white noise are initially applied onto the time domain seismic data set.

The complex frequencies reconstruction is considerably affected by the incorporation of noise in the observation. For low level of noise (20 dB, Figure 6.16(b)) the reconstruction is still correct with the smooth structures appearing, however, when the level of noise is

slightly larger (10 dB, Figure 6.16(a)) the reconstruction fails completely. It illustrates that the benefits of using complex frequencies have to be carefully considered because of their sensitivity to noise (as for low frequencies in general). We pursue the iterations with frequencies from 2 to 10 Hz with the 20 dB level of noise in the data set. The final reconstruction (starting from the complex frequency reconstruction) is pictured in Figure 6.17

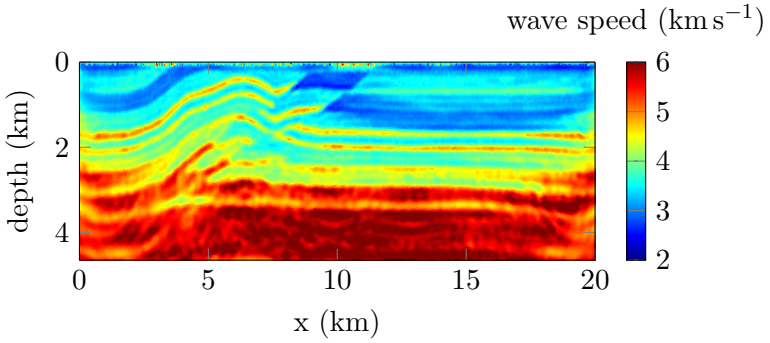


Figure 6.17: Reconstruction of the acoustic OT wave speed assuming constant density using complex frequencies from 2 to 10 Hz and starting with the reconstructed wave speed obtained from the complex frequencies iterations (Figure 6.16(b)). White noise is incorporated in the time domain seismic signals.

The final reconstruction captures the velocity structures and the effect of the noise appears in particular near the boundaries. We remark that the effect of noise is strongly related to the geometry of the subsurface coefficients to retrieved. For the reconstruction of the Marmousi wave speed in Section 6.2 we had a strong level of noise (5 dB) and the reconstruction was not very disturbed; here the noise has much more influence, because the geometry of the velocity requires lower frequency content.

Remark 6.1. *In this experiment, we have used the data with noise without any treatment. One could imagine to anticipate noise, and apply consequently a pre-processing of the data set to improve the accuracy. Hence the observed impact of noise could possibly be attenuated.*

6.3.5 Incorporating variable density

To test the performance of our algorithm, we now incorporate a variable density in the problem, so that we have a multi-parameter inverse problem. The density is created to have similar structures as the velocity and is pictured in Figure 6.18. From the velocity and the density we can extract the additional quantities that are useful for the parametrization, in particular the bulk modulus and the impedance, pictured in Figure 6.19. We observe that both density and wave speed have relatively low variations of profile, the minimum

and maximum values of each model lie in the same magnitude (10^3 m s^{-1} and 10^3 kg m^{-2} for the velocity and density respectively). However, the bulk modulus shows two orders of magnitude between the minimum and maximum value, one order for the impedance.

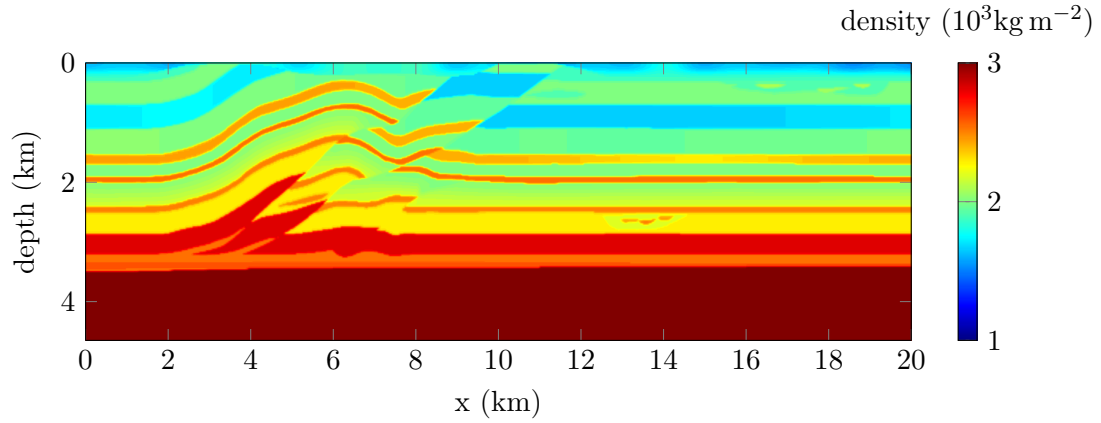
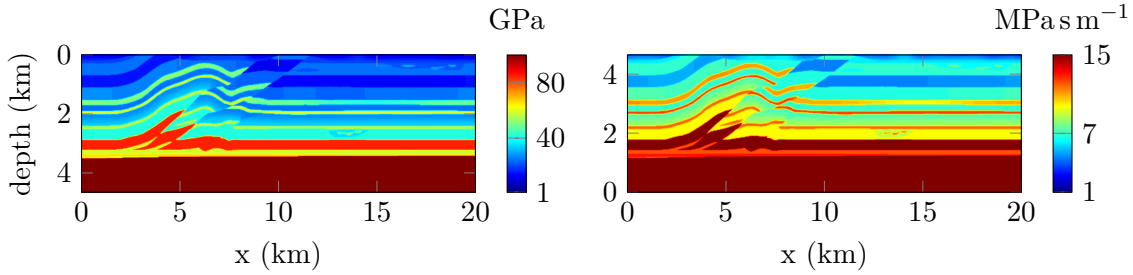


Figure 6.18: Density considered for the acoustic OT medium of size 20×5.65 km, it varies between 1500 and 3000 kg m^{-2}.



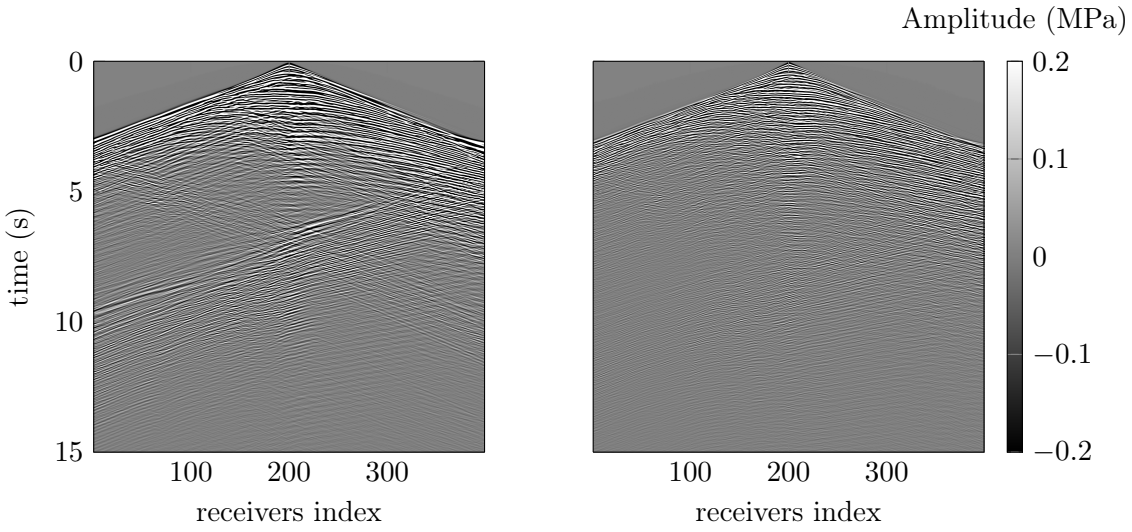
(a) Bulk modulus.

(b) Impedance.

Figure 6.19: Bulk modulus and impedance extracted from the OT wave speed and density, the bulk modulus is given by $\kappa = \rho c^2$, and the impedance $I = \sqrt{\kappa \rho}$.

We keep the same acquisition as above, using 199 sources with 399 receivers. The observation are obtained from time domain simulations and the seismic trace corresponding with a centrally located source is given in Figure 6.20, where we compare with the original experiment using a constant density. We see that the contrast is reduced when incorporating the variable density, in particular some of the central reflections do not appear clearly. In order to use complex frequencies, we do not incorporate noise.

The starting density profile used to start the subsurface model recovery is pictured in Figure 6.21, where none of the layers is assumed. It is simply a profile where the density increases with depth.



(a) Trace for constant density.

(b) Trace for variable density.

Figure 6.20: Comparison of time domain data trace for a central shot simulated using the OT model with constant or variable density. The devices record the amplitude of the pressure field.

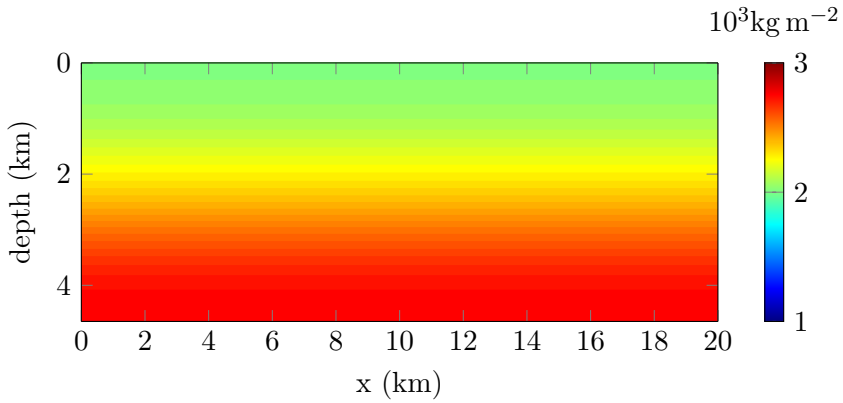


Figure 6.21: Initial density for the iterative minimization algorithm and reconstruction of the OT model. It is a one-dimensional variation, increasing with the depth.

6.3.5.1 Keeping the fixed initial density

Following the expertise developed in Chapter 5, Section 5.5 where the difficulties of recovering the density have been highlighted, we first disregard its reconstruction and focus on one single parameter. It means that the density is kept at the initial, smooth representation of Figure 6.21 all along the iterative procedure. Based on the wave equation (1.28), we select the inversion of the bulk modulus (more precisely its inverse, following the results of Chapter 5 on the parametrization).

We follow the similar frequency progression as in the previous test, starting with the

complex frequency set prescribed in Table 6.4. The reconstructed bulk modulus at this step is given in Figure 6.22.

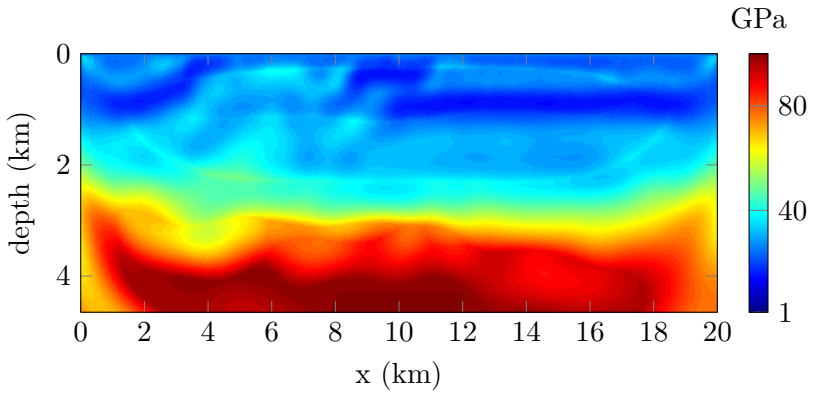


Figure 6.22: Reconstructions of the acoustic OT bulk modulus assuming variable density and using complex frequencies set depicted in Table 6.4.

We see that the bulk modulus is smoothly recovered from the complex frequencies sets. The incorporation of the density barely affects the resolution, and we retrieve very similar results as when the density was constant. Moreover, the density remains at its original smooth representation and is not updated, yet it does not affect the bulk modulus recovery. We pursue the iterative minimization starting with the complex frequencies result and using sequential frequencies from 2 to 10 Hz, with 1 Hz step. We still perform 20 iterations per frequency, the bulk modulus reconstructed is presented in Figure 6.23.

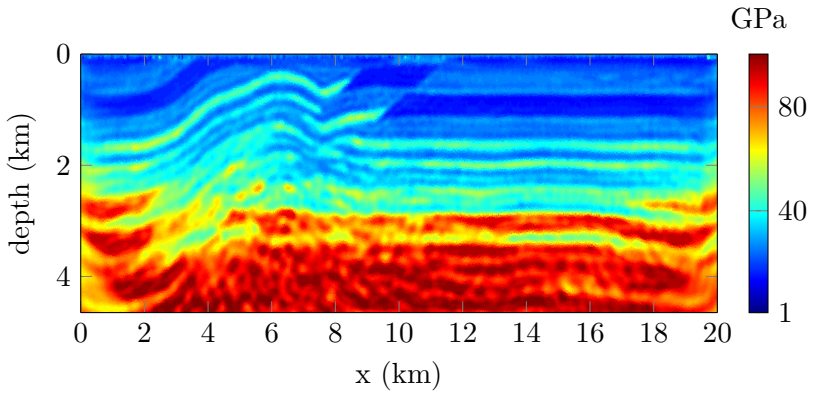
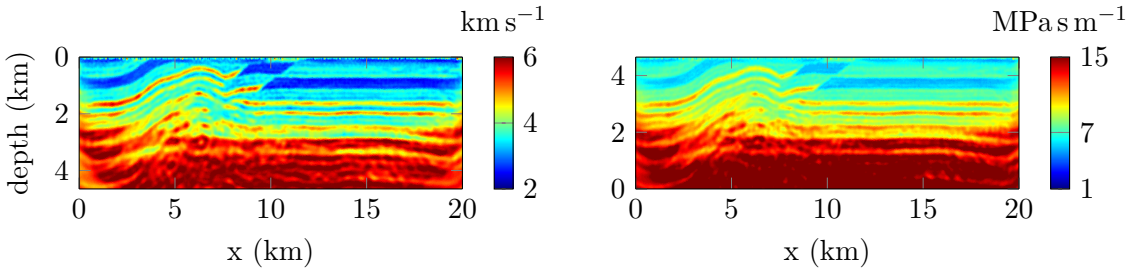


Figure 6.23: Reconstructions of the acoustic OT bulk modulus assuming variable density using frequencies from 2 to 10 Hz and starting with the reconstructed bulk modulus obtained from the complex frequencies iterations (Figure 6.22).

The lack of knowledge in the density does not affect the recovery. The resolution is accurate and the different layers are present. The amplitude of the bulk modulus, with the two orders of magnitude is partially retrieved with values slightly inferior to the exact ones.

We notice the lack of accuracy in both sides, particularly for the deeper part, similarly to the original experiment.

Remark 6.2 (Review of parametrization)**.** *We have chosen to invert the bulk modulus, based on the wave equation where it is separated from the density. However, one can process the iterative minimization choosing the velocity or the impedance as parameter of choice instead. From the experiments we have run in Chapter 5, we expect similar accuracy for the other parametrization (actually inverting the inverse of the selected parameter). It is confirmed by the experimental recoveries carried here, see Figure 6.24.*



(a) Reconstructed wave speed from inversion of the $1/c^2$.

(b) Reconstructed impedance from the inversion of $1/I^2$.

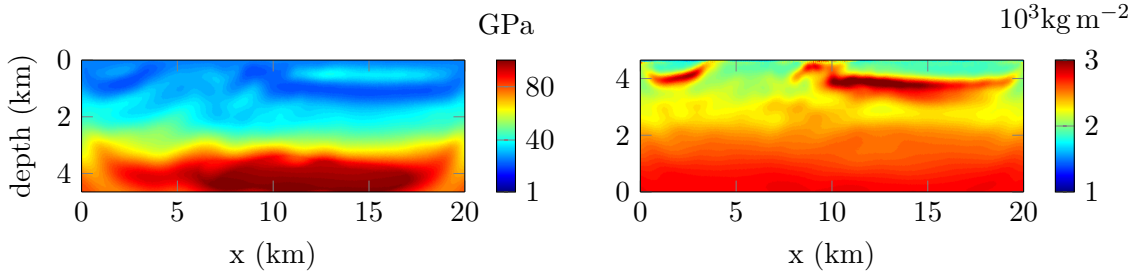
Figure 6.24: Reconstructions of the acoustic OT velocity and impedance assuming variable density. We compare the reconstruction depending on the choice of inverted parameter. The procedure is exactly similar as for the bulk modulus reconstruction, only the choice of parameter varies.

As expected, the accuracy is relatively similar for the two other selections, and the layers of the model appear with appropriate magnitude. We notice that the velocity makes appear sharper contrasts than the impedance, in particular in the upper region.

6.3.5.2 Simultaneous density reconstruction

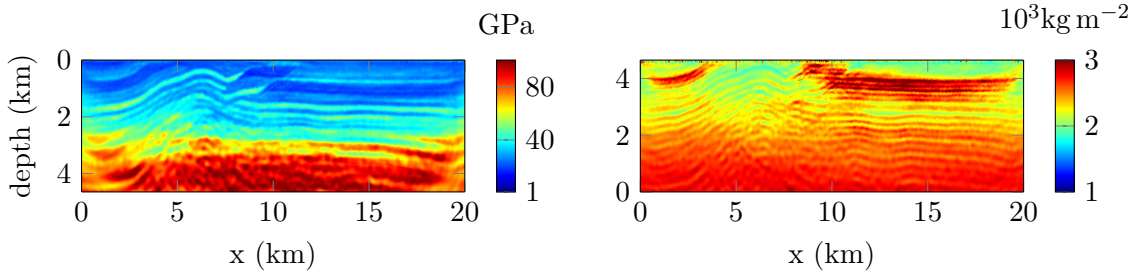
Let us try the multi-parameter reconstruction inverting simultaneously the density in addition to the bulk modulus. We follow the exact same steps, adding the reconstruction of the density in parallel with the bulk modulus. In Figure 6.25 we first show the recovery after the complex frequency set.

The reconstructed density shows some of the upper layer of the profile, but the magnitude is incorrect, the upper layer should have low density values while the recovery shows higher values. Concerning the bulk modulus, the smooth reconstruction remains relatively similar as when not inverting the density, however, the upper layer seems affected by the large (incorrect) increase of density. We pursue the reconstruction with frequencies from 2 to 10 Hz. The final reconstruction is shown in Figure 6.26.



(a) Bulk modulus. (b) Density.

Figure 6.25: Reconstructions of the acoustic OT bulk modulus and density using complex frequencies set depicted in Table 6.4.



(a) Bulk modulus. (b) Density.

Figure 6.26: Reconstructions of the acoustic OT bulk modulus assuming variable density using frequencies from 2 to 10 Hz and starting with the reconstructed modulus obtained from the complex frequencies iterations (Figure 6.22).

The final reconstruction confirms the partial results after the complex frequencies. The density layers are missing, at the exception of the near surface layers, but with inaccurate magnitudes. The reconstructed bulk modulus acquires the proper layers but suffers from the large magnitude of the density.

6.4 Acoustic 2D Sigsbee2A model

The acoustic two-dimensional Sigsbee2A model is a synthetic geophysical medium which was released by the Society of Exploration Geophysicists (SEG). On the contrary to the media that we have presented before, it contains an object with high contrast, which represents a challenging experiment.

6.4.1 Acquisition setup and data

The Sigsbee2A velocity model is illustrated in Figure 6.27. It has a size of 24.4×9.1 km. It consists of a smooth background where the speed varies from low to high values, with increasing depth. A strong contrasting object of high velocity is located in the middle.

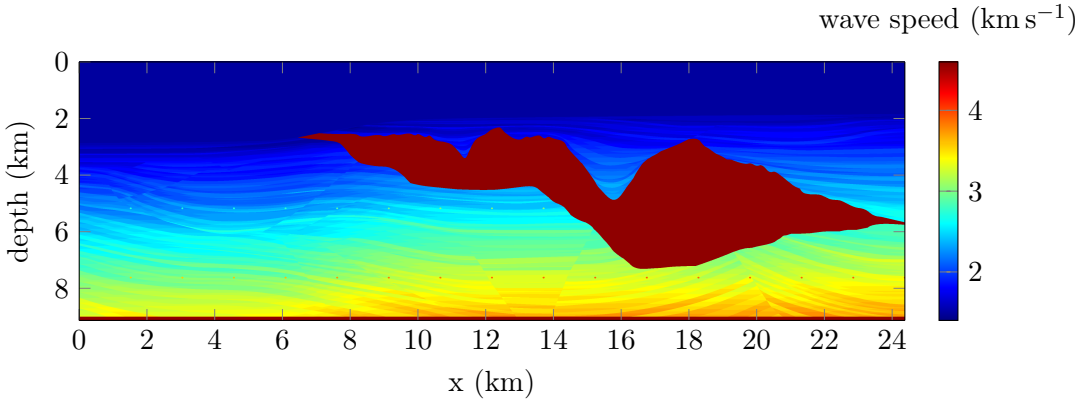


Figure 6.27: Wave speed for the acoustic Sigsbee2A medium of size 24.4×9.1 km, the velocity varies between 1430 and 4500 m s^{-1}.

Because of this large structure of high velocity, this model generates strong reflections. The shape of the object is also interesting with some sort of hole in the upper surface and a general width that strongly varies. The speed inside the structure is of 4500 m s^{-1} meanwhile the background velocity varies from 1500 to 3300 m s^{-1}. We take a constant density $\rho = 1000$ kg m^{-2}.

The seismic acquisition is conducted with 135 sources and 232 receivers record the pressure measurement for every source. Contrary to the previous test, case the data are directly generated in the frequency domain. Compared to the inverse algorithm, the order of discretization and mesh refinement are different. It allows us to avoid inverse crime, yet there is clearly much less noise and model error than for the previous examples where we started from time domain data.

For the reconstruction, we start with a model that does not have any information on the object, see Figure 6.28. This absence of knowledge is challenging for the algorithm as no expectation of an object involving strong reflections is anticipated. Regarding the background velocity, we roughly follow the variation of the actual one.

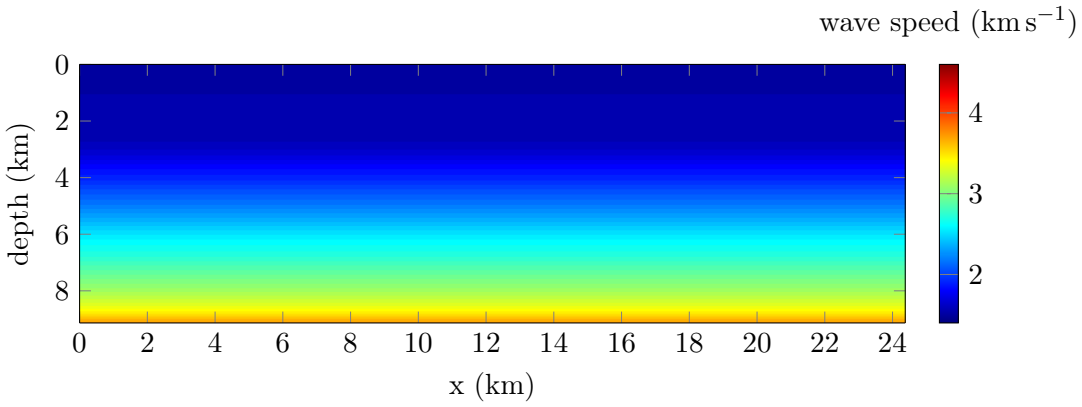


Figure 6.28: Initial wave speed for the iterative minimization algorithm and reconstruction of the Sigsbee2A model. The velocity is a one-dimensional profile of increasing velocity with depth.

6.4.2 Reconstruction using only Fourier frequency

The iterative minimization is conducted with 20 iterations per frequency, from 1 to 10 Hz and the least squares norm for the misfit function. The final reconstruction obtained is given in Figure 6.29.

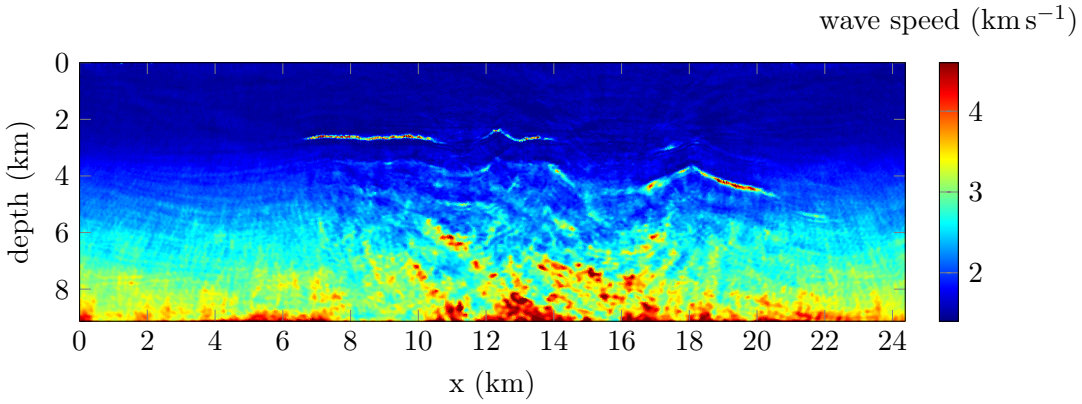


Figure 6.29: Reconstructions of the acoustic Sigsbee2A wave speed assuming constant density and using frequencies from 1 to 10 Hz. 20 iterations are performed per frequency (for a total of 200 iterations).

At the exception of a very vague idea of the upper interface between the background and the contrasting object, the minimization algorithm does not recover any information on the subsurface coefficient. The method fails to discover the large structure and the deepest parts are perturbed by artifacts. The failure of the recovery can be justified by the initial model we have taken where no information is assumed. We believe this initial guess to be outside the radius of convergence for the problem, preventing any possibility of reconstruction. For this reason the use of complex frequencies is justified, as it increases the size of the radius of convergence (see Chapter 5).

6.4.3 Reconstruction starting from complex frequency

The use of complex frequencies allows an increase of the size of the basin of attraction, as we have shown in Chapter 5. It is crucial in this case to identify the object, particularly because our initial guess does not have any knowledge of it. It should provide insight regarding the discovery of the obstacle. In this case, the frequency is given by the Equation (1.25): $-\omega^2 = (\sigma + 2i\pi f)$.

We initiate the algorithm starting with equating to zero the imaginary part, $f = 0$, so that only remains the damping parameter (i.e. we are positioned in the Laplace domain). Following the analysis of Section 5.2, the damping frequency varies from low to high values. Then we take $f \neq 0$, and in this context, the damping coefficient is chosen to vary from high to low values. Finally, we run the standard Fourier frequency set, by taking $\sigma = 0$. The summary and the progression of frequency is given in Table 6.5.

Set of iterations	**damping parameter (s)**	**Fourier frequency (f)**
1–5	$\{1, 2, 5, 7, 10\}$	0 Hz
6–8	$\{7, 5, 2\}$	1 Hz
9–11	$\{7, 5, 2\}$	2 Hz
12–21	0	{1,2,3,4,5,6,7,8,9,10} Hz

Table 6.5: Sets of frequencies employed for the reconstruction of the acoustic Sigsbee2A model. Every frequency is performed with 20 iterations of the minimization algorithm for a total of 420 iterations. The complex frequency is defined by $-\omega^2 = (\sigma + 2i\pi f)^2$.

After the initial iterations using the five first sets (keeping $f = 0$, see Table 6.5), the current reconstruction is pictured in Figure 6.30.

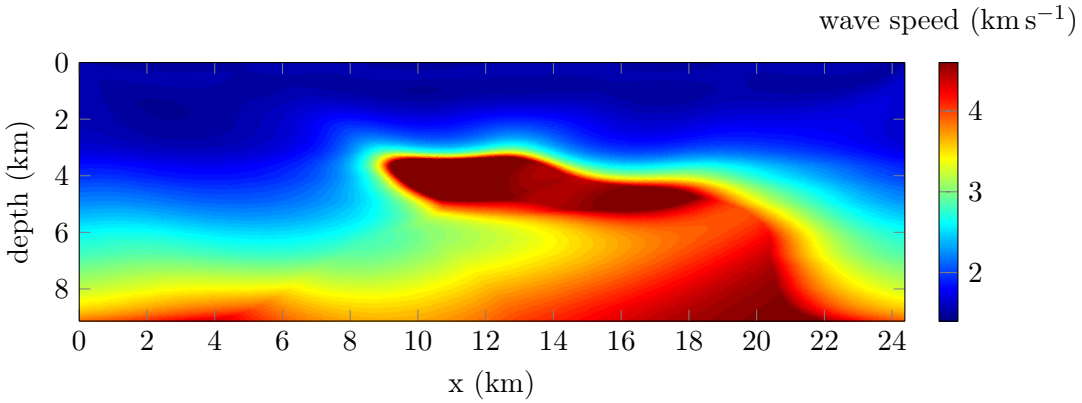


Figure 6.30: Current reconstructions of the acoustic Sigsbee2A wave speed after the iteration sets using only the damping coefficients (i.e. after the fifth set of iterations indicated in Table 6.5).

The object smoothly appears in the reconstruction, with velocity values well approximated. As expected, the use of damping frequency, by increasing the basin of attraction, allows insight regarding the strong reflection object lying in the area, which was impossible to discover before. It provides a relatively smooth model which is now an interesting initial base for higher frequencies iterations. Eventually, we show the final reconstructed wave speed in Figure 6.31.

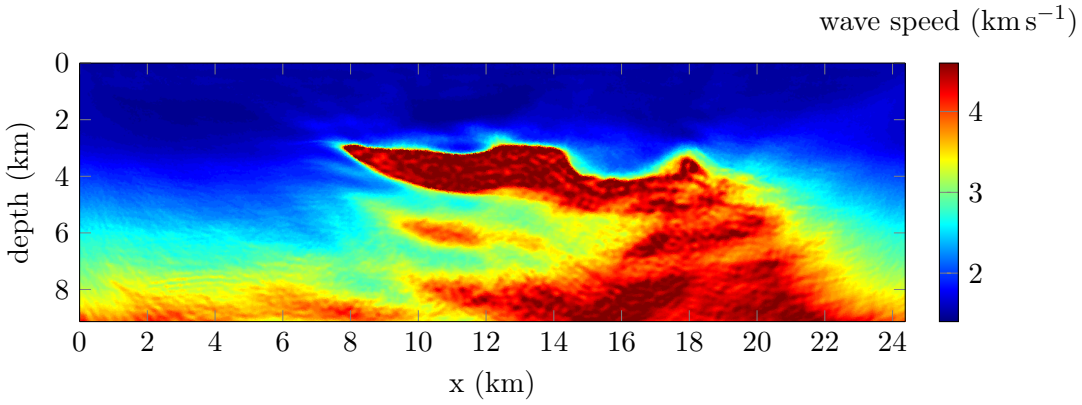


Figure 6.31: Final reconstruction of the acoustic Sigsbee2A wave speed when all sets of frequencies have been carried out (i.e. after the twenty-first set of iterations indicated in Table 6.5) for a total of 420 iterations.

The continuation of the reconstruction algorithm with higher frequencies sharpens the outline of the object. The upper interface is relatively well approximated, at the exception of the well. The deepest interface is much less accurate, and the object is globally much smaller than in reality. We also observe some artifacts appearing underneath the body. In this experiment, we have highlighted the difficulties arising when reconstructing strong reflection body, in particular when taking a blind guess for point of departure. We have demonstrated the usefulness of complex frequencies to allow the discovery of the object. This efficiency must be mitigated by the high impact on noise on the low and complex frequencies data. In Section 6.7, we further investigate such seismic reconstructions with toy experiments where the dimensions of the problem is shown to be crucial as well.

6.5 Acoustic 2D Seam model

6.5.1 Acquisition setup and data

We investigate the acoustic two-dimensional Seam model of size 35×15 km. This model originates from the Society of Exploration Geophysicists (SEG) in collaboration with oil

industries (SEAM referring to SEG Advanced Modeling Program[6]). The wave speed is given in Figure 6.32. The velocity model includes a salt dome near the surface. Similarly to the previous Sigsbee2A case, there is an object of high velocity immersed in a relatively smooth background. In this model we also observe deeper layers of velocity, underneath the salt dome. The higher celerity is located in a deep horizontal layer with value of 4800 $\mathrm{m\,s^{-1}}$, the celerity in the salt dome is of 4500 $\mathrm{m\,s^{-1}}$.

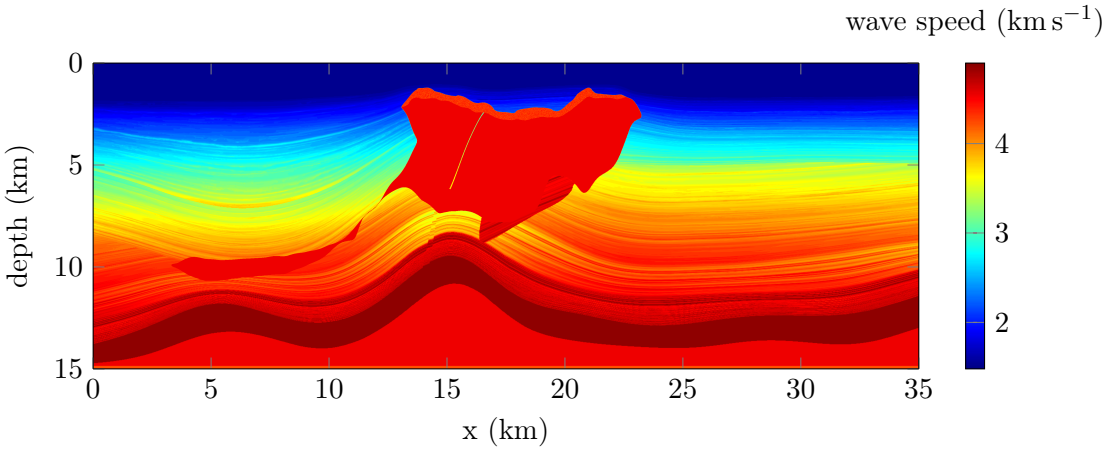


Figure 6.32: Wave speed for the acoustic Seam medium of size 35×15 km, the velocity varies between 1430 and 4800 $\mathrm{m\,s^{-1}}$.

The seismic acquisition is conducted with 174 sources located near the surface, equally distributed by 200 m. For every source, 349 receivers record pressure measurements. As usual the initial wave speed taken for the iterative reconstruction does not have any knowledge of the velocity structures, it is depicted in Figure 6.33. It is basically a smooth background.

Similarly to the Sigsbee2A experiment, the reconstruction of the salt dome is challenging and we demonstrate the necessity of using complex frequencies.

6.5.2 Reconstruction using only Fourier frequency

We start performing the iterative minimization in the Fourier domain only, with frequency from 1 to 10 Hz. We first compare the use of sequential frequency progression and group of frequencies. For the sequential progression, we use single frequencies from 1 to 10 Hz with 1 Hz increment. The second algorithm uses ten groups of frequency, where every group encompasses five consecutive frequencies with 0.2 Hz step, starting from the integer fre-

[6]Many models have arisen from this program (http://seg.org/News-Resources/Research-and-Data/SEAM), including elastic and anisotropic media. Here we only focus on the acoustic 2D, avoiding ambiguity for the keyword SEAM.

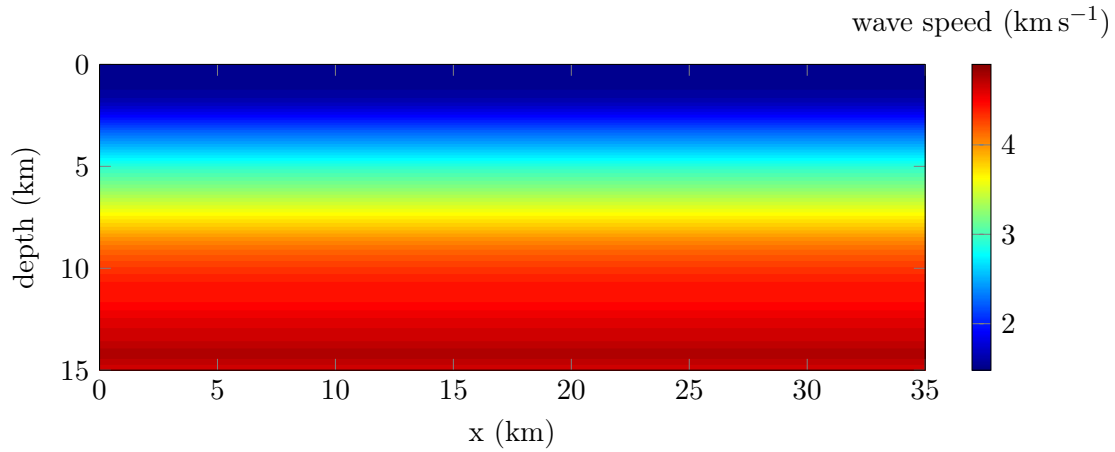
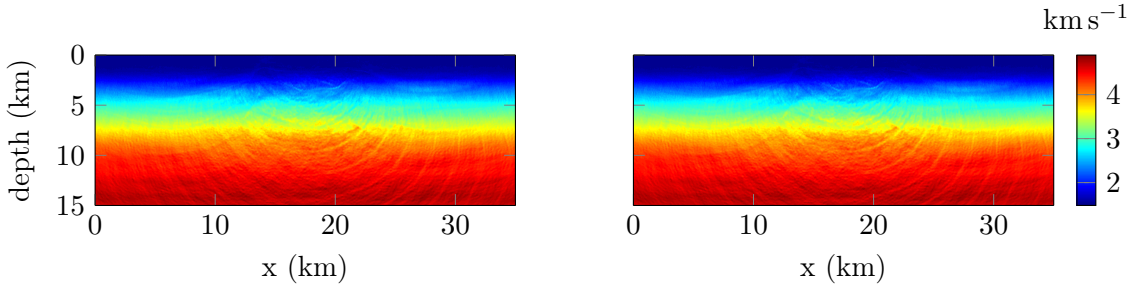


Figure 6.33: Initial wave speed for the iterative minimization algorithm and reconstruction of the Seam model. The velocity is a one-dimensional profile, increasing with depth.

quency; for example the first group employs the following frequencies: $\{1, 1.2, 1.4, 1.6, 1.8\}$ Hz. 20 iterations are processed for every set of frequencies (sequential or group) and the final recoveries are pictured in Figure 6.34.



(a) Reconstructed wave speed using sequential frequency progression.

(b) Reconstructed wave speed using progression of group of frequencies.

Figure 6.34: Reconstructions of the acoustic Seam wave speed using frequencies from 1 to 10 Hz. 20 iterations are performed per set of frequencies (for a total of 200 iterations).

There is no difference between the two methods, both fail. They do not recover any of the wave speed characteristics: the salt as well as the layers are missing. It tends to confirm the analysis we did in Chapter 5 that using several frequencies does not impact on the basin of attraction and does not help in the case where the initial model lacks of information (it can, however, help regarding noise robustness).

6.5.3 Reconstruction starting from complex frequencies

We proceed similarly to the Sigsbee2A medium, where the use of complex frequencies is motivated to increase the basin of attraction (cf Chapter 5). We expect that it will give us insight into the velocity structures, when our initial model does not have any prior information. Let us start with initial complex frequencies having a zero Fourier component, such that $f = 0$ in Equation (1.25). We follow the same sets as for the previous Sigsbee2A velocity reconstruction, which is reminded in Table 6.6. The reconstruction after all the five first sets (keeping $f = 0$ Hz) is presented in Figure 6.35.

Set of iterations	**damping parameter (σ)**	**Fourier frequency (f)**
1–5	$\{1, 2, 5, 7, 10\}$	0 Hz
6–8	$\{7, 5, 2\}$	1 Hz
9–11	$\{7, 5, 2\}$	2 Hz
12–21	0	{1,2,3,4,5,6,7,8,9,10} Hz

Table 6.6: Sets of frequency employed for the reconstruction of the acoustic Seam model. Every frequency is performed with 20 iterations of the minimization algorithm for a total of 420 iterations. The complex frequency is defined by $-\omega^2 = (\sigma + 2i\pi f)^2$.

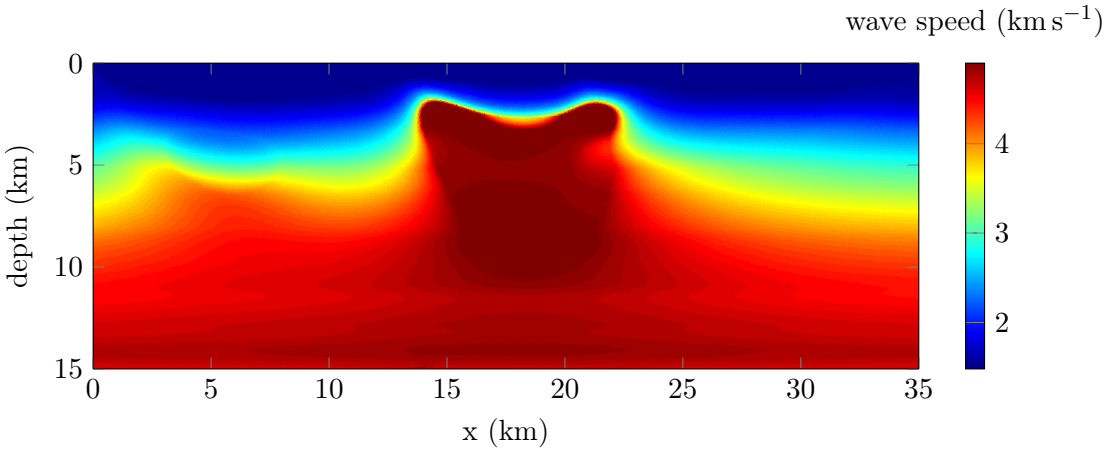


Figure 6.35: Current reconstruction of the acoustic Seam wave speed after the sets of iteration using only the damping coefficients (i.e. after the fifth set of iterations indicated in Table 6.6).

The use of the Laplace domain iterations gives a breakthrough on the salt dome position. It provides an initial smooth structure, which is revealed to be very appropriate. Once again it illustrates the benefit of the complex frequency when intending the reconstruction of model without knowledge on the subsurface structures. We then pursue the iterative reconstruction and show the final results obtained after the 10 Hz iterations in Figure 6.36

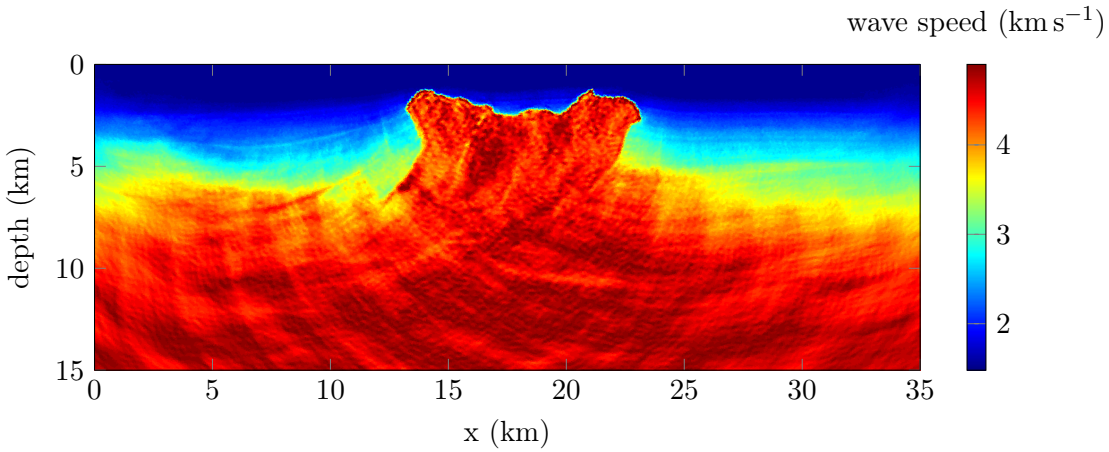


Figure 6.36: Reconstruction of the acoustic Seam wave speed when all sets of frequency have been carried out (i.e., after the twenty-first set of iterations indicated in Table 6.6, with maximum frequency of 10 Hz).

The use of complex frequencies has allowed the recovery of the salt dome in the model. In particular, the upper boundary of the salt is accurately defined, as well as some of its lateral boundaries. The values of the velocity are appropriately retrieved in the salt. We can barely distinguish the lower part of the salt dome but it is not very sharp. We can even notice the appearance of the little tail at the bottom left of the salt. However, the deepest layers are still missing, the maximal depth involved, 15 km, may be too demanding. In such context it would be interesting to see how the Hessian behaves and if it helps in the recovery of the salt or the deepest structures, as it should following our initial experiments of Section 4.7.

6.6 Acoustic 3D Louro model

6.6.1 Acquisition setup and data

The seismic inverse problem is conducted for the acoustic three-dimensional Louro model (courtesy of Total). It is a model of size $2.46 \times 1.56 \times 0.6$ km, characterized by several bodies which have high velocity values while the velocity background remains smooth. It is presented in Figure 6.37. In the figure, we select two sections in order to have a better visualization of the model: the vertical section is at fixed $y = 680$ m and the horizontal section is at fixed depth $z = 340$ m. For this experiment, the density is taken to be constant with $\rho = 1000$ kg m^{-3}.

We identify eight separate bodies of various sizes where some are located very near the

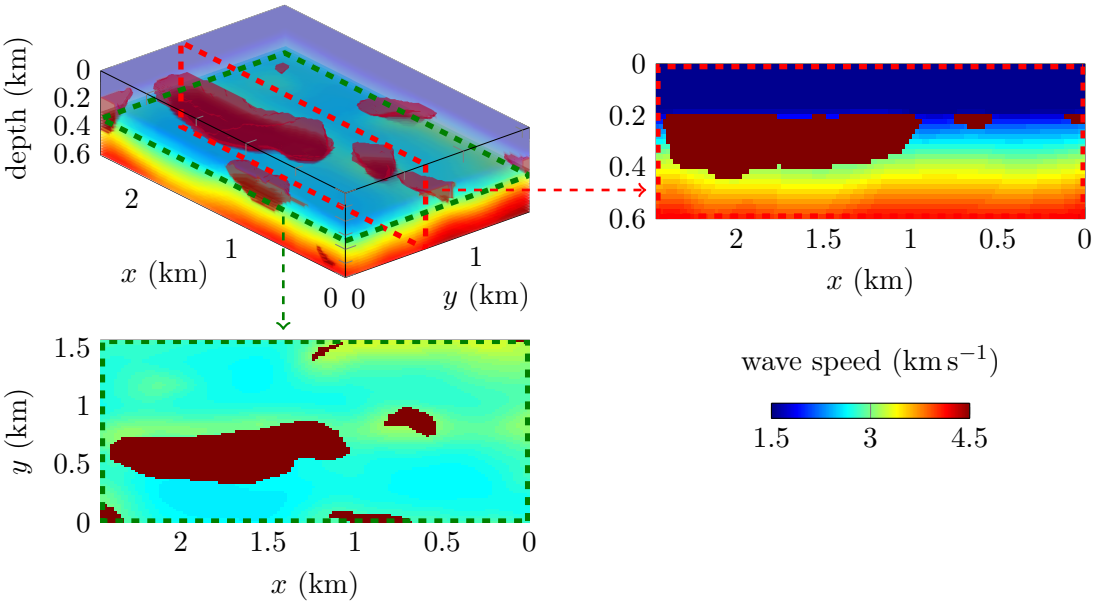


Figure 6.37: Wave speed for the three-dimensional acoustic Louro medium of size $2.46 \times 1.56 \times 0.6$ km (courtesy of Total). The velocity varies between 1500 and 4500 $\mathrm{m\,s^{-1}}$. Vertical section at $y = 680$ m is indicated with the red dashed frame and plotted in the right. The horizontal section at fixed depth $z = 340$ m is indicated with the green frame and shown at the bottom.

medium boundary. The velocity inside the contrasting objects is of 4500 $\mathrm{m\,s^{-1}}$ in a smooth velocity background which increases almost linearly with depth.

The seismic acquisition consists of a two-dimensional array of sources equally spaced. The source lattice is composed of 12 and 8 sources along the x and y axis respectively, for a total of 96 sources. More precisely the sources are positioned every 200 m along x and y axis, avoiding positions near the boundaries. Similarly the lattice for the receivers is composed of a map of 40×25 along the x and y axis respectively, they are positioned every 60 m in both directions. Data are generated in the time domain, creating three-dimensional seismic traces for every source. We illustrate the resulting data in Figure 6.38, for a centrally located shot. The 3D visualization of the trace is in the upper left part of the figure, two-dimensional sections are extracted at a fixed time ($t = 0.6$ s, bottom left) and for a fixed line of receivers positioned at $y = 855$ m (upper right). Naturally, the section with a fixed line of receivers produces a picture that resembles to a traditional two-dimensional trace. The propagating wave can be observed in the cross section at fixed time.

For the reconstruction, we start with a wave speed that does not have the knowledge of any of the salt domes, it is a smooth velocity background with a one-dimensional profile that increases with depth. It is presented in Figure 6.39. The background we have taken here is not even close to the original background velocity (true model removing the salt

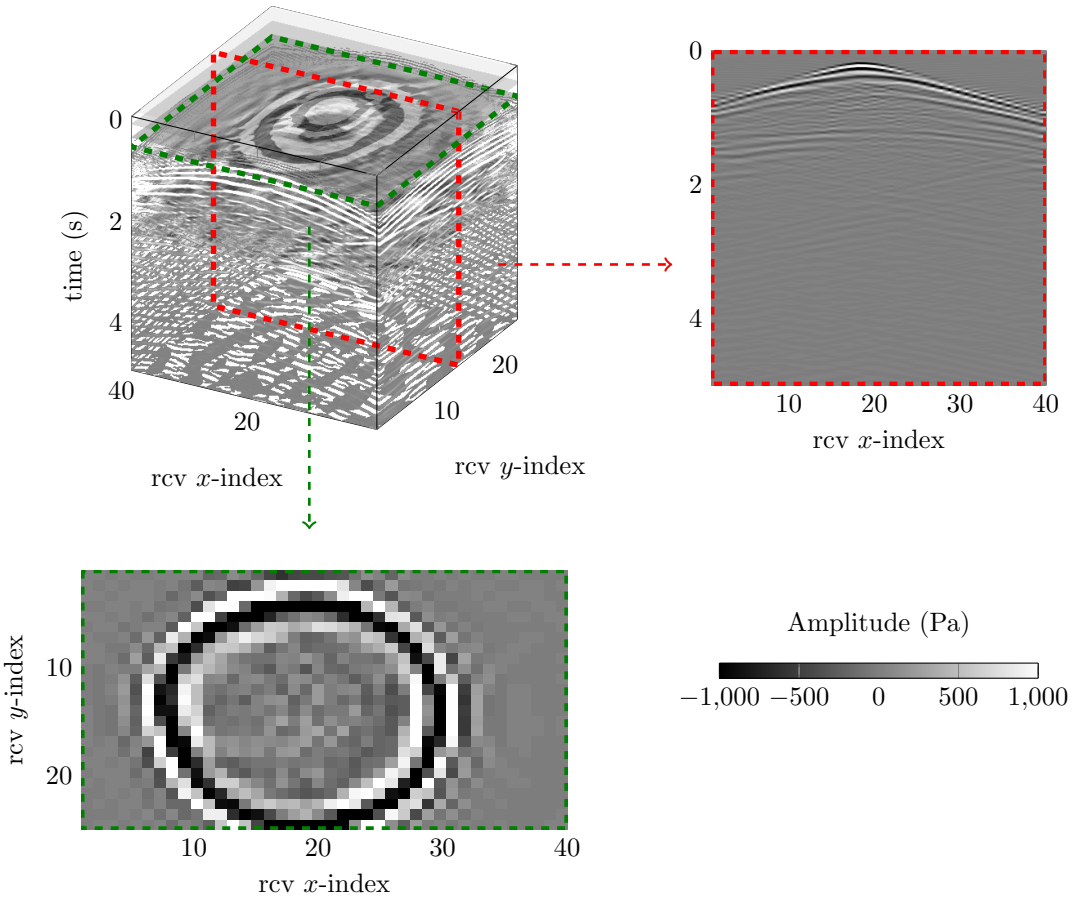


Figure 6.38: Three-dimensional trace for a centrally located source, the 3D visualization is plotted in the upper left figure. A two-dimensional section at a fixed time $t = 0.6$ s is pictured in the bottom left figure, the section is indicated in green in the 3D visualization. The upper right figure selects a fixed line receivers positioned at $y = 855$ m, it is indicated by the red dashed line in the 3D visualization.

domes), we have taken much lower velocity values.

6.6.2 Homogeneous density

We perform a Fourier transform to obtain the data in the frequency domain, after incorporating a level of noise of 20 dB in the time domain trace, following the signal to noise ratio of Equation (6.1). The iterative reconstruction is conducted with sequential frequencies from 2 to 10 Hz with 1 Hz increment. 20 iterations are processed per frequency for a total of 180 iterations. The final reconstruction is presented in Figures 6.40 and 6.41, where we have used finite differences and finite elements for the discretization of the inverse wave problem, respectively.

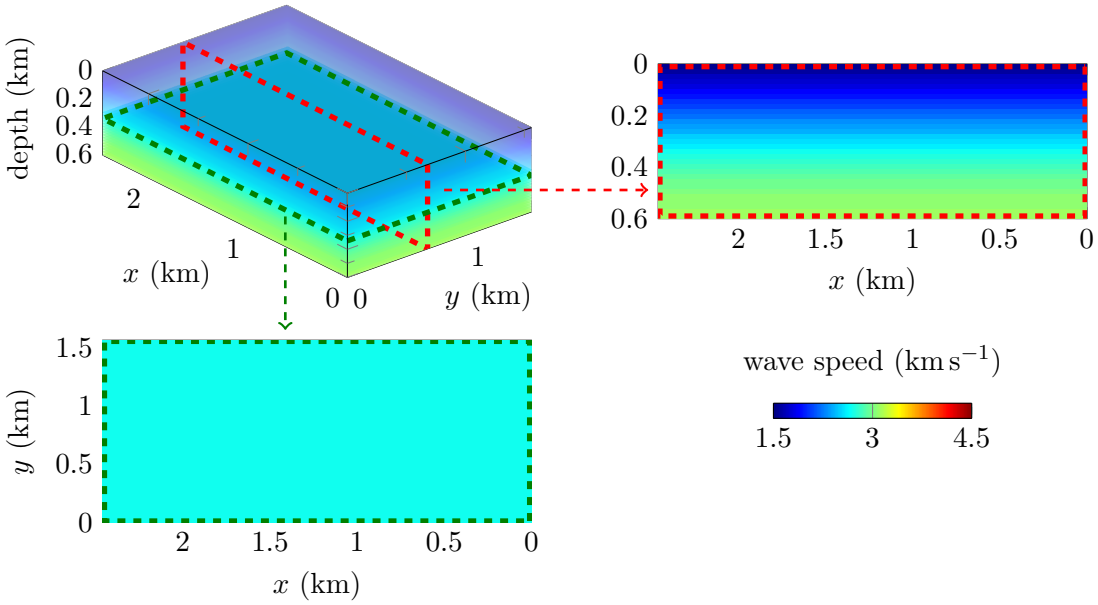


Figure 6.39: Initial wave speed for the iterative minimization algorithm and reconstruction of the Louro model. Vertical section at $y = 680$ m and horizontal section at fixed depth $z = 340$ m are detailed.

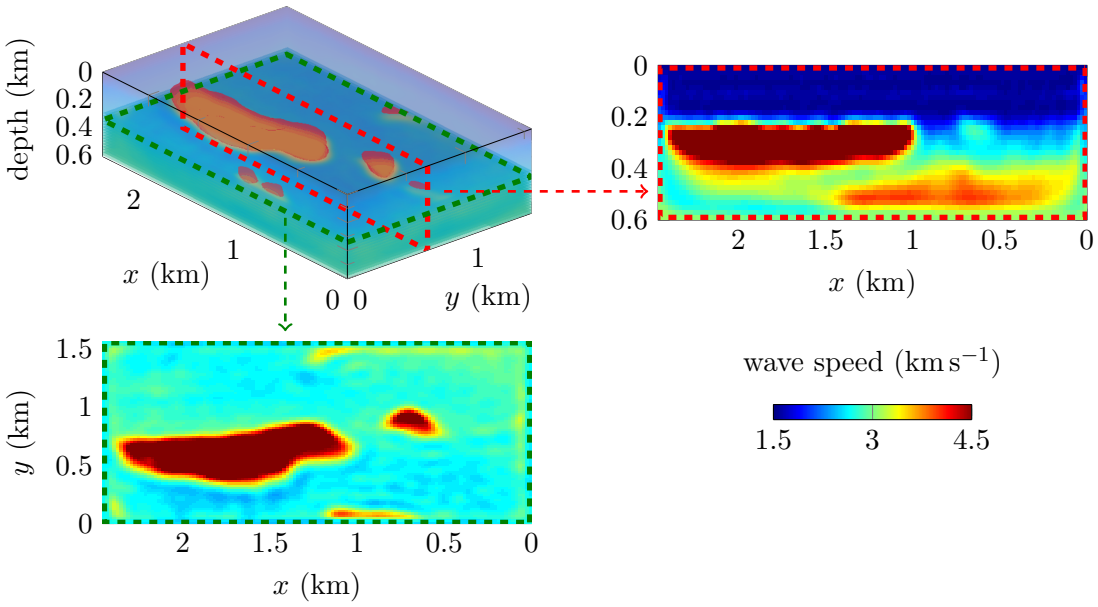


Figure 6.40: Reconstruction of the Louro three-dimensional acoustic model after 10 Hz frequency iterations. Vertical section at $y = 680$ m and horizontal section at fixed depth $z = 340$ m are detailed. The method uses finite differences for the discretization.

The FWI iterative minimization scheme is able to capture the main salt domes for both discretizations. Several bodies appear at their righteous position, with slightly smaller shapes than the original ones. The two central bodies are particularly well recovered, as

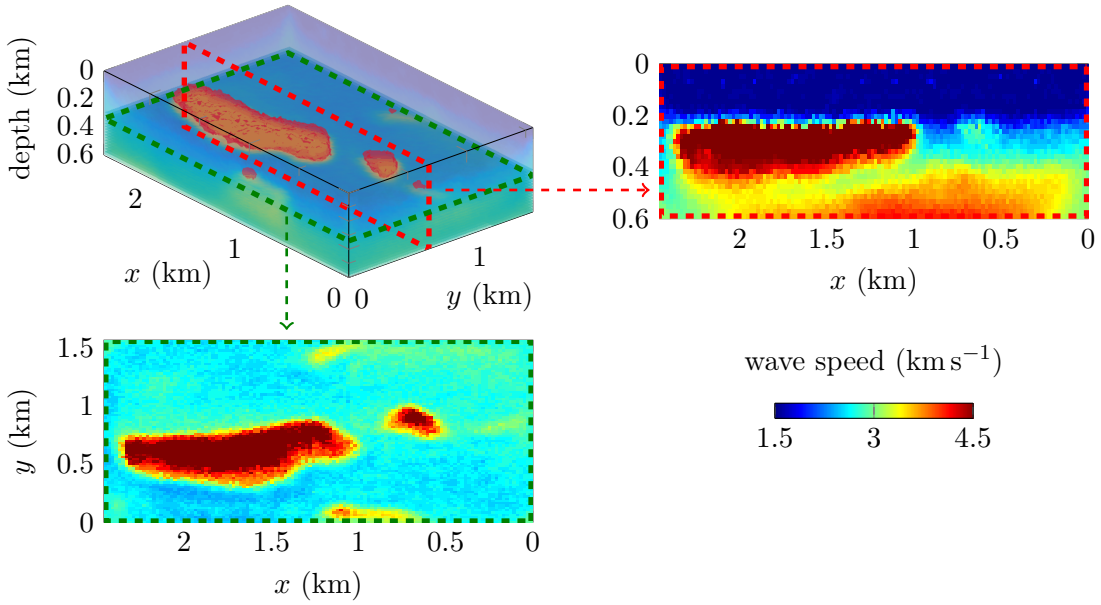
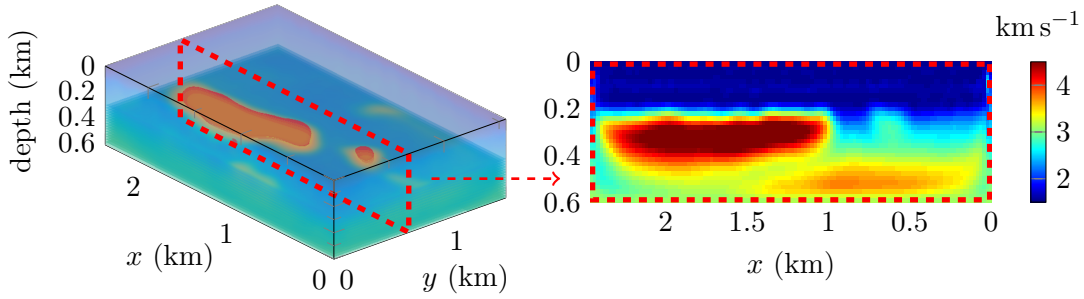


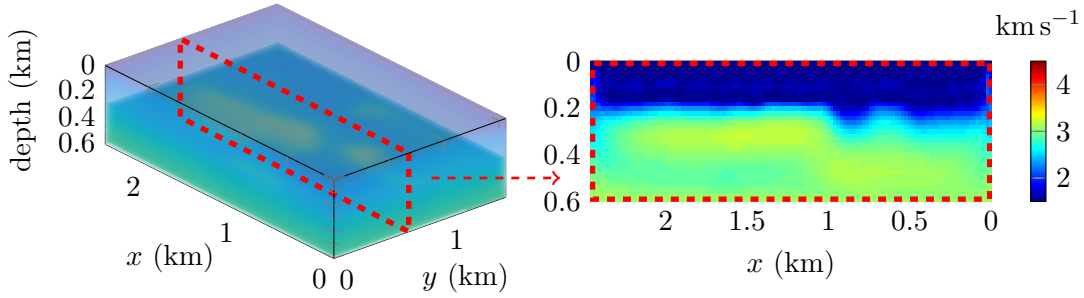
Figure 6.41: Reconstruction of the Louro three-dimensional acoustic model after 10 Hz frequency iterations. Vertical section at $y = 680$ m and horizontal section at fixed depth $z = 340$ m are detailed. The method uses finite elements for the discretization. The mesh is composed of about 300 000 cells, leading to about 430 000 degrees of freedom.

illustrated by the horizontal section. The smallest one of the two is not large enough, as one can barely see it on the vertical section. The salt bodies that are on the sides are harder to recover, due to their limited illumination, yet some are visible on the three-dimensional visualization. The wave speed in the domes is accurately recovered with the value of 4500 $\mathrm{m\,s^{-1}}$. Concerning the background, it is correctly retrieved at the exception of the area under the large salt dome (see the vertical sections). The FE method appears less smooth, which seems to be inherited from the underlying tetrahedral partition. Here the original mesh does not assume any of the bodies. It would be interesting to re-mesh the domain in parallel to the iterative reconstruction, to account for the ongoing salt domes recovery and see if it improves the resolution. This idea of mesh adaptation is part of future development.

Remark 6.3 (Parametrization)**.** *In Chapter 5, Section 5.5, we have illustrated how the parametrization affects the reconstruction for acoustic media. We have shown that the use of the slowness squared* $(1/c^2)$ *is more appropriate, and demonstrated the benefits on the Marmousi experiment. It is the parametrization employed for the current experiments. Here we can compare the effect for this three-dimensional test case, where salt domes are present. In Figure 6.42 we compare the reconstruction depending on the parametrization. The algorithm is exactly the same, only the inverted parameter changes, we compare the inversion of the velocity* (c) *and the slowness* $(1/c)$*. The reference results of Figure 6.40 used the slowness squared* $(1/c^2)$*.*



(a) Using slowness parametrization ($1/c$).



(b) Using velocity parametrization (c).

Figure 6.42: Reconstruction of the Louro three-dimensional acoustic model after 10 Hz frequency iterations depending on the parametrization used when conducting the iterative minimization. Vertical sections at $y = 680$ m are given on the right.

We observe exactly the same behavior as in the Marmousi experiment comparing the parametrization, see Figure 5.37. Namely, the slowness inversion provides a smooth recovery, where the 3D visualization shows some limitation in the shape of the salt domes. The velocity inversion fails in recovering the appropriate wave speed and even the large central dome is barely visible. It confirms the experimental observation we have provided in Section 5.5, and the necessity for slowness squared inversion in geophysical applications.

Remark 6.4. *Despite the presence of salt domes, the algorithm does not require complex frequencies, contrary to the two-dimensional situations (Seam and Sigsbee2A in particular). We believe that the size of the bodies and domain (much smaller than the 2D ones) is the main reason for such behavior. We further illustrate this aspect of body reconstruction with some experiments in Section 6.7.*

Remark 6.5. *This type of experiments, with contrasting objects, are well adapted for mesh deformation. We have mentioned the possibility of simultaneous re-meshing during the iterative procedure. Another possibility is the use of segmentation or level-sets method. Moreover, it would be interesting to investigate how such method can perform when no expectation on the domes is initially taken.*

6.6.3 Incorporating variable density

In our initial experiment, we have assumed a constant density with value 1000 kg m^{-3}. We will now incorporate a variable density, encompassing the similar structures as the wave speed. This density is shown in Figure 6.43. Here the density varies from 1000 to 2500 kg m^{-3}.

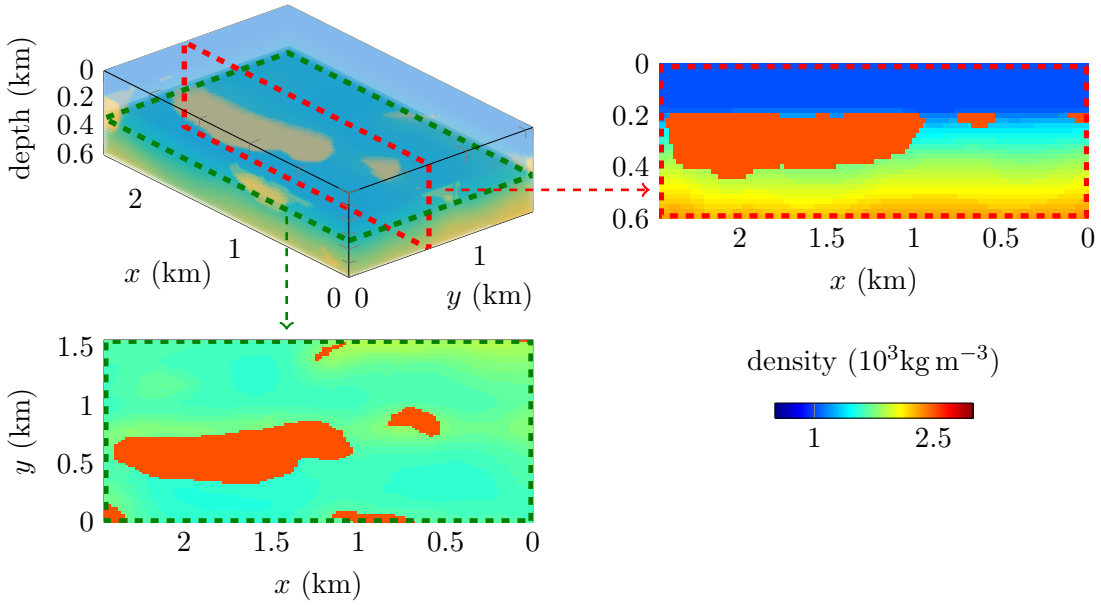
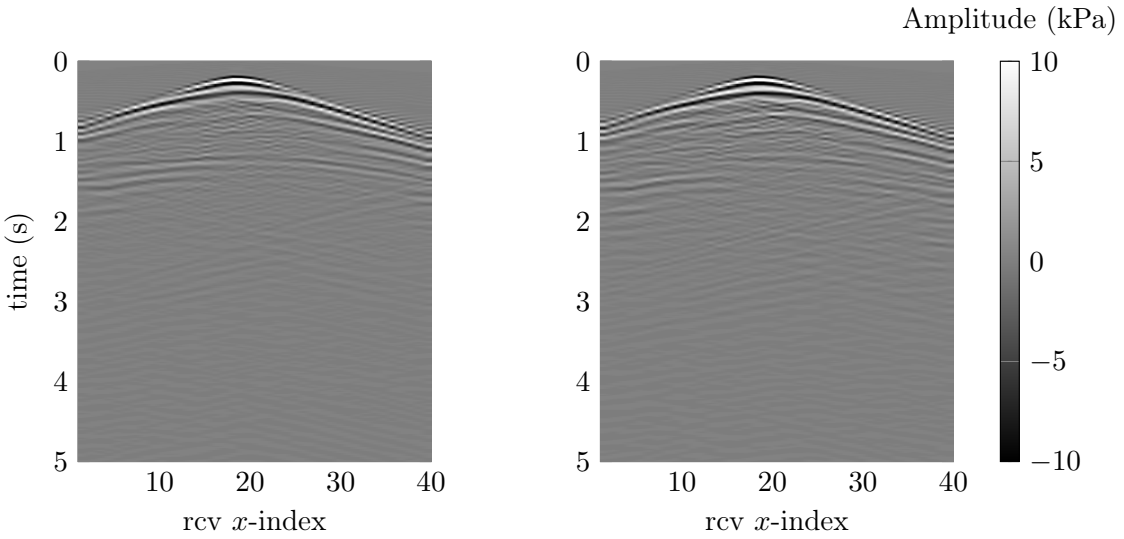


Figure 6.43: Density for the three-dimensional acoustic Louro medium of size $2.46 \times 1.56 \times 0.6$ km. The density varies between 1000 and 2500 kg m^{-3}. Vertical section at $y = 680$ m is indicated with the red dashed frame and plotted in the right. The horizontal section at fixed depth $z = 340$ m is indicated with the green frame and shown at the bottom.

In this context, the wave propagation is dictated by the Equation (1.28). The time domain data are reproduced appropriately. In Figure 6.44, we compare the two-dimensional sections of the seismic trace for a fixed line of receivers. We barely see any difference with the use of variable or constant density.

For the reconstruction we now have two parameters to recover: the bulk modulus and the density. We keep the same starting velocity we pictured in Figure 6.39. For the density we take an initial model without any information, consisting in our usual one-dimensional variation, see Figure 6.45.

For the iterative scheme, we proceed similarly to the constant density situation, but increasing the maximum frequency to 15 Hz, and we use 20 iterations per frequency. Following the analysis proposed in Chapter 5, Section 5.5, we decide not to invert the density, and to keep it as its original representation of Figure 6.45 along with the iterations. In



(a) Section of the seismic trace using constant density.

(b) Section of the seismic trace using variable density.

Figure 6.44: Comparison of two-dimensional sections of the trace for a centrally located source, it corresponds to a fixed line of receivers positioned at $y = 855$ m.

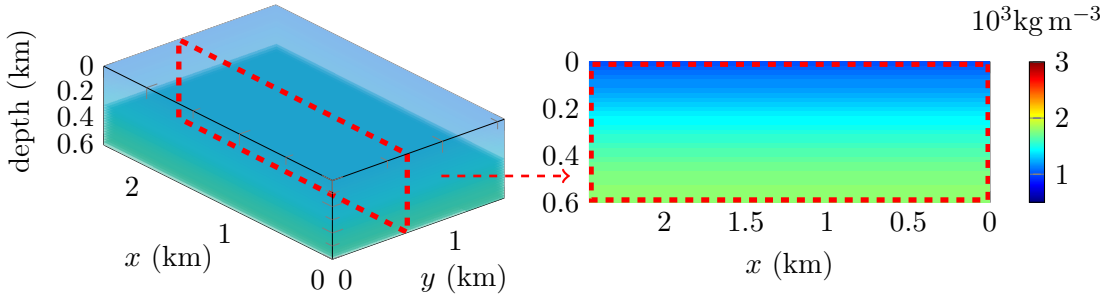


Figure 6.45: Initial density for the iterative minimization algorithm and reconstruction of the Louro model. Vertical section at $y = 680$ m is given on the right.

Figure 6.46, we show the final reconstruction after 15 Hz iterations.

Despite the absence of knowledge on the density, the method is able to recover the main salt domes. The resolution is not as precise as for the experiment with constant density but we still obtain the appropriate wave speed, and some of the smaller domes (see the 3D visualization of Figure 6.46). Clearly the incorporation of a variable, unknown density complicates the reconstruction but the overall results remain correct. This can be explained by the weak sensibility of the data for the density, observed in the comparison proposed in Figure 6.44.

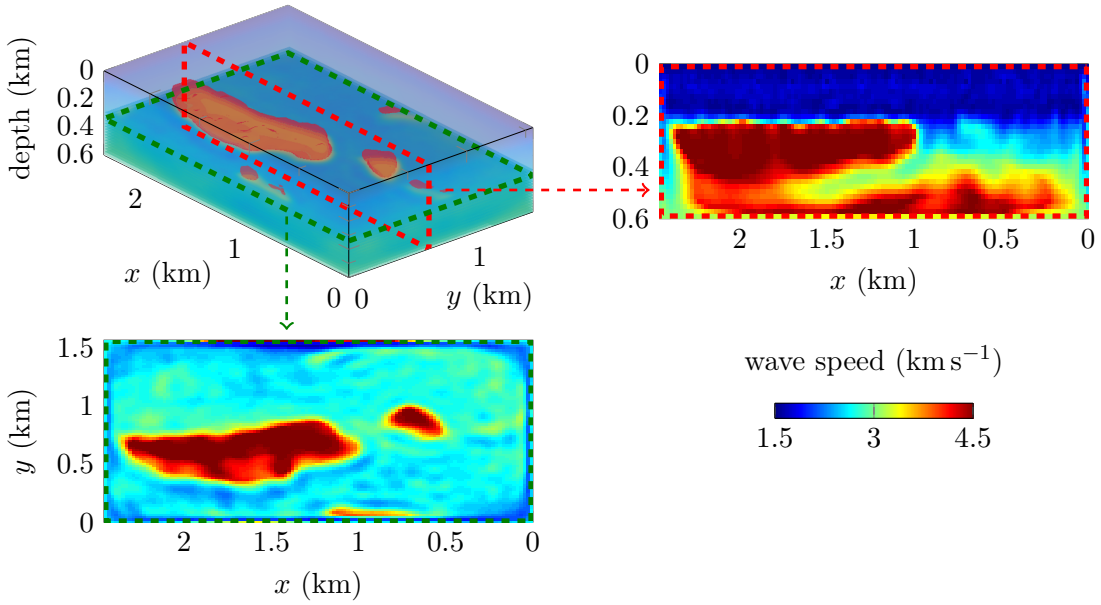


Figure 6.46: Reconstruction of the Louro three-dimensional acoustic model with variable density after 15 Hz frequency iterations. Vertical section at $y = 680$ m and horizontal section at fixed depth $z = 340$ m are detailed. The method uses finite differences for the discretization.

6.7 Perspective 1: geometrical understanding, salt dome reconstruction

The presence of high contrast objects in the subsurface area affects deeply the reconstruction algorithm. As it has been shown in Chapter 5, we know that such geometry generates a decrease of the size of the basin of attraction, and we require the use of lower or complex frequencies to overcome this issue. It is exactly what we observed in the numerical experiments for the reconstruction of Seam and Sigsbee2A models, see Sections 6.5 and 6.4 respectively. However, in the latest three-dimensional salt dome experiment we have processed, see Section 6.6, we did not require any low or complex frequency and the objects appeared appropriately. Following this observation we conjecture that there is an additional parameter that makes the salt reconstruction more difficult. Here, we investigate how the scale of the model affects such recovery.

For this purpose, we design a two-dimensional acoustic test case where salt domes are included. We will play with the size of the model (width and length) and illustrate how it affects the recovery. To remain consistent the number of sources and receivers is adjusted depending on the size of the medium. Yet we always keep the same distance between two consecutive sources: 80 m, similarly we have 40 m between consecutive receivers. Then the number of sources and receivers is automatically given by the size of the domain. The

initial model for the reconstruction consists of the background velocity, which is illustrated in Figure 6.47.

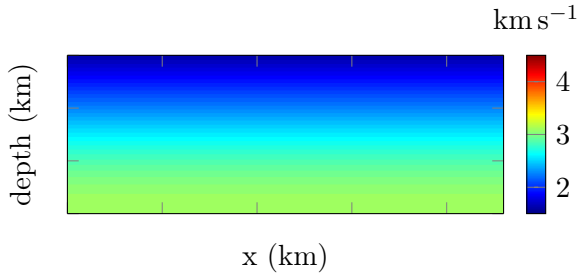


Figure 6.47: 2D initial wave speed for the reconstruction of synthetic salt domes subsurface media. The length and width of the model vary depending on the experiments.

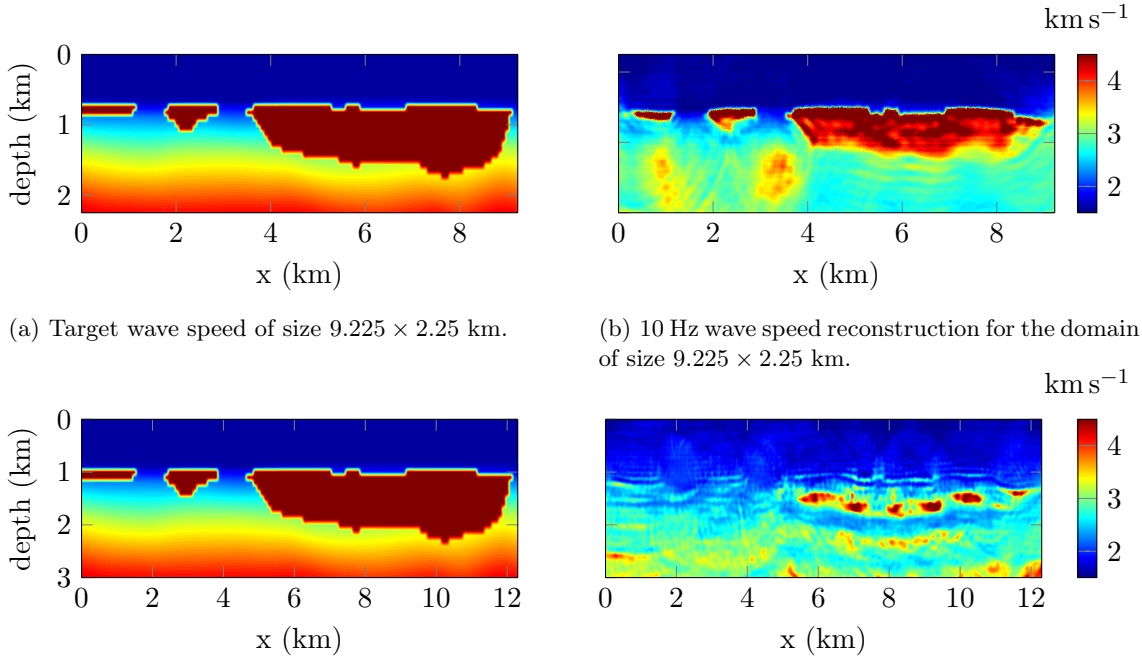
For the reconstruction, synthetic data are generated in the frequency domain (without noise to only study the effect of the salt domes). The iterative minimization is conducted with frequencies from 1 to 10 Hz, with 1 Hz increment and processing 20 iterations per frequency.

6.7.1 First realization

We start with the same targeted velocity, simply playing with the size of the domain of interest. In Figure 6.48 we compare two realizations of the same experiment, where the domain of interest is either 9.225×2.25 km or 12.3×3 km.

We see that the size of the domain of interest plays a fundamental role in the recovery. In the small domain, we see a smooth recovery of the salt and the three bodies appear. Even if the background is inaccurate, the high contrast objects are identified. However, when considering a slightly larger domain, the method fails and no information is obtained.

Remark 6.6 (Low frequency compensation)**.** *In order to overcome the impossibility of retrieving the obstacles, we have successfully used complex frequencies in previous experiments. Similarly, one can use lower frequency to extract the missing information. In the Figure 6.49 we illustrate the recovery for the larger domain (*12.3×3 km*) incorporating data for nine frequencies from* 0.1 *to* 0.9 Hz*. Then the salt is perfectly designed and even the background velocity is well approximated. Nonetheless, such low frequencies are impossible to attain in practice.*



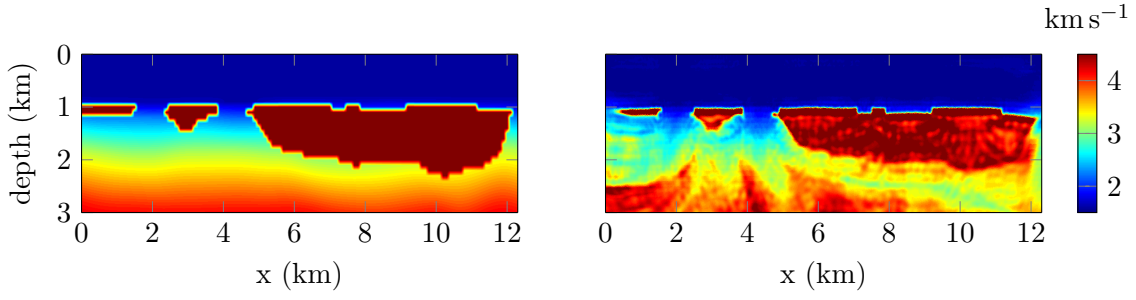
(a) Target wave speed of size 9.225×2.25 km.

(b) 10 Hz wave speed reconstruction for the domain of size 9.225×2.25 km.

(c) Target wave speed of size 12.3×3 km.

(d) 10 Hz wave speed reconstruction for the domain of size 12.3×3 km.

Figure 6.48: 2D reconstruction of wave speed including salt domes with data covering frequencies from 1 to 10 Hz. The models on the left represent the target model (from which the observations are acquired). The test on the top considers the domain with size 9.225×2.25 km; the test on the bottom consider the domain with size 12.3×3 km. The initial model used for the algorithm is given in Figure 6.47.



(a) Target wave speed.

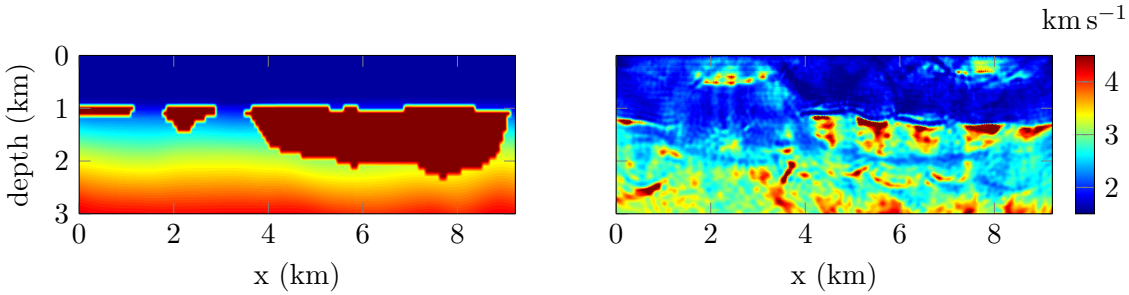
(b) 10 Hz reconstruction.

Figure 6.49: 2D reconstruction of wave speed including salt domes with data covering frequencies from 0.1 to 10 Hz for the domain of size 12.3×3 km. The initial model used for the algorithm is given in Figure 6.47. The algorithm takes full benefits of the low frequency content in the data.

6.7.2 Salt position

When increasing the dimension of the domain, we have increased the dimension of the salt dome, and the depth at which it is positioned. We shall see that the latest has tremendous

importance. In Figure 6.50, we experiment a test case for a domain of size 9.225×3 km for which the larger salt dome is located between about 1000 and 2000 m depth.

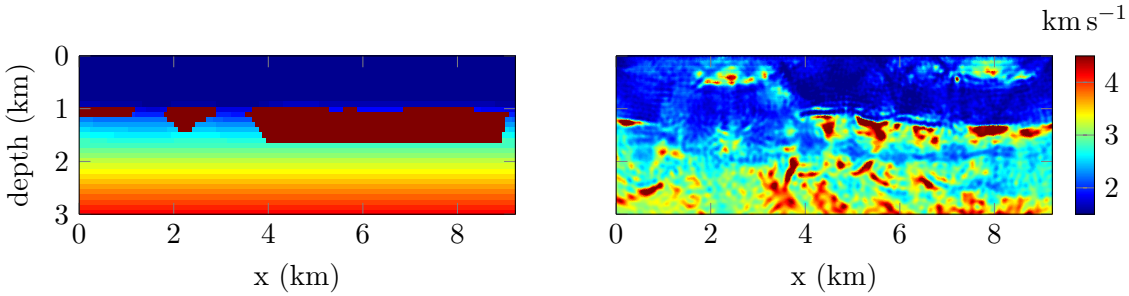


(a) Target wave speed.

(b) 10 Hz reconstruction.

Figure 6.50: 2D reconstruction of wave speed including salt domes with data covering frequencies from 1 to 10 Hz. The models on the left represent the target model (from which the observations are acquired). The domain has size 9.225×3 km. The larger salt dome is approximately located between 1000 and 2000 m depth. The initial model used for the algorithm is given in Figure 6.47.

In this case the algorithm fails to recover the salt. We now reproduce the test by changing the upper and bottom limit of the salt. In Figure 6.51, we show the reconstruction when the bottom of the salt has been moved up to about 1500 m. In Figure 6.52, we show the reconstruction when the upper boundary of the salt has been moved up to about 500 m.



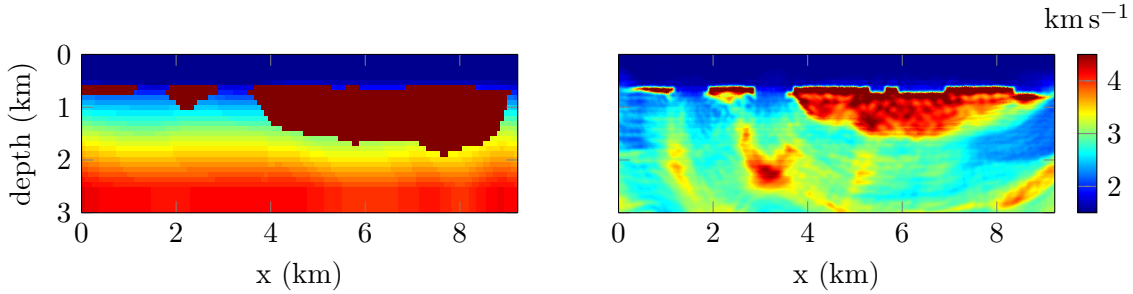
(a) Target wave speed.

(b) 10 Hz reconstruction.

Figure 6.51: 2D reconstruction of wave speed including salt domes with data covering frequencies from 1 to 10 Hz. The models on the left represent the target model (from which the observations are acquired). The domain has size 9.225×3 km. The larger salt is approximately located between 1000 and 1500 m depth. The initial model used for the algorithm is given in Figure 6.47.

Reducing the lower boundary of the salt dome is not sufficient to allow the recovery. When the upper boundary is closer to the surface however, the algorithm is able to detect the object, which starts to appear. The underneath limit is not approximated but we see a clear improvement. We can conjecture that the depth at which the salt is located is crucial, as well as its general dimension.

Remark 6.7. *The relation between salt position, reconstruction and frequency is obviously*



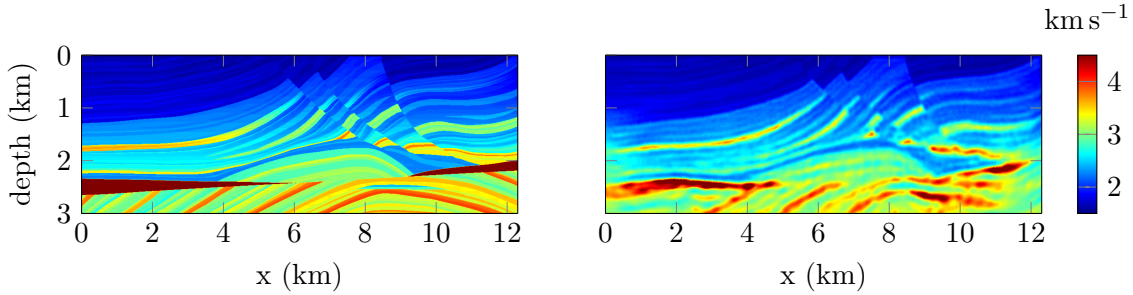
(a) Target wave speed.

(b) 10 Hz reconstruction.

Figure 6.52: 2D reconstruction of wave speed including salt domes with data covering frequencies from 1 to 10 Hz. The models on the left represent the target model (from which the observations are acquired). The domain has size 9.225×3 km. The larger salt dome is approximately located between 500 and 2000 m. The initial model used for the algorithm is given in Figure 6.47.

tightly correlated with the wavelength, which needs to be sufficiently large to allow the recovery. Therefore, reducing the frequency, by increasing the wavelength, is a substitution to identify the salt dome. We expect some kind of relation between the largest wavelength and the location of the upper and bottom interface of the body to allow the reconstruction. For the same reason, if the velocity has higher value, then the shape of the obstacle is easier to retrieve.

Remark 6.8. *In Figure 6.53, we reconstruct the Marmousi model, where the dimensions have been artificially increased to* 12.3×3 km. *Here the algorithm does not suffer from the increase in dimension and the reconstruction is accurate. It completely demonstrates the results of Chapter 5, where the size of the basin of attraction in the direction of one large reflector is much smaller than in the direction of the Marmousi structures. Hence, the Marmousi model is less affected by the increase in dimensions.*



(a) Target wave speed.

(b) 10 Hz reconstruction.

Figure 6.53: 2D reconstruction of the Marmousi wave speed with synthetic data covering frequencies from 1 to 10 Hz. The dimensions are artificially increased to the size 12.3×3 km. The initial model used for the algorithm is given in Figure 6.4.

In this short benchmark, we have exposed the specific complexity of recovering high contrast objects from surface observations. The size and position of the object are fundamental for the recovery. The use of low (or complex) frequencies, by increasing the wavelength, is an alternative but applications cannot usually rely on them, due to the noise. If no prior information is known on the subsurface, there is no hope of anticipating this kind of object and the algorithm necessarily fails. Finally, we believe that this simple experiment should also be compared with the introduction of the Hessian in the procedure, which may provide interesting benefits.

6.8 Perspective 2: reconstruction with attenuation

In Section 1.7, we have explained how attenuation can be incorporated in the formulation of the wave equation. Assuming the Kolsky–Futterman model for attenuation, the velocity becomes complex valued and the attenuation factor Q appears. We can remind the expression for the wavenumber in this context (following Equation (1.31))

$$k(x,\omega) = \frac{\omega}{c(\mathbf{x}) - \dfrac{\mathrm{i}\,\mathrm{sgn}(\omega)c(\mathbf{x})}{2Q(\mathbf{x})}}.$$

Accounting for attenuation is clearly a key in applications of FWI and we investigate how the iterative minimization can perform in this situation. In terms of algorithm, the adjoint state method (see Section 4.4) naturally provides the gradient with respect to Q.

The computation of the gradient requires the derivation of the wave operator with respect to the parameter, see Equation (4.22). Let us illustrate the technicality with an acoustic medium with constant density, where the waves follow the Helmholtz equation. The velocity is taken with an imaginary part and the derivation of the wave operator with respect to c and Q (following the notation of the Kolsky–Futerman model Equation (1.31)) gives

$$\begin{cases} \dfrac{\partial \mathcal{P}}{\partial c} = -\dfrac{2\omega^2}{c^3\left(1 - \dfrac{\mathrm{i}\,\mathrm{sgn}(\omega)}{2Q}\right)^2}, \\ \dfrac{\partial \mathcal{P}}{\partial Q} = -\dfrac{\mathrm{i}\,\mathrm{sgn}(\omega)\omega^2}{c^2Q^2\left(1 - \dfrac{\mathrm{i}\,\mathrm{sgn}(\omega)}{2Q}\right)^3}. \end{cases}$$

It shows that when accounting for attenuation, the gradient with respect to the wave speed must be appropriately adjusted as well.

Here we design a three-dimensional toy example to study the behavior of the attenuation factor reconstruction. We basically take the background velocity of the Louro model, introduced in Figure 6.37, and design a cross-shaped obstacle. The model is of size $2.46 \times 1.56 \times 0.6$ km, the wave speed and attenuation factor are pictured in Figures 6.54 and 6.55 respectively. The background consists in a constant attenuation factor and an increasing velocity with depth. Inside the cross-shaped body, the wave speed and attenuation are constant with values 4000 $\mathrm{m\,s^{-1}}$ and $Q = 80$.

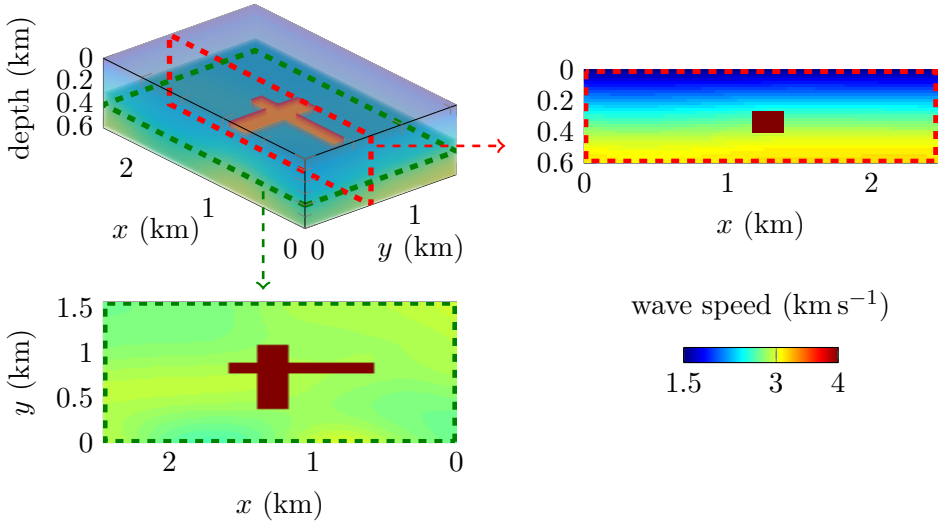


Figure 6.54: Wave speed for the three-dimensional experiment including attenuation of size $2.46 \times 1.56 \times 0.6$ km. The velocity inside the body is constant with value 4000 $\mathrm{m\,s^{-1}}$. Vertical section at $y = 680$ m is indicated with the red dashed frame and plotted in the right. The horizontal section at fixed depth $z = 400$ m is indicated with the green frame and shown at the bottom.

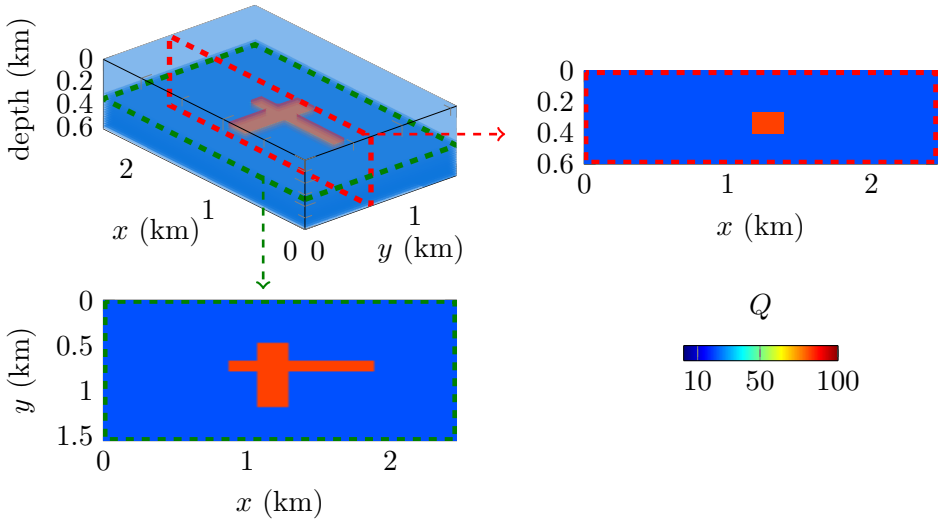


Figure 6.55: Quality factor for three-dimensional experiment. Vertical section at $y = 680$ m is indicated with the red dashed frame and plotted in the right. The horizontal section at fixed depth $z = 400$ m is indicated with the green frame and shown at the bottom.

For the reconstruction we consider synthetic data where the initial models match with the backgrounds. Hence the initial wave speed is similar as the Louro experiment, see Figure 6.39; the initial attenuation factor is constant with value 20 (value of the background in Figure 6.55). Hence we simply try to retrieve the subsurface body. We perform 20 iterations using sequential frequencies from 2 to 10 Hz, using 1 Hz step. Synthetic data have been generated prior to the reconstruction. In Figures 6.56 and 6.57 we show the reconstructed wave speed and attenuation factor.

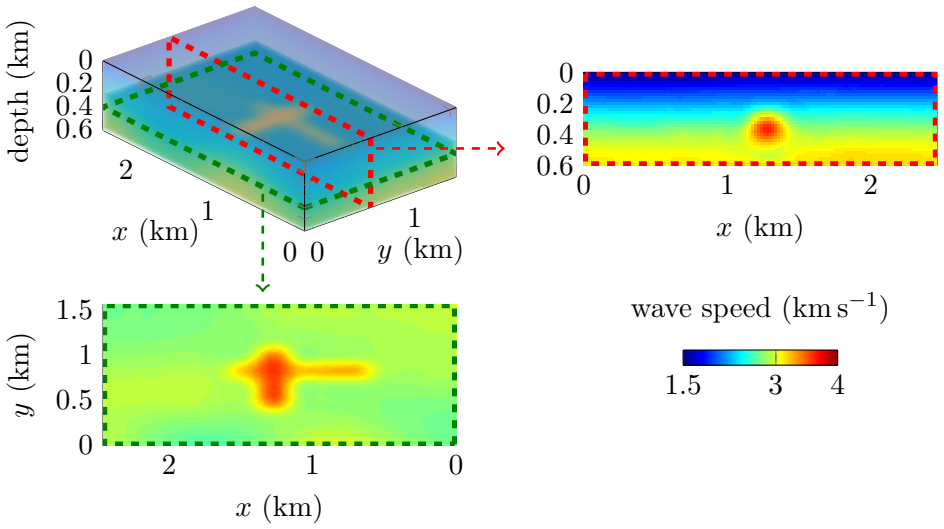


Figure 6.56: Reconstructed wave speed of the three-dimensional acoustic model with attenuation and the subsurface cross-shaped body. Vertical section at $y = 680$ m and horizontal section at fixed depth $z = 400$ m are detailed.

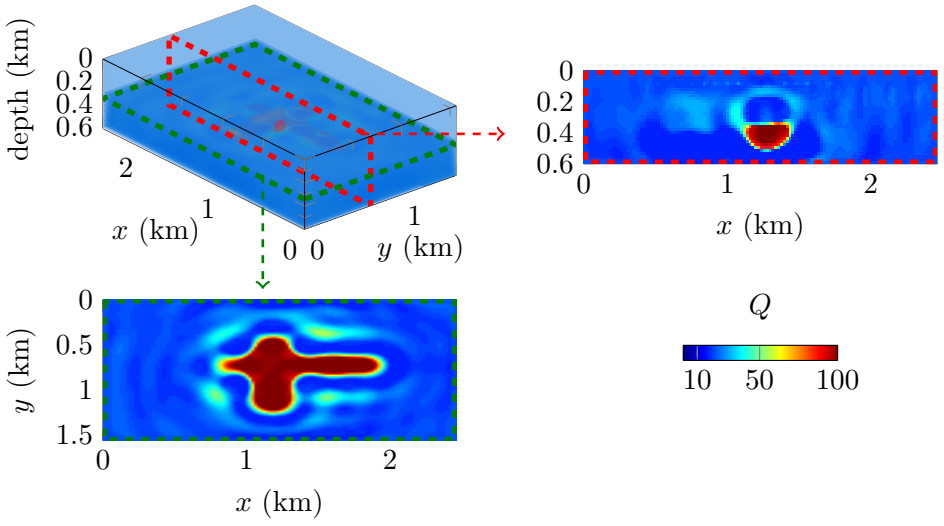


Figure 6.57: Reconstructed attenuation factor of the three-dimensional acoustic model with attenuation and the subsurface cross-shaped body. Vertical section at $y = 680$ m and horizontal section at fixed depth $z = 400$ m are detailed.

The subsurface body appears in the recovered wave speed and attenuation factor. The

position is well approximated for the velocity, but the values are noticeable lower than expected. For the attenuation factor, we observe some circular artifacts appearing around the shape. The value is well approximated but it seems that the position of the cross is underneath the actual one (it is particularly visible on the vertical section of Figure 6.57).

In this preliminary experiment, we have shown how the recovery of attenuation is possible in the algorithm. We have obtained interesting results in terms of reconstruction but this is a synthetic experiment. It has to be enriched with more realistic ingredients. In particular, here, both forward and inverse problems account for the attenuation with the Kolsky–Futterman model. We need to incorporate consequent model error to test the reliability of the method, and the use of time domain noisy data. Hence we see this test as an initial step towards more challenging experiments.

6.9 Elastic 2D Marmousi2 model

The elastic Marmousi2, introduced by Martin et al. (2006) is an extension of the acoustic Marmousi model. It has already been thoroughly employed in Chapter 5, and numerical reconstructions have been carried out with a comparison on the elastic parametrization in Section 5.6. The elastic coefficients, P-velocity, S-velocity and density have been presented in Figure 5.44, the starting models are in Figure 5.14. This model is of size 17 km by 3.5 km.

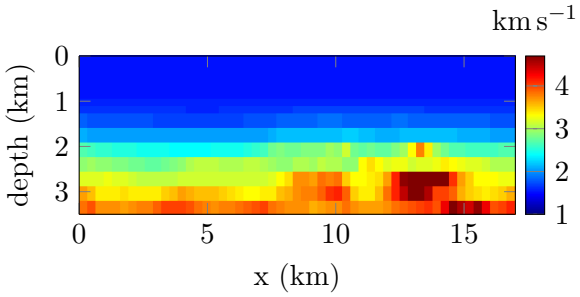
In this section, we only reproduce the experiment incorporating piecewise constant domain partition in parallel with the iterations. We do not consider surface waves for simplicity and follow the same frequency progression as in Section 5.6. In Table 6.7 we provide the selected relation between frequency and model representation.

frequency iterations	**model representation (number of unknowns)**
1 Hz	530
2 Hz	2 226
3–4 Hz	9 116
5–6 Hz	36 975
7–8 Hz	148 750

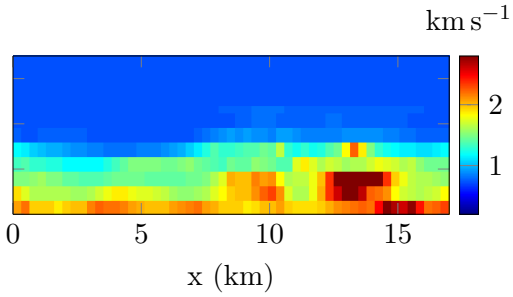
Table 6.7: Relation between frequency and scale selected for the multi-level reconstruction of the elastic Marmousi2 model. No partition is employed for the last frequencies, 9 and 10 Hz.

During the iterative minimization, we do not reconstruct the density, which is kept as its initial representation, see Figure 5.14. The progression of frequency and model partitioning

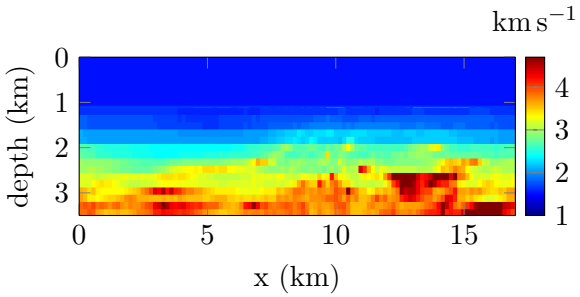
leads to a step by step reconstruction, starting with coarse representation where only rough recovery can be expected. In Figure 6.58, we show the results after 1 and 2 Hz iterations, to illustrate the scale progression.



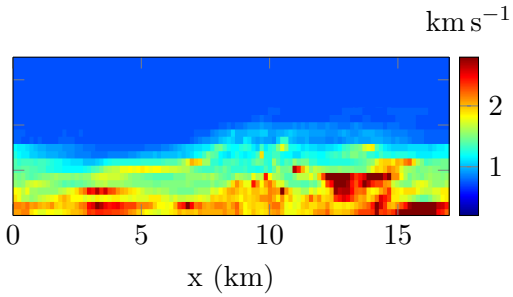
(a) P-wave speed after 1 Hz iterations.



(b) S-wave speed after 1 Hz iterations.



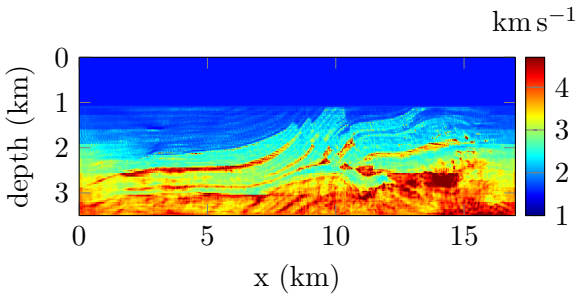
(c) P-wave speed after 2 Hz iterations.



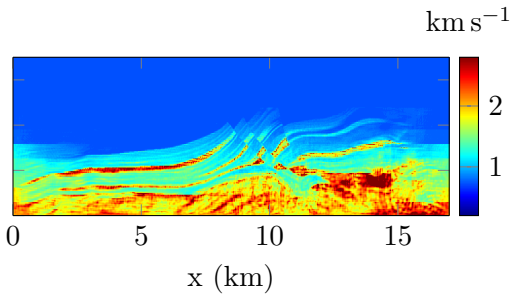
(d) S-wave speed after 2 Hz iterations.

Figure 6.58: Progressive reconstruction of the elastic Marmousi2 model, where scale and frequency are paired according to Table 6.7. The density is kept fixed along the iterations as its initial representation. In this experiment surface waves are not considered.

The evolution of scale provides the progressive recovery of the structures. The initial frequencies, where the number of model coefficients remains low is still able to anticipate the variation. We show the results after the final 10 Hz iterations in Figure 6.59.



(a) P-wave speed after 10 Hz iterations.



(b) S-wave speed after 10 Hz iterations.

Figure 6.59: Reconstruction of the elastic Marmousi2 model after 10 Hz iterations.

We see that the multi-level reconstruction eventually gives the correct structures for both velocities with similar accuracy. Compared to our original experiment without model partition, the deepest areas of the velocities are better recovered, in particular regarding the amplitude of the velocity.

6.10 Elastic 2D Pluto model

In this section, we investigate the reconstruction of an elastic medium encompassing salt domes. We compare the consideration, or not, of surface waves.

6.10.1 Acquisition setup and data

Let us introduce the elastic Pluto model, which is a two-dimensional medium composed of salt domes. It is a medium of size 31.16×7.08 km, represented with the P-wave speed, S-wave speed and density pictured in Figures 6.60, 6.61 and 6.62 respectively. It is composed of three separate bodies, the largest is positioned in the center while two smaller ones are located on the sides.

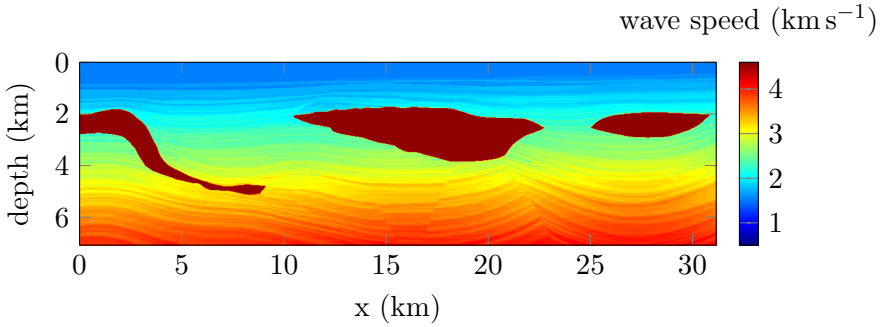


Figure 6.60: P-wave speed for the elastic Pluto medium of size 31.16×7.08 km, the velocity varies between 1500 and 4500 m s^{-1}.

In this experiment, the seismic acquisition is composed of 150 sources and 310 receivers for every source. We generate synthetic data in the frequency domain using different order and meshes for the method of discretization. As usual we start from initial models where no information are assumed on any of the domes. We take smooth velocity backgrounds, illustrated in Figure 6.63.

We do not consider the density for the reconstruction, and it is kept as its original representation of Figure 6.63. We try to confirm the demonstration we have initiated that inaccurate density does not prevent from the recovery of the other coefficients. Because of

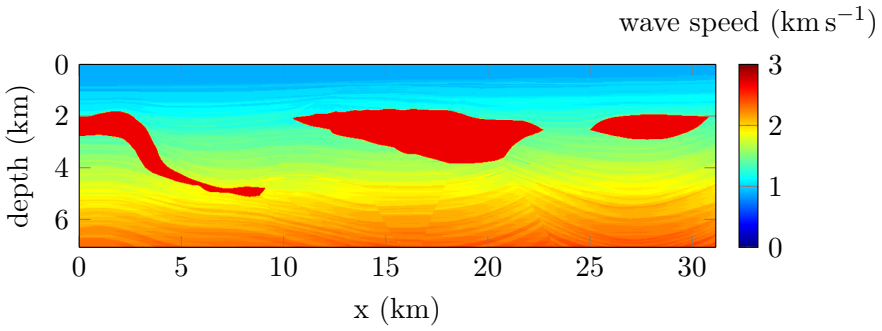


Figure 6.61: S-wave speed for the elastic Pluto medium of size 31.16×7.08 km, the velocity varies between 880 and 2700 m s^{-1}.

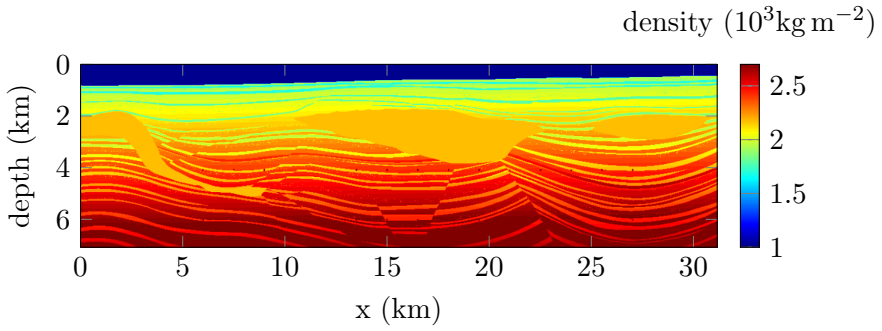


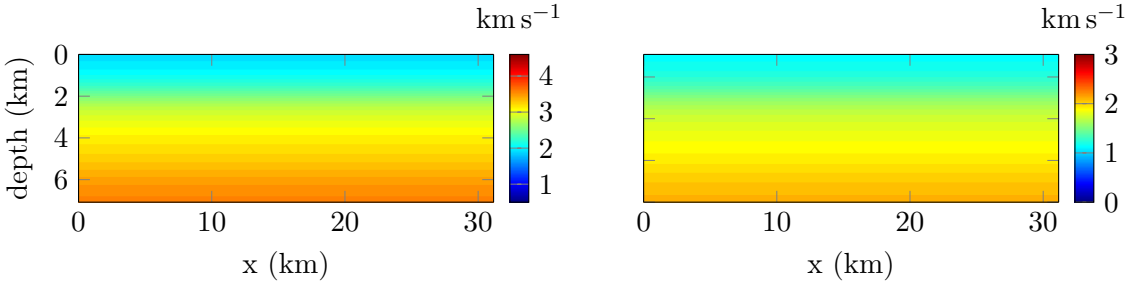
Figure 6.62: Density for the elastic Pluto medium of size 31.16×7.08 km, it varies between 1000 and 3000 kg m^{-2}.

the salt domes, we can anticipate that the use of complex (or very low) frequencies will be required, following the expertise carried out in Chapter 5 and the acoustic experiments of Sections 6.4, 6.5 and 6.7.

6.10.2 Ignoring surface waves

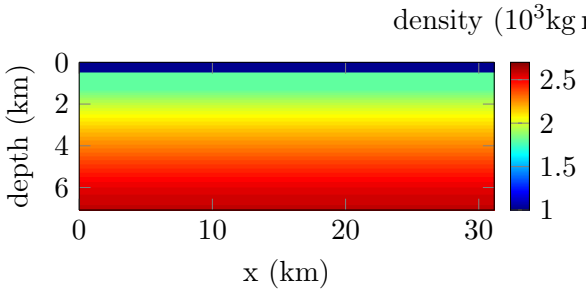
Let us start by neglecting surface waves, for a simpler framework, as illustrated in Section 5.6 with the Marmousi2 reconstruction, and mentioned in Gélis et al. (2007); Yuan et al. (2015). It means that both the synthetic data generated and the inverse algorithm do not account for them. We select sequential frequencies from 1 to 10 Hz, with 1 Hz step and perform 20 iterations per frequency. In Figure 6.64, we show the reconstructed P- and S-wave speeds.

As expected, the algorithm fails to recover the subsurface information, due to the high velocity domes. It confirms the acoustic situation where high contrast objects have required the use of complex frequencies when no information is initially known on the subsurface area. Therefore, we reproduce the experiments starting with complex frequencies. The



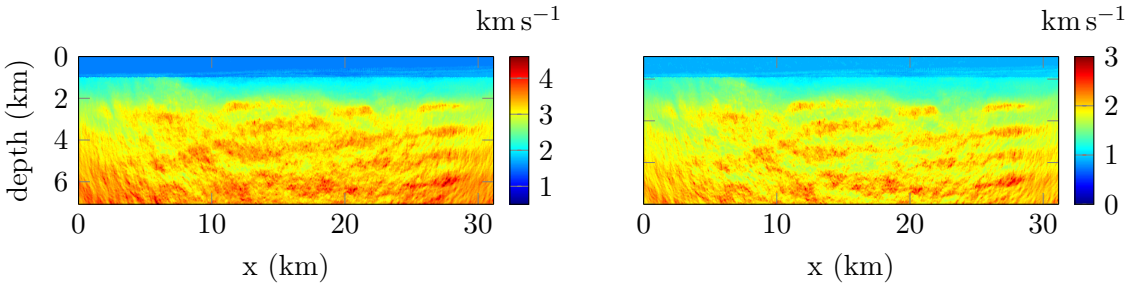
(a) Initial P-wave speed.

(b) Initial S-wave speed.



(c) Initial density.

Figure 6.63: Initial models for the reconstruction of the elastic Pluto medium of size 31.16×7.08 km. All starting models use a one-dimensional profile which only varies with depth.



(a) Reconstructed P-wave speed.

(b) Reconstructed S-wave speed.

Figure 6.64: Reconstructions of the elastic Pluto models using frequencies from 1 to 10 Hz. 20 iterations are performed per frequency (for a total of 200 iterations).

complete set of frequencies is given in Table 6.8, 20 iterations are performed per frequency.

We also employ the multi-level algorithm so that frequency and scale progress simultaneously, starting with coarse model representation. In Table 6.9, we give the relation between the set of frequency and the number of coefficients to represent the parameter. In this experiment, the mesh of the domain is generated with about 140 000 cells, without accounting for any of the domes shapes.

Set of iterations	**damping parameter (σ)**	**Fourier frequency (f)**
1 to 4	2,5,7,10	0 Hz
5 to 7	7,5,2	1 Hz
8 to 10	7,5,2	2 Hz
11 to 20	0	1 to 10 Hz

Table 6.8: Set of complex frequency employed for the reconstruction of the elastic Pluto model. Every set is composed of one single frequency defined by $-\omega^2 = (\sigma + 2i\pi f)^2$. For instance the very first frequency is only composed of the damping with $\sigma = 2$; the seventh set is $(\sigma = 2, f = 1\text{Hz})$. 20 iterations of the minimization algorithm are performed per set, for a total 400 iterations.

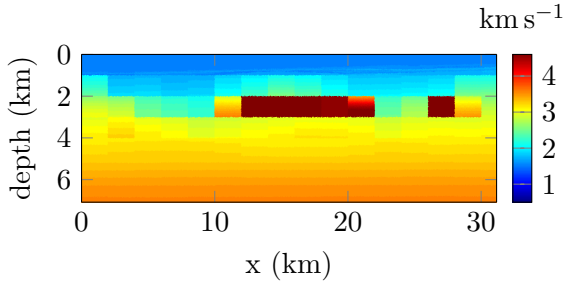
frequency set	**model representation (number of unknowns)**	**percentage compared to the original scale**
1 to 4	75	0.05%
5 to 7	160	0.12%
8 to 10	234	0.20%
11	384	0.30%
12 and 13	702	0.50%
14 and 15	3 744	3.00%
16 to 20	11 232	8.00%

Table 6.9: Relation between frequency set and scale selected for the reconstruction of the acoustic Pluto model without accounting for surface waves. The corresponding frequencies in the set are given from Table 6.8.

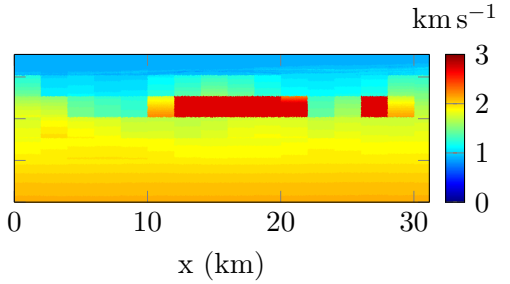
In Figure 6.65, we picture the reconstruction at different steps of the complex frequency sets. Figures 6.65(a) and 6.65(b) show the reconstruction of the velocities after the iterations where only the damping is used (Laplace domain reconstruction, after the fourth set of Table 6.8). Figures 6.65(c) and 6.65(d) show the final reconstruction after all complex frequencies have been processed (after the tenth set of Table 6.8).

We see that the objects appear smoothly. The complex frequencies have increased the size of the radius of convergence, and allow an initial recovery of the parameters. The reduction of the number of coefficients provides a coarse reconstruction, which is actually well adapted to the recovery of subsurface domes. We pursue the iterative minimization, incorporating increasing frequency parameters f, see Table 6.8. The final reconstructions of the P- and S-wave speeds, after the maximum 10 Hz iterations, are pictured in Figure 6.66.

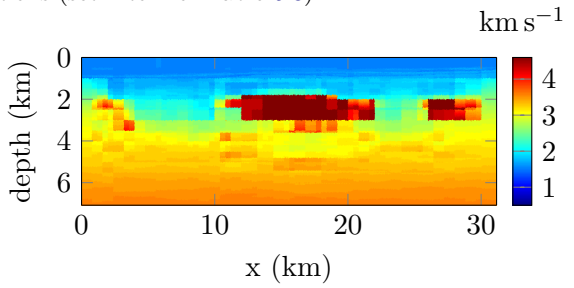
The three subsurface salt domes are captured, benefiting from the complex frequencies initial iterations. The three objects are appropriately designed and the value of velocities is well approximated, both for the P- and S-wave speeds. We observe that the overall shapes of the objects are slightly smaller than the actual ones, the underneath interface being harder



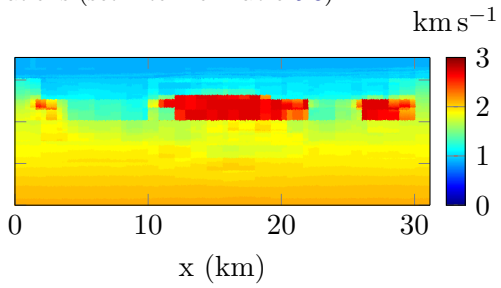
(a) Reconstructed P-wave speed after damping iterations (set 1 to 4 of Table 6.8).



(b) Reconstructed S-wave speed after damping iterations (set 1 to 4 of Table 6.8).

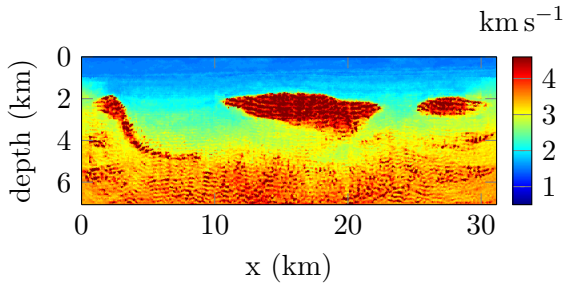


(c) Reconstructed P-wave speed after complex frequencies iterations (set 1 to 10 of Table 6.8).

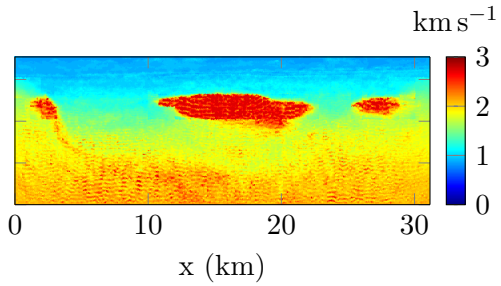


(d) Reconstructed S-wave speed after complex frequencies iterations (set 1 to 10 of Table 6.8).

Figure 6.65: Reconstructions of the elastic Pluto models after complex frequencies have been used (set 1 to 10 in Table 6.8).



(a) Reconstructed P-wave speed.



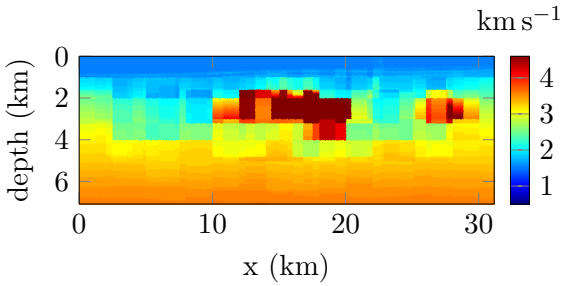
(b) Reconstructed S-wave speed.

Figure 6.66: Final reconstructions of the elastic Pluto models after 10 Hz iterations, using the progression of frequency and scale prescribed in Table 6.8).

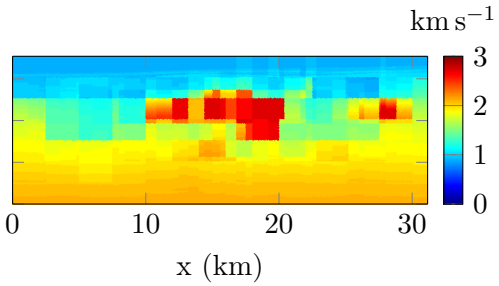
to obtain. It is certain that the need for complex frequencies may be complicated to fulfill when noise perturbs the data. An alternative is to start from an initial guess including a priori knowledge. We also note that keeping the inaccurate density does not prevent from the recovery. Regarding the compression, we have largely decreased the number of coefficients to represent the model (see Table 6.9) but the resolution does not suffer from it and remains accurate.

6.10.3 Incorporating surface waves

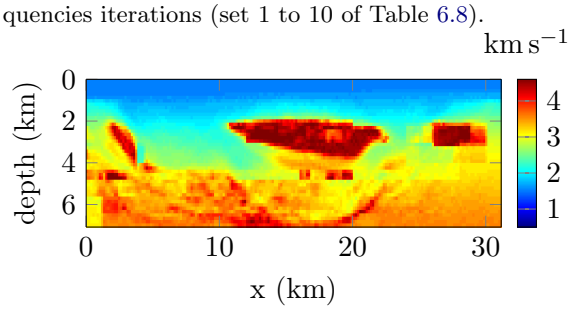
We reproduce the experiment but considering the surface waves. We follow the same set of complex frequencies given in Table 6.8 and we still use 20 iterations per frequency. In Figure 6.67, we picture the reconstruction of the P- and S-wave speeds. We show the reconstruction after the complex frequencies have been processed, and the final reconstruction.



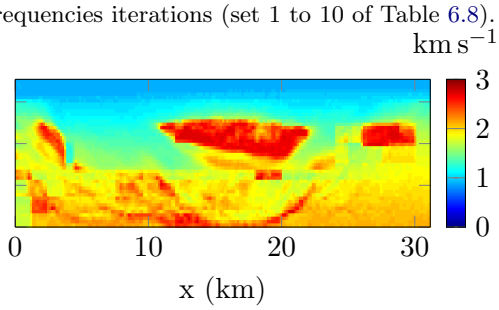
(a) Reconstructed P-wave speed after complex frequencies iterations (set 1 to 10 of Table 6.8).



(b) Reconstructed S-wave speed after complex frequencies iterations (set 1 to 10 of Table 6.8).



(c) Reconstructed P-wave speed after the final 10 Hz iterations (all sets of Table 6.8 are processed).



(d) Reconstructed S-wave speed after the final 10 Hz iterations (all sets of Table 6.8 are processed).

Figure 6.67: Reconstructions of the elastic Pluto models with surface waves. The progression of frequency and scale is prescribed in Table 6.8.

Contrary to the Marmousi2 experiment (see Section 5.6), the incorporation of surface waves does not affect the recovery and we still retrieve the salt domes. In particular, the complex frequency sets seem to provide a larger dome compared to when we neglect the surface waves. Yet the final reconstruction shows low resolution regarding the deepest structures, more precisely, one can see that the dome on the left is less accurate. We can conjecture that it is strongly related with the pattern of the subsurface structures. Indeed, having high contrast object generates few, reflections back to the surface and having several layers (as in the Marmousi2 medium) induces much more reflections.

6.11 Elastic 3D Epati model

6.11.1 Acquisition setup and data

We now study a three-dimensional elastic medium. With such media, we attain the current computational limitations of the frequency domain approach, imposed by the numerical cost of the matrix factorization. Here, more than the efficiency of the recovery, we aim at illustrating that our software is able to handle large scale computations, as an initial step towards even larger domains. We study the Epati model (courtesy of Total), which is a model of size 1.8 km by 1.4 km by 1.2 km. It is a relatively small model in geophysics, but already challenging for the frequency approach. The corresponding P-wave speed, S-wave speed and density are presented in Figures 6.68, 6.69 and 6.70 respectively where we show a three-dimensional visualization and a two-dimensional section for fixed $y = 700$ m. The resolution of the wave equation is conducted with about ten millions degrees of freedom. Note that we do not change the discretization during the frequency progression, even if it would be easily imaginable to employ a coarser discretization for low frequency (hence reducing the computational cost consequently).

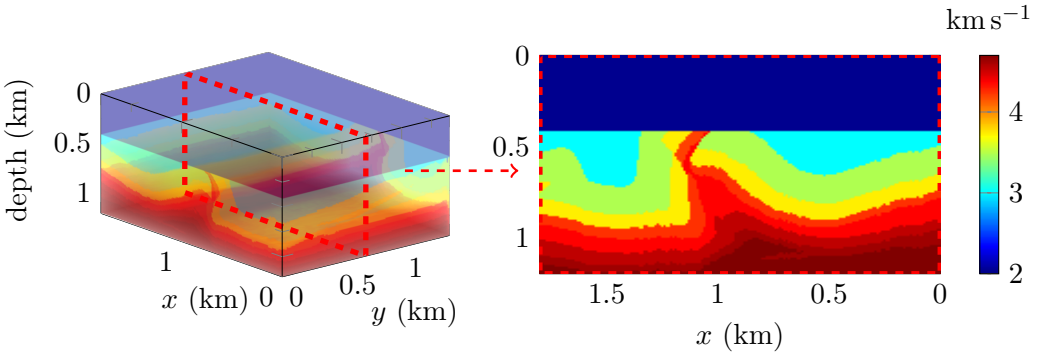


Figure 6.68: P-wave speed for the three-dimensional elastic Epati medium of size $1.8 \times 1.4 \times 1.2$ km (courtesy of Total). The velocity varies between 2000 and 4700 m s^{-1}. Vertical section at $y = 700$ m is indicated with the red dashed frame and pictured in the right.

It is interesting to notice that, in addition to difficulties due to the large scale, the P- and S-wave speeds have very different profiles. The P-wave speed presents a cone of increasing velocity while the S-wave speed has a structure with layers of constant velocity values. This is a major difference compared to the previous elastic experiments where the P- and S-wave speeds have similar patterns (see Marmousi2 and Pluto experiments of Sections 6.9 and 6.10).

The seismic acquisition is conducted with 176 sources positioned in an two-dimensional

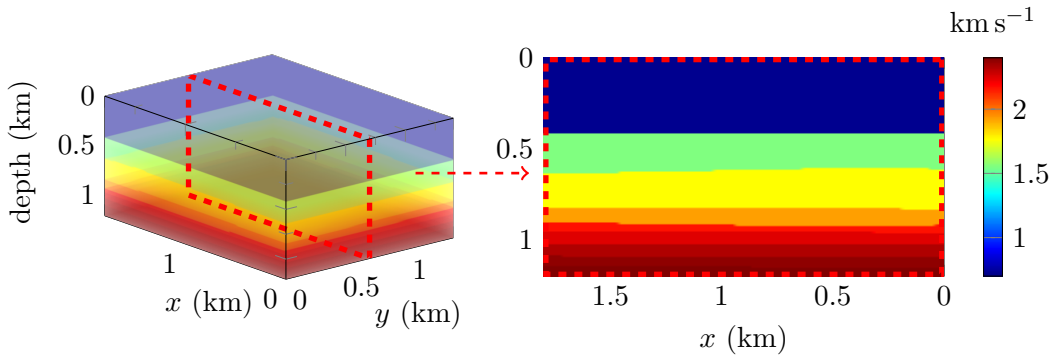


Figure 6.69: S-wave speed for the three-dimensional elastic Epati medium of size $1.8 \times 1.4 \times 1.2$ km (courtesy of Total). The velocity varies between 700 and 2400 m s^{-1}. Vertical section at $y = 700$ m is indicated with the red dashed frame and pictured in the right.

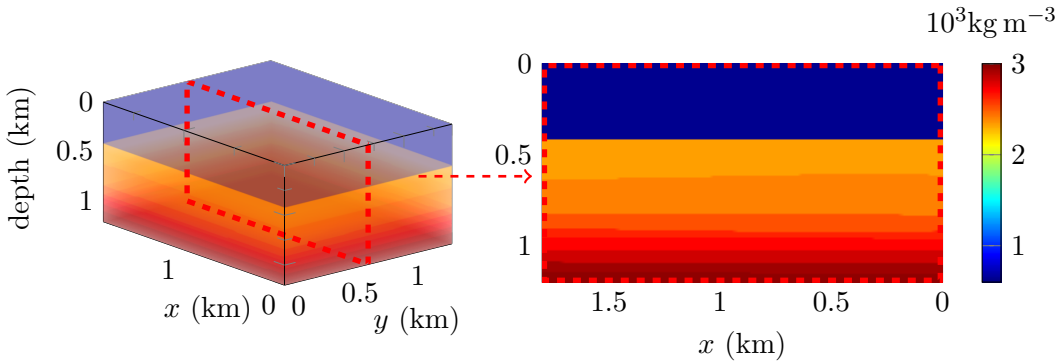


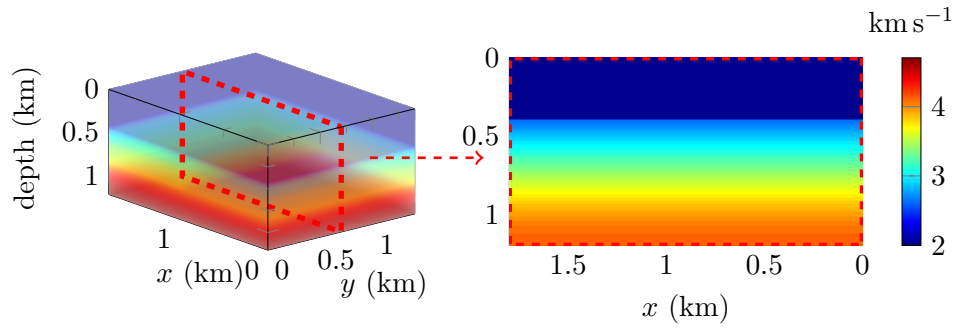
Figure 6.70: Density for the three-dimensional elastic Epati medium of size $1.8 \times 1.4 \times 1.2$ km (courtesy of Total). The velocity varies between 600 and 3000 kg m^{-3}. Vertical section at $y = 700$ m is indicated with the red dashed frame and pictured in the right.

lattice. Every source is separated by 80 m on both the x and y axis. Similarly, 936 fixed receivers are incorporated with 40 m distance in both directions.

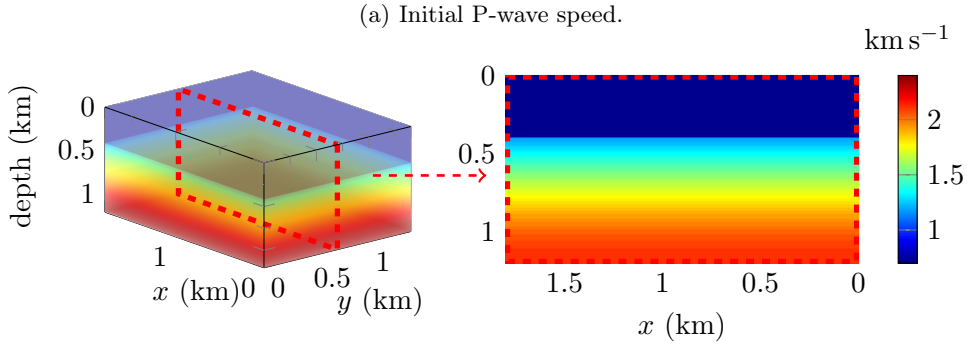
6.11.2 Reconstruction using synthetic data

Synthetic data are generated directly in the frequency domain, and we incorporate noise in the data. For the reconstruction, the initial models have no information on the subsurface structures and correspond with depth varying profiles, see Figure 6.71.

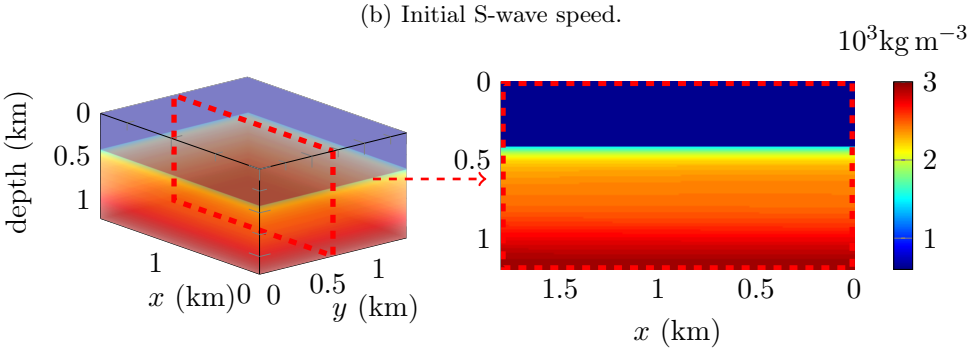
Following the previous elastic experiments, we do not account for the density reconstruction and utilizes the multi-level algorithm to associate frequency and scale progression (with a piecewise constant representation of the model). The available frequencies cover a set from 2 to 14 Hz. Here we use sequential progression selecting integer frequencies and



(a) Initial P-wave speed.



(b) Initial S-wave speed.



(c) Initial density.

Figure 6.71: Initial P-wave speed, S-wave speed and density for the reconstruction of the three-dimensional elastic Epati medium. Initial models consist of one-dimensional profiles with variation in depth only.

15 iterations are performed for each. This progression, in association with the number of coefficients to represent the models, is given in Table 6.10.

The final reconstructions after the 14 Hz iterations of our FWI minimization algorithm are presented in Figures 6.72 and 6.73, with the P- and S-wave speeds respectively.

Providing synthetic data, the reconstruction accurately recovers the variation of wave speeds, despite the difference between the profiles and the inaccurate density. The reconstructed P-wave speed has the increasing cone of high velocity (see the vertical section in

frequency iterations	**model representation (number of unknowns)**
2 Hz	60
3 Hz	616
4–5 Hz	5 610
7–7 Hz	47 250
8–10 Hz	378 000

Table 6.10: Relation between frequency and scale selected for the multi-level reconstruction of the elastic Epati model. No partition is employed for the last frequencies, from 11 to 14 Hz.

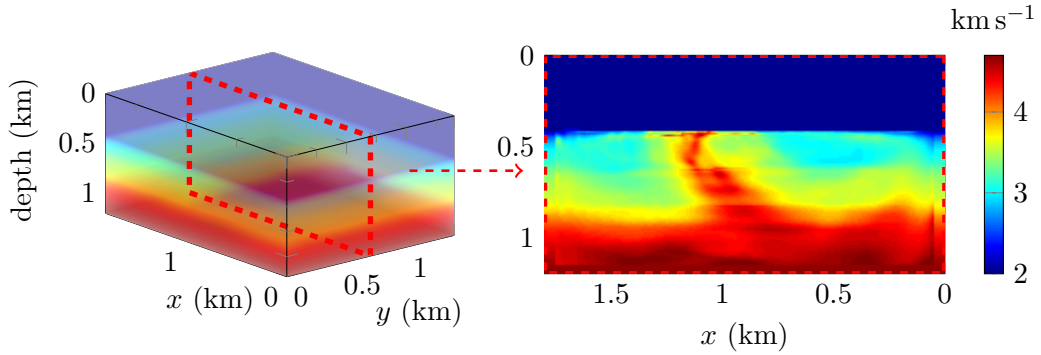


Figure 6.72: Reconstructions of the elastic Epati P-wave speed using frequencies from 2 to 14 Hz.Vertical section at $y = 700$ m is indicated with the red dashed frame and pictured in the right.

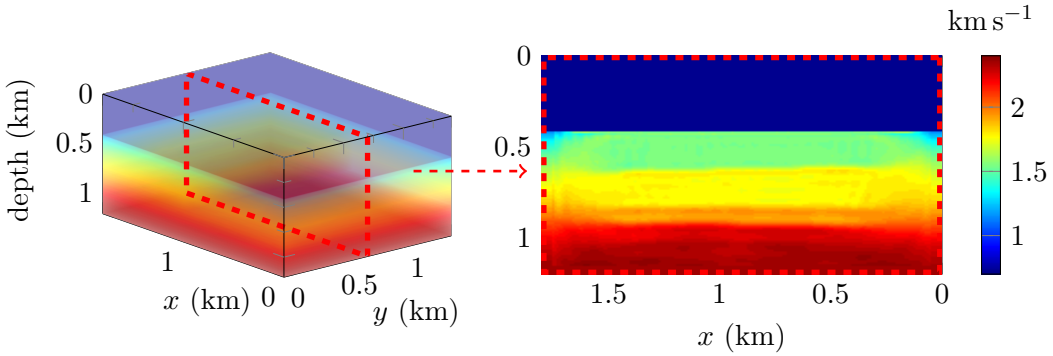


Figure 6.73: Reconstructions of the elastic Epati S-wave speed using frequencies from 2 to 14 Hz.Vertical section at $y = 700$ m is indicated with the red dashed frame and pictured in the right.

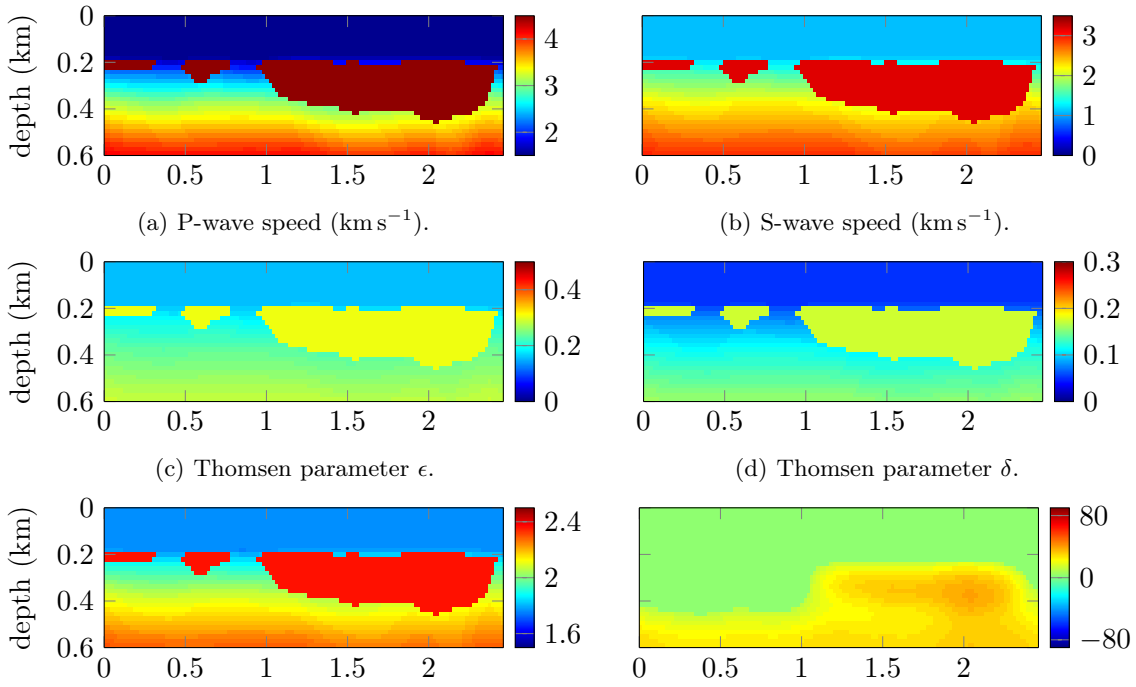
Figure 6.72). The reconstructed S-wave speed makes appear the different layers of constant speed, with accuracy. As usual, both sides have limited resolution due to the limited illumination. Once again, despite the unknown density, the recovery of the velocities is achieved.

This three-dimensional elastic test case shows some promises. There is clearly a need

to continue the experiment decreasing the accuracy of the observed data but we see that the optimization scheme can be conducted for large scale problems. The frequency domain approach certainly carries a computational burden regarding the matrix factorization but with novel linear algebra techniques, there is a promising road ahead.

6.12 Perspective 3: towards TTI reconstruction

We ave throughly benchmarked acoustic and elastic media, in two and three dimensions. The natural extension is to move towards TTI subsurface media. Such media are represented with six subsurface parameters in 2D and eight in 3D, see Section 1.4. Here we develop a test case to try out the algorithm in this context. We remain in two dimensions and design a domain of size 2.46 by 0.6 km. The medium is composed of the following parameters: the P-wavespeed, the S-wavespeed, the density, the Thomsen parameters (ϵ and δ) and the angle θ; they are pictured in Figure 6.74.



(a) P-wave speed ($\mathrm{km\,s^{-1}}$).

(b) S-wave speed ($\mathrm{km\,s^{-1}}$).

(c) Thomsen parameter ϵ.

(d) Thomsen parameter δ.

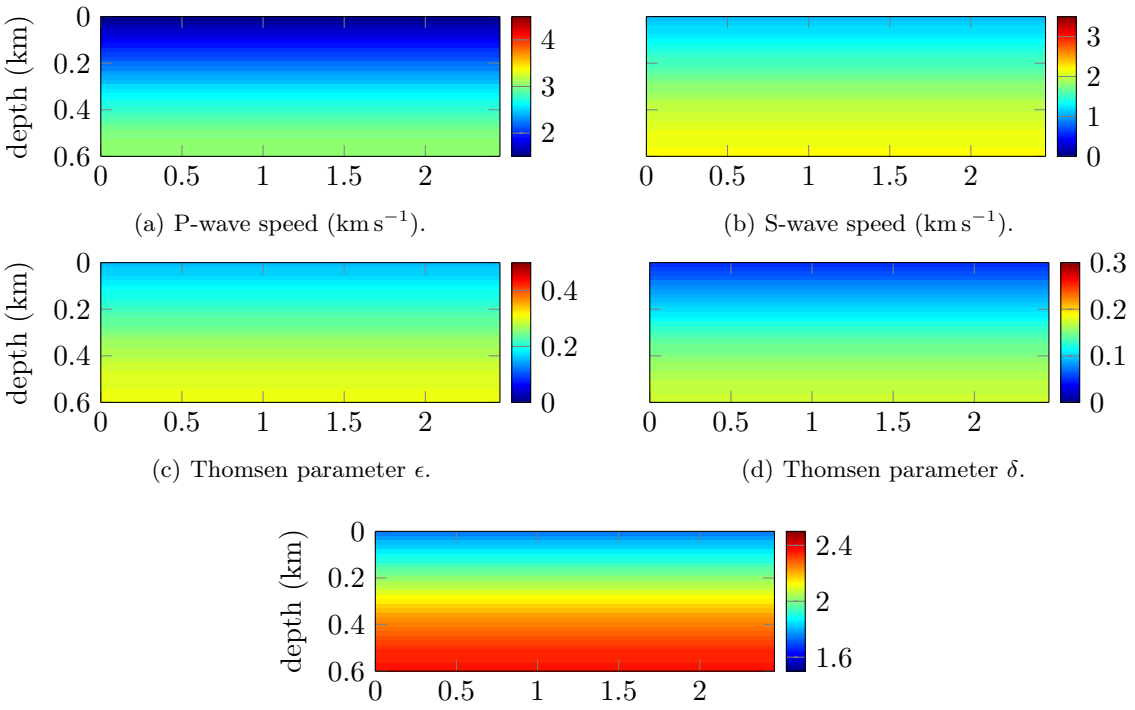
(e) Density ($10^3\mathrm{kg\,m^{-2}}$)

(f) TTI angle θ (°).

Figure 6.74: Experimental TTI models of size 2.46 by 0.6 km. The x axis is indicated in km.

Our algorithm is designed to handle the anisotropic wave equation and we attempt TTI reconstruction in this simple framework, just to verify that it behaves appropriately. We generate synthetic data in the frequency domain, without any noise and perform the iterative minimization with frequencies from 1 to 10 Hz. The initial models are given in

Figure 6.75: we start from one-dimensional profile. Moreover the initial angle θ is set to 0°. We follow a naive approach for the reconstruction where all parameters are inverted simultaneously.



(a) P-wave speed ($\mathrm{km\,s^{-1}}$).

(b) S-wave speed ($\mathrm{km\,s^{-1}}$).

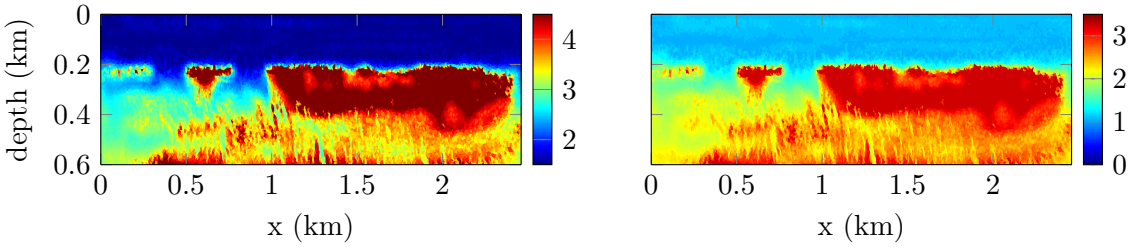
(c) Thomsen parameter ϵ.

(d) Thomsen parameter δ.

(e) Density ($10^3 \mathrm{kg\,m^{-2}}$).

Figure 6.75: Initial models used for the reconstruction of the TTI medium. The profiles consist in one-dimensional variations with depth. The x axis is indicated in km. The initial angle θ is constant with $\theta = 0°$.

After the final 10 Hz iterations, only the P- and S-wave speeds have been accurately approximated, they are pictured in Figure 6.76. We do not present the other parameters because the recovery fails, yet we see that the high contrast object appears in the reconstruction of the velocities. It seems that having inaccurate representation for the density, Thomsen parameters and angle still allows some insight into the subsurface structures.



(a) P-wave speed ($\mathrm{km\,s^{-1}}$).

(b) S-wave speed ($\mathrm{km\,s^{-1}}$).

Figure 6.76: Reconstruction of the P- and S-wave speeds of the TTI experiments. The other parameters are inverted as well but the algorithm fails to recover any useful information.

This is a very simple experiment and we cannot draw many conclusions out. Our code is able to handle anisotropy and we are now facing many questions regarding the reconstruction. The parametrization should be clarified to see if inverting from the stiffness tensor or the physical parameters makes a difference. Also it appears that the data are more sensitive to some parameters (the velocities here) so we should design the method to focus on those parameters. We have also use a small, noise-free, experiment and stepping to credible applications necessitates increased robustness and understanding.

6.13 Conclusion

We have benchmarked the FWI algorithm developed during this project and tested its robustness. Several experiments have been carried out, with acoustic and elastic media, in two and three dimensions. We have designed a multi-level multi-frequency algorithm based on our previous stability and convergence reasonings, so that frequency and model partitions evolved simultaneously. The benefits may be limited in the simplest cases (e.g., acoustic Marmousi), but it has proven to give some acute insight. It is in particular naturally adapted for salt domes reconstruction. It can also be appropriate when dealing with noise, because the compression can be assimilated to some regularization.

Concerning the recovery of the density, which is known to be a challenging problem, we have found that keeping an inaccurate representation does not prevent the algorithm from retrieving the subsurface structures. In particular, we have kept the density at some one-dimensional representation, and the reconstruction of the velocity (for both acoustic and elastic media) has only been marginally affected. Yet, there is a need for an alternative method for the approximation of the density.

We have shown that low or complex frequencies are required to initiate the recovery, due to their effect on the size of the basin of attraction. In particular, it has a crucial impact when salt domes are investigated. However, these frequencies are usually unavailable due to the noise. Hence there is a need to improve the robustness. It can consist of alternative techniques to build initial models (we have started with absolutely no information) or alternative algorithm. Here, we mention the MBTT algorithm (Migration Based Travel Times), developed by Chavent and Clément (1992); Clément et al. (2001) which shows some promises in avoiding low frequencies (Tcheverda et al., 2016). We plan its implementation in a very near future to evaluate its performance. Regarding salt domes reconstruction, we think that an appropriate way to handle it would be the use of level-sets methods (Lewis and Vigh, 2016; Kadu et al., 2017). However, without any a priori knowledge, it is hard to anticipate what to look for, and a more appropriate approach should be needed initially.

Because of the inherent optimization method involved, several possibilities are available, and their performance should be investigated. It includes the selection of the norm for the misfit function and the choice of descent direction to conduct the minimization. Here it would be interesting to pursue the benchmark in regards to the use of Hessian, or its approximation (such as pseudo-Hessian, Choi et al. (2008); Jun et al. (2015)) to identify how and when it is astute. Following the same idea, the use of regularization terms in the cost function (e.g., Qiu et al. (2016); Brandsberg-Dahl et al. (2017)) or following optimal control framework as recently developed for geophysical reconstruction in Métivier et al. (2016); Qiu et al. (2017), is an evolution that needs to be processed.

We have carried out a three-dimensional elastic test case which is an initial, yet fundamental, step towards more realistic applications of harmonic FWI. It must be put in perspective with new linear algebra techniques, to allow the handling of large matrix factorization. Finally, we have probed the algorithm to work with attenuation and anisotropy. These are crucial steps towards geophysical prospecting. They still require a thorough analysis to comprehend what can or cannot be expected in complex media and how it behaves with noise. Here we believe that the numerical algorithm cannot be sufficient and that a more global, analytical understanding can help to develop the appropriate workflow, to help focus on the parameter of interest.

Chapter 7

The perspectives of reconstructions from Cauchy data

Abstract

We investigate the seismic marine subsurface reconstruction in the frequency domain with the use of Cauchy data. It is motivated by the dual sensors acquisition devices, where pressure and vertical velocity are measured. Conditional Lipschitz stability for the associated inverse problem can be obtained for piecewise linear model representations. It involves a novel cost function derived from the Green's identity. We illustrate the effectiveness through computational experiments in three dimensions starting with time domain data. The reconstruction is carried out with single and multi frequency data. We also show that the use of Cauchy data with this new cost function leads to promising results, when minimal information is known on the acquisition set (namely we only require the position of the receivers but not the sources position, nor their wavelet).

Disclaimer: *This is an ongoing work in part of collaboration with Profs. Giovanni Alessandrini, Hélène Barucq, Maarten V. de Hoop, Romina Gaburro and Eva Sincich. we mention the two related proceedings: Alessandrini et al. (2017a,b).*

Contents

We study the possibility of reconstructions in the marine geophysical context when Cauchy data are available. Seismic marine acquisition standardly relies on measurements obtained from hydrophones, but perspectives have arisen with the new types of acquisition devices that allow multi components data recordings. In particular, the *dual sensors*[7] are able to capture the vertical (or normal) velocity in addition to the pressure, see Carlson et al. (2007); Tenghamn et al. (2007). It increases the quantity of information and provides benefits in terms of noise reduction and for image processing, as indicated in Whitmore et al. (2010); Rønholt et al. (2015). In this chapter, we investigate the application of our Full Waveform Inversion (FWI) algorithm with this type of data in order to identify the possible gains.

Instead of the normal velocity measures, we first focus on the use of Cauchy data, motivated by mathematical reasons, but having in mind the data probed by dual sensors devices for applications. However, our numerical experiments are conducted with the derivative of the pressure with respect to the vertical direction, which accordingly relates to the vertical velocity from the Euler's equations, as illustrated in Carlson et al. (2007). We note that compared to the Dirichlet-to-Neumann map employed in Chapter 3 (data which, as mentioned, are not observed directly in seismic acquisition anyway), the Cauchy data do not suffer from eigenfrequencies.

The inverse problem associated with Cauchy data has been proven to have a conditional Lipschitz stability, when the model is represented by piecewise linear functions. The result holds in particular in the case of partial data, using a free surface and absorbing boundary conditions on some portion of the boundary. For the analysis, we refer to Alessandrini et al. (2017c,b). In Section 7.1, we review the formulation of the problem and identify the appropriate cost function to employ. We also defined the associated adjoint state for the gradient computation. In Section 7.2, we experiment the reconstruction for a three-dimensional acoustic situation. First, the reconstruction is conducted with data at one single frequency and a fixed piecewise linear representation of the velocity. Then we allow a multi-frequency reconstruction with scale evolution. Finally, we show that from the novel cost function that is introduced, minimal information is actually required on the seismic acquisition set, see Section 7.3. We show that we can artificially incorporate sources for the simulations, which further allows a reduction of the computational time.

[7] Dual sensors are developed by Petroleum Geo-Services, PGS, see https://www.pgs.com/publications/feature-stories/dual-sensor-stands-high-after-ten-years/

7.1 Iterative minimization from Cauchy data

7.1.1 Framework

We consider the domain $\Omega \subset \mathbb{R}^2$ or $\mathbb{R}^3$, with boundary $\Gamma = \Gamma_1 \cap \Gamma_2$. The distinction of boundaries is made so that Γ_1 denotes the free surface boundary and Γ_2 the others. The pressure field p satisfies the wave equation,

$$\begin{cases} -\omega^2 c^{-2} p(\mathbf{x}) - \Delta p(\mathbf{x}) = f(\mathbf{x}) & \text{in } \Omega, \\ p(\mathbf{x}) = 0, & \text{in } \Gamma_1, \\ \partial_\nu p(\mathbf{x}) - \mathrm{i}\omega c^{-1}(\mathbf{x}) p(\mathbf{x}) = 0, & \text{in } \Gamma_2, \end{cases} \tag{7.1}$$

We have taken the free surface condition on Γ_1 and absorbing boundary conditions on Γ_2, (Engquist and Majda, 1977). Here, f denotes the source. For the reconstruction, Cauchy data (p and $\partial_\nu p$) are acquired on a portion Σ of the domain which is the receivers location. In particular, we make the important assumption that Σ is located below the sources. The associated forward problem at frequency ω (see Equation (4.2)) is

$$\mathcal{F}_\omega \ : \ c \quad \rightarrow \quad \Big\{ p(\mathbf{x}) \mid_\Sigma ; \ \partial_\nu p(\mathbf{x}) \mid_\Sigma \Big\}.$$

We assume the source and receivers to be located slightly underneath the surface, and we further consider the knowledge of the upper part of the medium where they lie, which is water. This setup is sketched in Figure 7.1 in two dimensions. In the numerical experiment we later consider the derivative with respect to z of the pressure field instead of the normal derivative. This is actually closer to the actual measurements of dual sensors, where the vertical velocity is related to the derivative of the pressure with the Euler's equations.

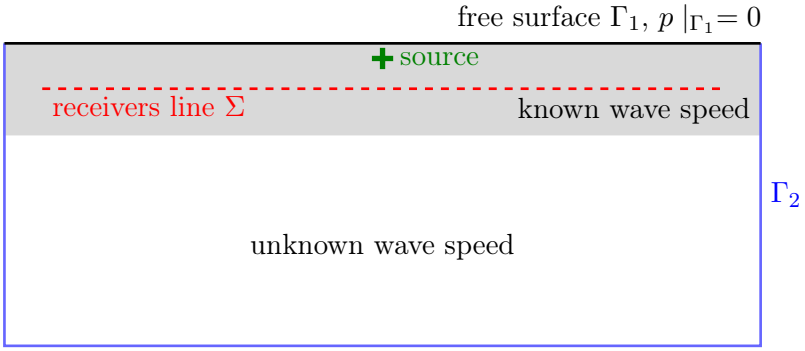


Figure 7.1: Illustration of the configuration in two dimensions. We apply a Dirichlet boundary condition on Γ_1 and absorbing boundary conditions on Γ_2, following Problem (7.1). We note that $\Gamma = \Gamma_1 \cap \Gamma_2$. The sources that probe the Cauchy data lie in between the receivers and the free surface, in a layer of known velocity (water).

7.1.2 Piecewise linear model representation

The unknown parameter, the velocity, is represented following a piecewise linear model representation. The domain Ω is partitioned with N non-overlapping subdomains $\mathcal{D}_k$ such that

$$\Omega = \big\{ \cup_{k=1}^{N} \mathcal{D}_k, \quad \mathcal{D}_i \cap \mathcal{D}_j = \emptyset \ \ \forall i \neq j \big\}. \tag{7.2}$$

Following the partition, the velocity is represented via piecewise linear functions, such that

$$c(\mathbf{x}) = \sum_{k=1}^{N} \mathcal{H}_k(\mathbf{x}) \chi(\mathcal{D}_k),$$

where $\chi(\mathcal{D}_k)$ represents the characteristic function over the subdomain D_k and $\mathcal{H}_k(\mathbf{x})$ is a linear function. Denoting the space variable $\mathbf{x} = \{x_1, \ldots, x_r\}$, where r is the dimension (2 or 3 here), this function is given by

$$\mathcal{H}_k(\mathbf{x}) = \alpha_{k,0} + \sum_{j=1}^{r} \alpha_{k,j} x_j.$$

$\alpha_{k,j}$ is defined for $k \in \{1, \ldots, N\}$, $j \in \{1, \ldots, r\}$. The number of scalar coefficients $\alpha_{k,j}$ is given by $(r+1)N$. Therefore, we have $4N$ coefficients in three dimensions and $3N$ in two dimensions. We refer to Section 3.4 of Chapter 3 for additional details and illustrations.

Remark 7.1 (Relation with the forward discretization)**.** *The Galerkin discretization of the wave equation generates a mesh of the domain Ω, such that*

$$\Omega = \big\{ \cup_{k=1}^{N_K} K_k, \quad K_i \cap K_j = \emptyset \ \ \forall i \neq j \big\},$$

where the N_K cells are usually triangles for a two-dimensional domain and tetrahedra in three dimensions. This domain decomposition can have no relation with the piecewise linear partition in N subdomains given in the Equation (7.2). Yet we naturally define the subdomains for the piecewise linear approximation in terms of the mesh cells and we take each $\mathcal{D}_k$ as the union of K_i,

$$\mathcal{D}_k = \bigcup_{i=1}^{N_{\mathcal{D}_k}} K_i, \qquad \forall k \in \{1, \ldots, N\}.$$

We have introduced $N_{\mathcal{D}_k}$, the number of mesh cells in the subdomain $\mathcal{D}_k$, which varies with k. Eventually, we have the relation between the total number of mesh cells (N_K) and the number of piecewise linear partitions (N),

$$\sum_{k=1}^{N} N_{\mathcal{D}_k} = N_K.$$

7.1.3 Misfit functions

For the iterative minimization we denote the decomposition of the data with d_D and d_N. d_D refers to the pressure observations and d_N stands for the normal derivative observations. Similarly, we require two restriction operators to apply onto the simulated pressure field p (following what is introduced in Section 4.3). We define

$$\begin{cases} \mathcal{R}_D(p(\mathbf{x})) = p(\mathbf{x}) \mid_\Sigma, \\ \mathcal{R}_N(p(\mathbf{x})) = \partial_\nu p(\mathbf{x}) \mid_\Sigma . \end{cases}$$

We note that $\mathcal{R}_D$ is the same restriction operator as $\mathcal{R}$ in Chapter 4 , meanwhile $\mathcal{R}_N$ encompasses a normal derivative.

With the Cauchy data, we first naively derive a cost function based on the traditional least squares, by incorporating the norm of the normal derivative such that

$$\mathcal{J}_{LS}(c) = \frac{1}{2}\sum_{k=1}^{n_{src}} \|\mathcal{R}_D(p^{(s_k)}(c)) - d_D^{(s_k)}\|^2 + \frac{\eta}{2}\|\mathcal{R}_N(p^{(s_k)}(c)) - d_N^{(s_k)}\|^2, \tag{7.3}$$

where the index (s_k) denotes the sources. We have also introduced a weighting parameter η to penalize the data, and give more or less influence on the normal derivative data. By taking $\eta = 0$, we revert back to the original marine seismic FWI, with pressure data only.

In Alessandrini et al. (2017c,b), the conditional stability of the inverse problem associated with the Problem (7.1) is obtained, in the case of piecewise linear model representation. It is a particularly powerful result because it accounts for the geophysical context: the partial data, the free surface and the absorbing boundary conditions (contrary to the more standard Dirichlet-to-Neumann data case we have proven in Chapter 3). On the other hand, it does not use the cost function (7.3), but instead defines the following,

$$\mathcal{J}_G(c) = \frac{1}{2}\sum_{k=1}^{n_{src}}\sum_{l=1}^{n_{src}} \|d_N^{(s_k)T}\mathcal{R}_D(p^{(s_l)}(c)) - d_D^{(s_k)T}\mathcal{R}_N(p^{(s_l)}(c))\|^2. \tag{7.4}$$

Hence we have a new cost function based on the correlation of observation with simulation. The understanding of this cost function is extracted from the second Green's identity which states that, for two regular enough functions ϕ, ψ,

$$\int_V \phi\Delta\psi - \psi\Delta\phi \,\mathrm{d}V = \int_S \phi\partial_\nu\psi - \psi\partial_\nu\phi \,\mathrm{d}S,$$

where V is a volume with boundary S. We notice that the left hand side equates to zero if both ϕ and ψ solves the Helmholtz equation with same velocity c. The cost function can be recognized in the right-hand side but requires a little effort in the context of partial data.

7.1.4 Discretized gradient computation with adjoint state

The computation of the gradient follows the adjoint state method we prescribed in Section 4.4 where the appropriate references have been given.

7.1.4.1 Least squares cost function

From the least squares cost function of Equation (7.3), the adjoint state is derived following the exact same path as in Section 4.4, incorporating the additional norm of the difference for the normal derivative data. It is straightforward to see that now, the adjoint state $\gamma^{(s_k)}$ (associated with the source s_k) is selected to solve the problem (analog to Equation (4.21))

$$\mathcal{P}^*\gamma^{(s_k)} = -\mathcal{R}_D^*(\mathcal{R}_D(p^{(s_k)}) - d_D^{(s_k)}) - \eta\mathcal{R}_N^*(\mathcal{R}_N(p^{(s_k)}) - d_N^{(s_k)}).$$

One adjoint state needs to be computed per source and the gradient of the cost function follows (similarly to Equation (4.22)),

$$\nabla\mathcal{J}_{LS} = \sum_{k=1}^{n_{src}} \mathrm{Re}\left(\langle \partial_c(\mathcal{P})p^{(s_k)}, \gamma^{(s_k)}\rangle\right),$$

where $\mathcal{P}$ correspond to the wave operator.

7.1.4.2 Green's cost function

Let us now study the new cost function given in the Equation (7.4). The adjoint state is computed in the same manner, without any more technicality. It is given as the solution of the adjoint of the forward problem, where the right-hand side is the derivative of the cost function with respect to p. According to the double sum in the Equation (7.4), there is now two indexes associated with the adjoint: one related to the observation source and one for the simulation source, denoted by s_k and s_l respectively. The adjoint state $\gamma^{(s_k,s_l)}$ solves the problem

$$\mathcal{P}^*\gamma^{(s_k,s_l)} = -\left(d_N^{(s_k)T}\mathcal{R}_D - d_D^{(s_k)T}\mathcal{R}_N\right)^*\left(d_N^{(s_k)T}\mathcal{R}_D(p^{(s_l)}) - d_D^{(s_k)T}\mathcal{R}_N(p^{(s_l)})\right). \tag{7.5}$$

The gradient follows to be expressed by

$$\nabla\mathcal{J}_G = \sum_{k=1}^{n_{src}}\sum_{l=1}^{n_{src}} \mathrm{Re}\left(\langle \partial_c(\mathcal{P})p^{(s_l)}, \gamma^{(s_k,s_l)}\rangle\right).$$

At this point, we would require the computation of the adjoint state for every couple (s_k, s_l), which would represent an important computational burden. This can actually be

immensely reduced. We have

$$\nabla \mathcal{J}_G = \sum_{k=1}^{n_{src}} \sum_{l=1}^{n_{src}} \mathrm{Re}\left(\langle \partial_c(\mathcal{P}) p^{(s_l)}, \gamma^{(s_k,s_l)} \rangle \right) = \sum_{l=1}^{n_{src}} \mathrm{Re}\left(\langle \partial_c(\mathcal{P}) p^{(s_l)}, \sum_{k=1}^{n_{src}} \gamma^{(s_k,s_l)} \rangle \right).$$

Let us introduce

$$\sum_{k=1}^{n_{src}} \gamma^{(s_k,s_l)} = \gamma^{(s_l)}.$$

From the linearity of the wave equation (7.5), we see that $\gamma^{(s_l)}$ solves the problem

$$\mathcal{P}^* \gamma^{(s_l)} = -\sum_{k=1}^{n_{src}} \left(d_N^{(s_k)T} \mathcal{R}_D - d_D^{(s_k)T} \mathcal{R}_N \right)^* \left(d_N^{(s_k)T} \mathcal{R}_D(p^{(s_l)}) - d_D^{(s_k)T} \mathcal{R}_N(p^{(s_l)}) \right).$$

One needs to solve only one adjoint state per simulated source. Hence, the cost for the computation of the gradient of this cost function is similar to the cost of the least squares one.

Remark 7.2. *We notice that in the variational formulation of the problem, the right-hand side is tested against basis functions (see Chapter 2). It provides a natural framework because we have, by definition,* $\forall \psi, \phi \in H^1(\Sigma)$,

$$< \mathcal{R}_N^* \phi, \psi \mid_\Sigma > = < \phi \mid_\Sigma, \mathcal{R}_N \psi >,$$

where $< \cdot, \cdot >$ *is the dual pairing* $H^{-1/2}(\Sigma)$, $H^{1/2}(\Sigma)$. *Therefore, the normal derivative (encompassed in* $\mathcal{R}_N$*) can be considered on the test functions.*

7.2 Acoustic 3D model reconstruction

7.2.1 Acquisition setup and Cauchy data

To evaluate the performance of the reconstruction from Cauchy data, we consider the acoustic three-dimensional model of Figure 7.2 (courtesy of Statoil). It is a model of size $2.54 \times 1.44 \times 1.22$ km, where we also visualize selected sections at fixed $y = 1125$ m and at fixed depth $z = 800$ m. The density is set to a constant with $\rho = 1000\ \mathrm{kg\,m^{-3}}$. The profile is particularly interesting because there is a deep layer where the wave speed is reduced, and then it increases again.

The data are generated in the time domain, where we have considered a regular lattice for the sources and receivers. We take a total of 160 sources and 1376 receivers for each source. The resulting three-dimensional seismic traces for a centrally located shot are pictured in Figures 7.3(a) and 7.3(b) for the pressure observation and the its derivative

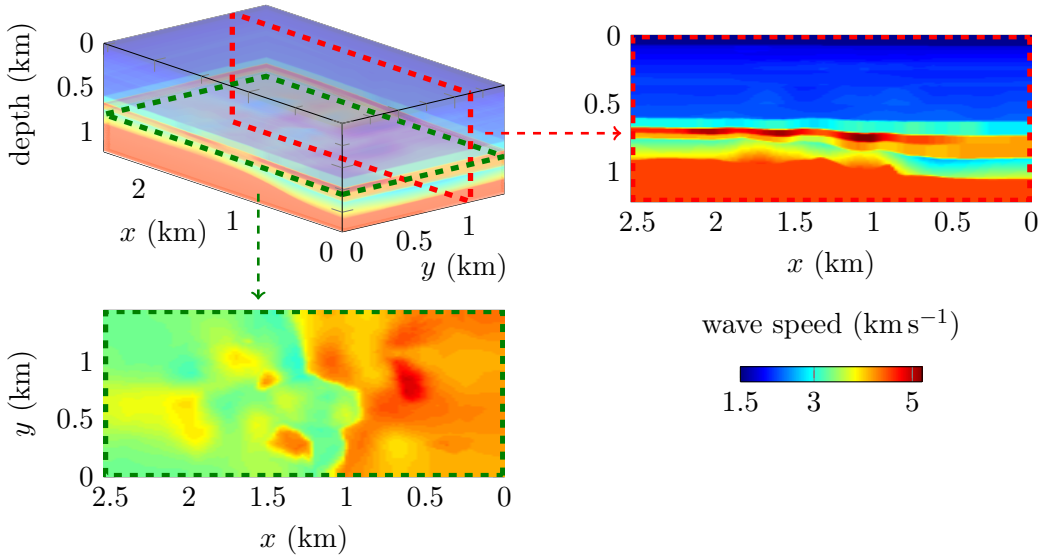


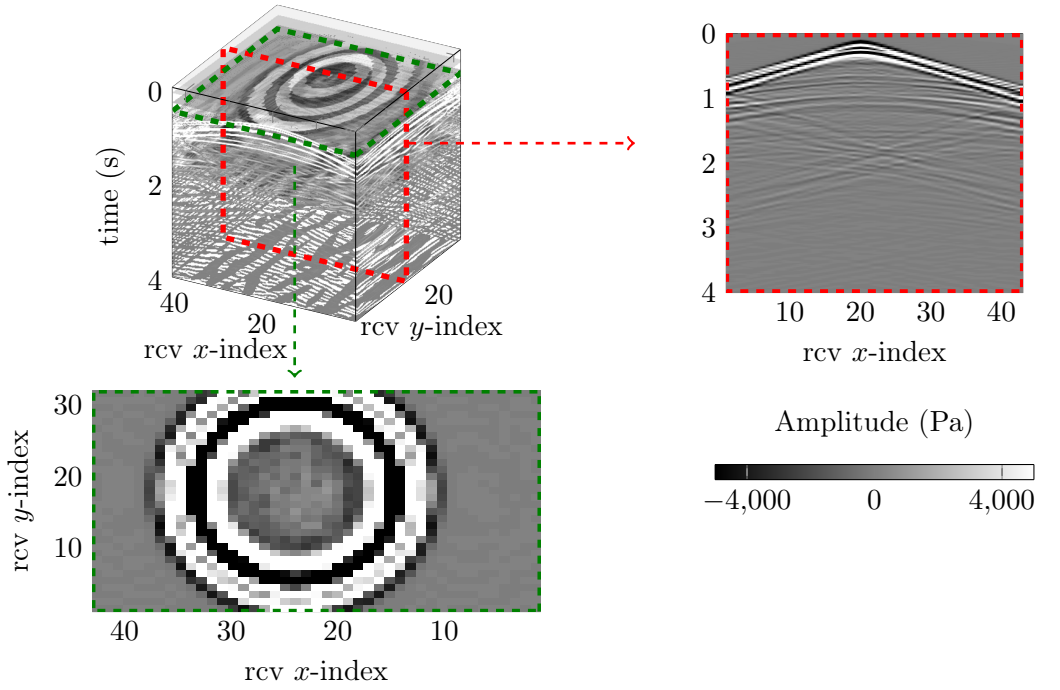
Figure 7.2: Three-dimensional acoustic wave speed (courtesy of Statoil) of size $2.54 \times 1.44 \times 1.22$ km, the velocity varies between 1500 and 5200 $\mathrm{m\,s^{-1}}$. Vertical section at $y = 1125$ m is indicated with the red dashed frame and plotted in the right. The horizontal section at fixed depth $z = 800$ m is indicated with the green frame and shown at the bottom. The model is represented from 1 527 168 nodal values.

with respect to z, which can be assimilated to the vertical velocity (using the Euler's equation, Carlson et al. (2007)). We present the three-dimensional traces and associate two two-dimensional sections, at fixed time $t = 0.5$ s (bottom left) and for a fixed line of receivers positioned at $y = 695$ m (upper right).

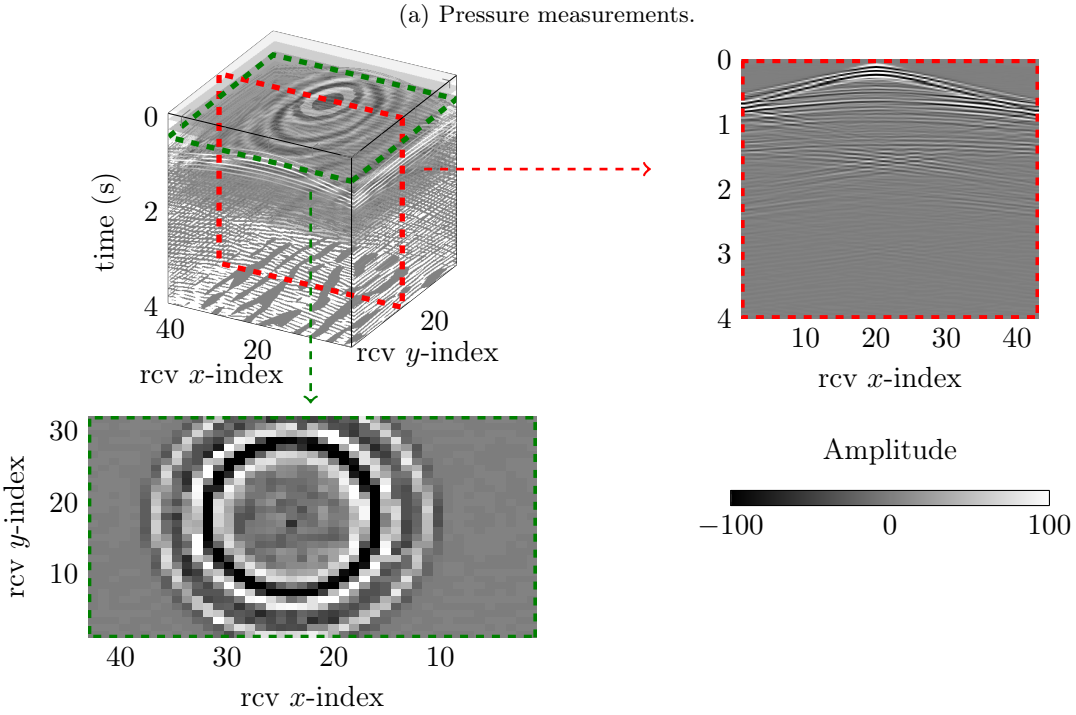
We observe a large difference of scale between the two types of data, with one order of magnitude difference in the amplitude. It justifies the scaling parameter η in the least-squares misfit function Equation (7.3), which allows us to adjust the influence of the different types of data.

7.2.2 Single frequency reconstruction

From the observed data in the time domain, we apply a Fourier transform to extract the frequency content that we use in our harmonic FWI algorithm. We first try to use the data extracted for one single frequency equal to 10 Hz. The initial model for the reconstruction consists in a smooth background: a one-dimensional velocity profile, increasing with depth, as depicted in Figure 7.4. For the discretization, we employ continuous Galerkin discretization, the number of cells in the mesh is about 3×10^5 and the number of degrees of freedom is 7×10^5.



(a) Pressure measurements.



(b) Vertical derivative measurements.

Figure 7.3: Three-dimensional trace associated with a centrally located source. A two-dimensional section at a fixed time $t = 0.5$ s is pictured in the bottom left of each figure, and is indicated in green in the 3D visualization. The upper right of each figure selects a fixed line receivers positioned at $y = 695$ m, which is indicated by the red dashed line in the 3D visualization.

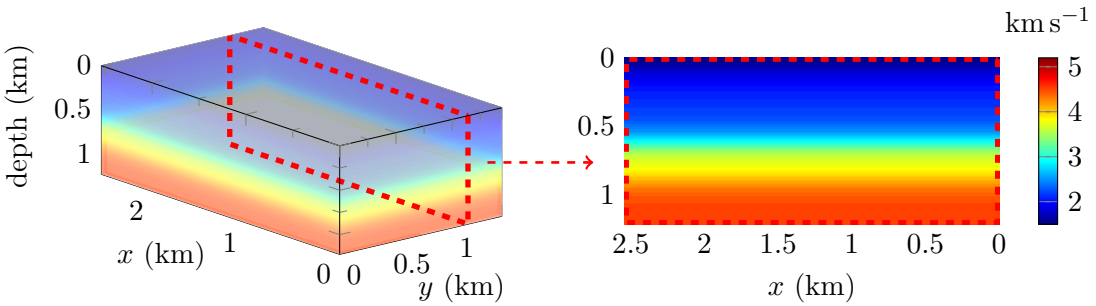


Figure 7.4: Initial wave speed for the iterative minimization algorithm and reconstruction of the Statoil model with single frequency data. Vertical section at $y = 1125$ m and horizontal section at fixed depth $z = 800$ m are detailed.

Following the stability result, we employ a domain partition where piecewise linear representation is applied, see Subsection 7.1.2. The original model representation (Figure 7.2) is made of 1.5×10^6 nodal values and the mesh we employ to discretize the domain is of 3×10^5 tetrahedra. We largely reduce this number for the piecewise linear representation and select 1089 subdomains. Because our initial guess has no information on the subsurface, the choice of subdomain follows a structured decomposition of the domain. Following the piecewise linear representation, it leads to $4 \times 1089 = 4356$ coefficients to represent the velocity. It is 0.3% of the nodal representation and 1.3% in regards to the number of cells. Hence, we have drastically reduced the number of unknowns in the problem.

We perform the iterative minimization of the cost function established from the Green's identity, Equation (7.4). We only use 10 Hz data and the piecewise linear representation. In Figure 7.5, we show the reconstruction after 175 iterations.

We are able to recover the variations of the velocity from single frequency data at 10 Hz. In particular, in the vertical section of Figure 7.5, we see that the profile with a layer where the velocity decreases is well captured. Despite the lack of initial information and the single frequency content, the method benefits from the Cauchy data and proves successful reconstruction. We have drastically reduced the number of coefficients (0.3% compared to Figure 7.2), it is visible in the reconstruction but the resolution is appropriate, it can further easily be improved, see Remark 7.3.

We can also evaluate the performance with the comparison of data at the receivers location. In Figures 7.6 and 7.7, we show the data of the map of receivers for a centrally localized shot. We compare the Fourier transform of the observed measurements with the records simulated from the initial and reconstructed models.

We can observe the improvement provided by the iterative reconstruction and verify that both types of data have been ameliorated. In order to have a better visualization,

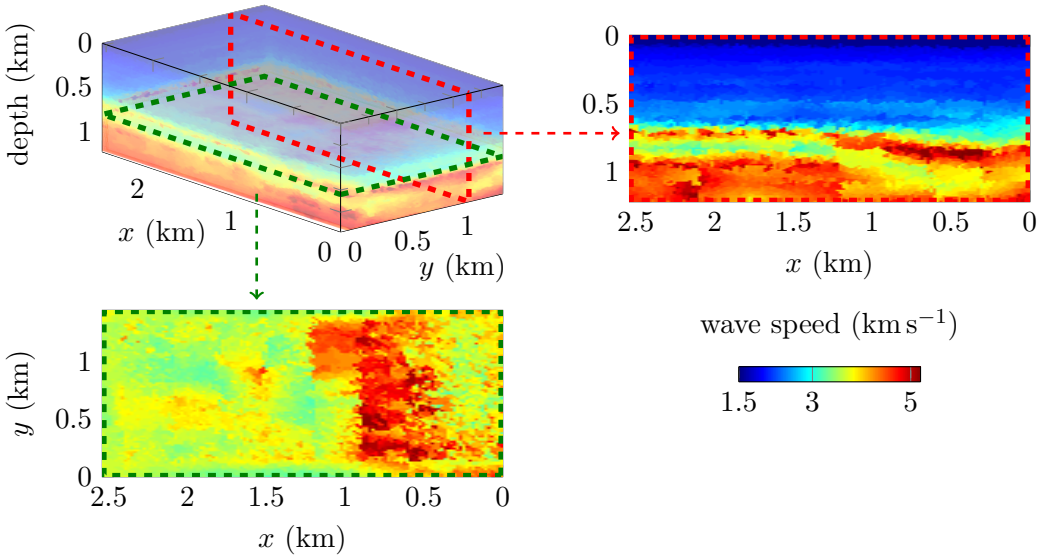
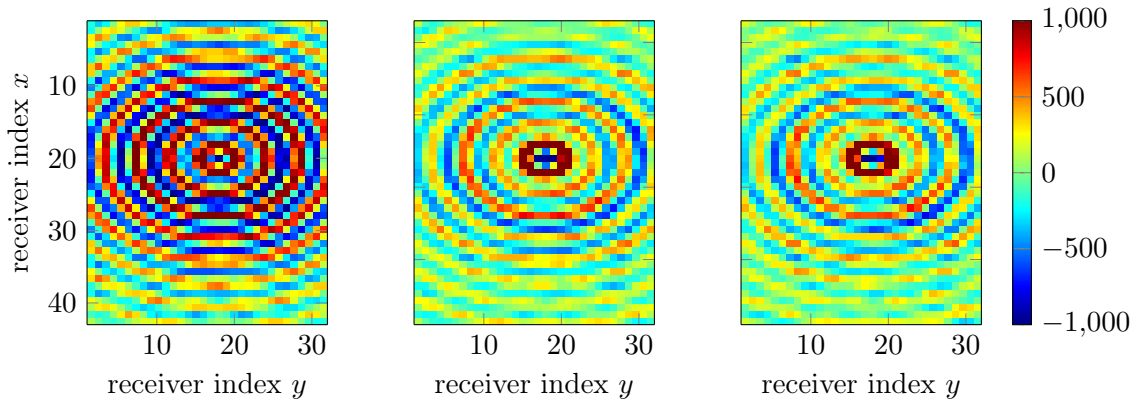


Figure 7.5: Reconstruction of the acoustic 3D Statoil wave speed from Cauchy data at 10 Hz only, with the initial guess of Figure 7.4. The model is represented via 1089 structured subdomains where piecewise linear functions are applied, for a total of 4356 coefficients. The vertical section at $y = 1125$ m and horizontal section at fixed depth $z = 800$ m are detailed.



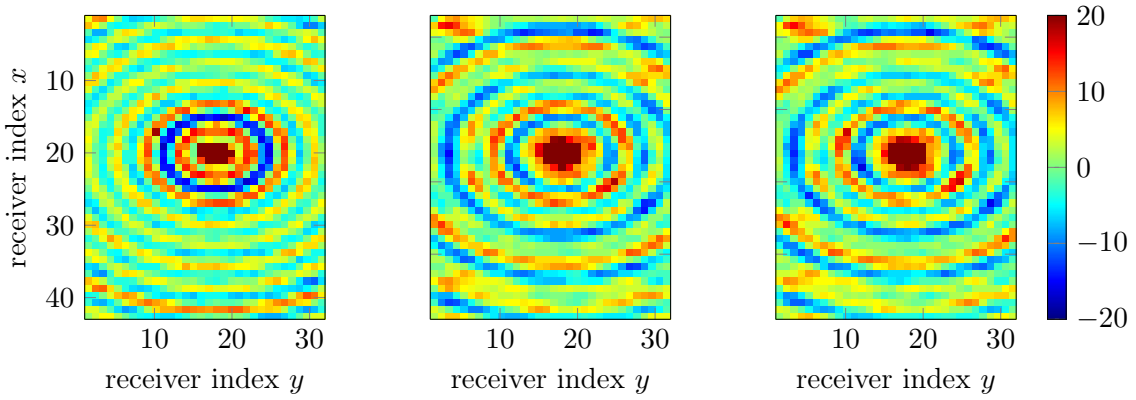
(a) Simulation from the initial model.

(b) Fourier transform of the time domain observation.

(c) Simulation from the reconstructed model.

Figure 7.6: Comparison of the real part of the 10 Hz frequency pressure data for a centrally located source.

we plot the information for a fixed receivers index y in Figure 7.8, where we select the tenth index. The data from the reconstructed model match the observed measurements, at the exception of the central area, near the source position. Both types of data are well approximated.

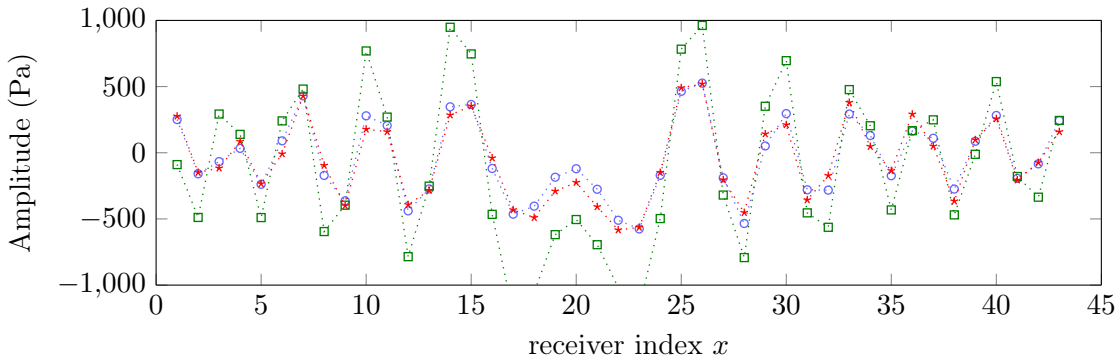


(a) Simulation from the initial model.

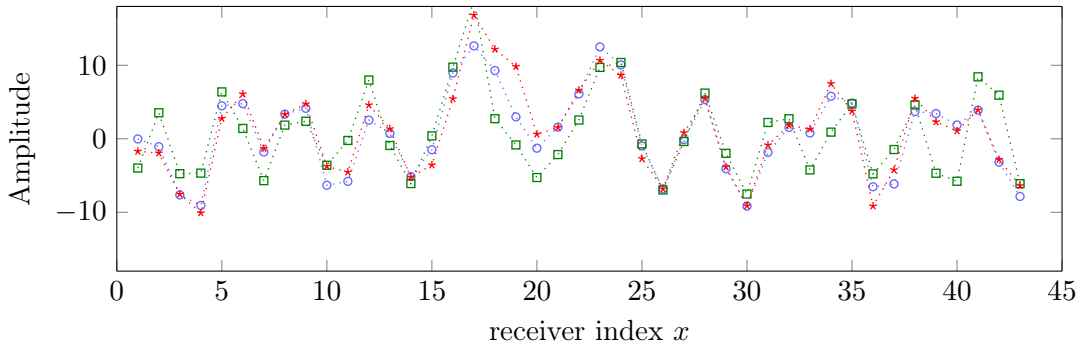
(b) Fourier transform of the time domain observation.

(c) Simulation from the reconstructed model.

Figure 7.7: Comparison of the real part of the 10 Hz frequency vertical data for a centrally located source.



(a) Pressure data.



(b) Vertical derivative data.

Figure 7.8: Real part of the 10 Hz data recorded by a single line of receivers associated with a centrally located source. The blue circles (···○···) represent the Fourier transform of the observed data, the green squares (···□···) are the simulation using the initial model and the red crosses (···×···) are the simulation from the final reconstruction.

7.2.3 Multi-frequency reconstruction

The reconstruction from single frequency data has been successfully carried out and we now investigate the, more traditional, multi-frequency situation. From the time domain observation, we apply the Fourier transform for frequencies from 3 to 15 Hz, using 1 Hz step. Because we have access to lower frequency, we can take a starting model with even less accuracy, see Figure 7.9. Here the one-dimensional variation in depth assumes a much lower velocity background compared to the starting velocity of Figure 7.4.

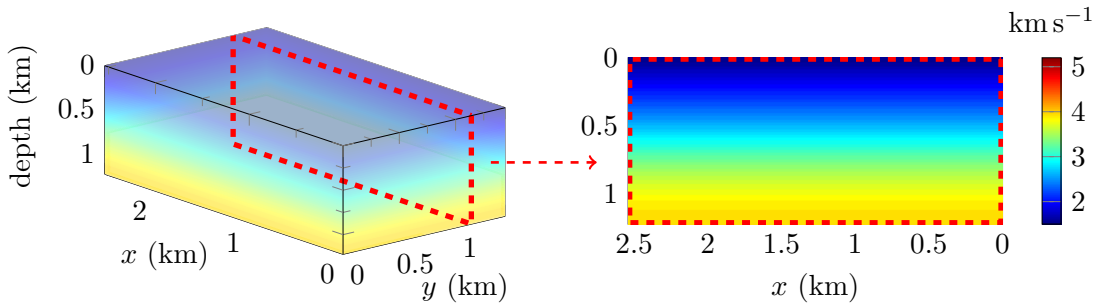


Figure 7.9: Initial wave speed for the iterative minimization algorithm and reconstruction of the Statoil model with multi frequency data. The vertical section at $y = 1125$ m and horizontal section at fixed depth $z = 800$ m are detailed.

For the reconstruction, we employ the multi-level algorithm (see Section 6.1) with a simultaneous progression in scale and frequency. Contrary to the experiments performed in Chapter 6, we now consider piecewise linear functions over the partitions. Table 7.1 provides the relation between frequency and scale during the iterative procedure.

frequency iterations	**number of coefficients in the model representation**
3 Hz	600
4 Hz	840
5 Hz	1 764
6 Hz	3 520
7 Hz	6 760
8–15 Hz	52 000

Table 7.1: Relation between frequency and scale selected for the reconstruction of the acoustic Statoil model. The representation employs piecewise linear functions, so that the number of coefficients corresponds to $4N$ where N is the number of subdomains in the representation.

We perform 20 iterations per frequency and compare the results depending on three situations. The reconstruction using Cauchy data and the cost function $\mathcal{J}_G$ of Equation (7.4) is pictured in Figure 7.10. Then we depict the reconstruction using Cauchy data with the least squares functional $\mathcal{J}_{LS}$ (of Equation (7.3)) and the reconstruction using the pressure

data only in Figure 7.11.

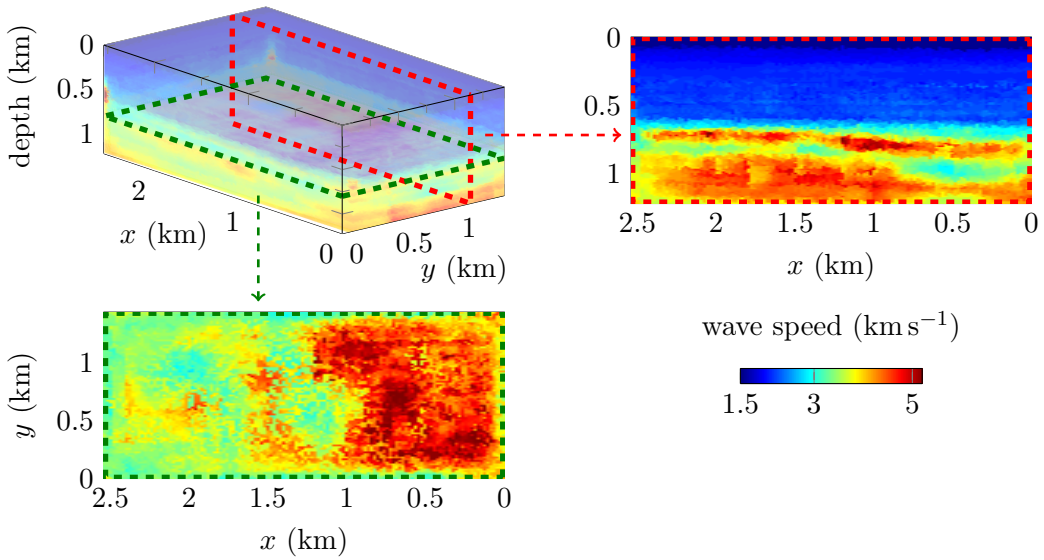
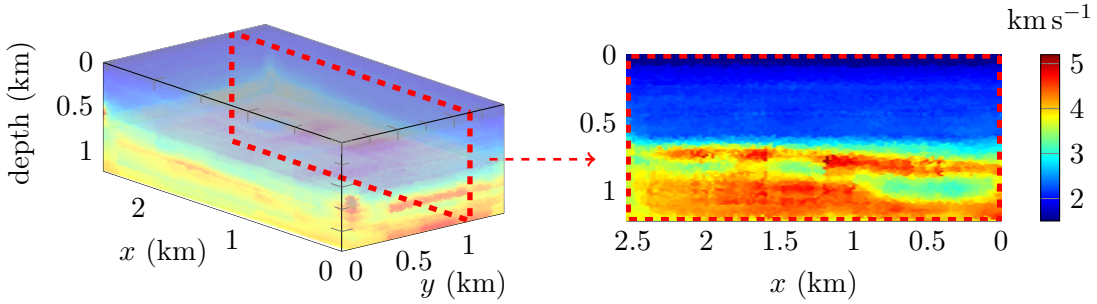
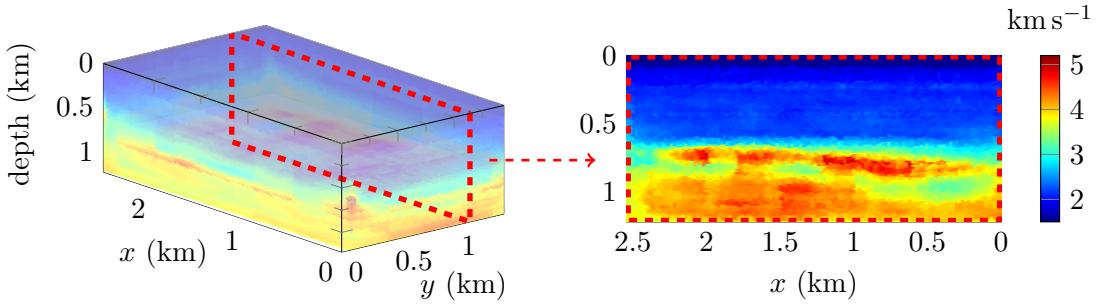


Figure 7.10: Reconstruction of the acoustic 3D Statoil wave speed from Cauchy data using frequencies from 3 to 15 Hz and the cost function $\mathcal{J}_G$ (Equation (7.4)). The initial guess is pictured in Figure 7.9. The reconstruction uses simultaneous progression in frequency and scale, prescribed in Table 7.1.



(a) Reconstruction using Cauchy data and the least squares cost function $\mathcal{J}_{LS}$ given in the Equation (7.3).



(b) Reconstruction using the pressure data only.

Figure 7.11: Comparison of reconstructions of the acoustic 3D Statoil wave speed from Cauchy data using frequencies from 3 to 15 Hz. The initial guess is pictured in Figure 7.9. The reconstruction uses simultaneous progression in frequency and scale, prescribed in Table 7.1.

We see that the method is able to recover the model appropriately in all situations. The vertical sections are accurately designed, in particular the layer of low velocity in between larger values is retrieved. When we only use the pressure data (i.e. the standard marine FWI from hydrophone measurements), see Figure 7.11(b), we obtain limited accuracy on the sides and the model looks not as sharp as when using the Cauchy data. When comparing the reconstruction with Cauchy data obtained from the two choices of cost functions, Figures 7.10 and 7.11(a), the resolution is similarly satisfying and we barely see any difference.

Remark 7.3 (Smoothing the images)**.** *We have seen in the previous experiments that the overall recovery suffers from the drastic reduction in the number of coefficients, and from the tetrahedral mesh. It results in some coarse aspect in the visualization of the final reconstruction. An easy way to overcome this is to apply a smoothing filter onto the final image. This can be done, for example, with the* `imgaussfilt` *function of MATLAB®, which applies a Gaussian smoothing filter. In Figure 7.12, we show its application onto the reconstruction of Figure 7.10.*

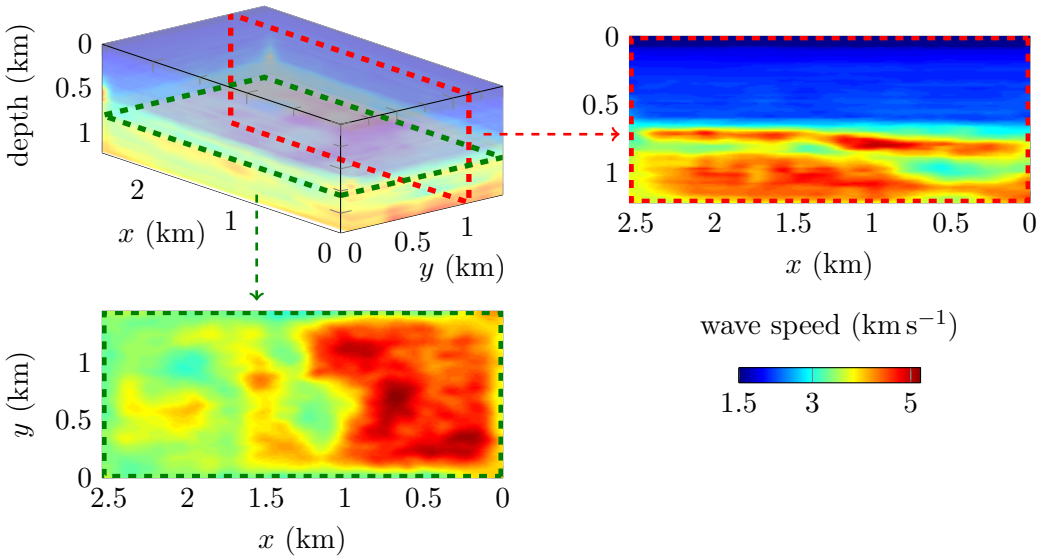


Figure 7.12: Application of a Gaussian smoothing filter onto the velocity reconstruction shown in Figure 7.10. The smoothing has been realized using the function `imgaussfilt` in MATLAB®.

The smoothing of the final reconstruction is an effortless procedure which provides a very astute visualization of the final results. Here we realize even more the accuracy obtained from the minimization of the Cauchy data.

7.3 Perspective from the Green's cost function

In our previous experiment, we have compared the reconstructions from Cauchy data depending on the choice of cost function ($\mathcal{J}_{LS}$ or $\mathcal{J}_G$), and when using the pressure data only. We have observed some improvements from the use of Cauchy data but the choice of cost functions seems to provoke few modification. It is however important to remain in the analytical framework to validate the conditional Lipschitz stability result. Also, by taking a closer look at the architecture of $\mathcal{J}_G$ in the Equation (7.4), tremendous opportunities appear.

Indeed, let us uncouple the set of sources for the simulations and the observations, we write

$$\mathcal{J}_G(c) = \frac{1}{2}\sum_{k=1}^{n_{src}^{obs}}\sum_{l=1}^{n_{src}^{sim}} \|d_N^{(s_k)T}\mathcal{R}_D(p^{(s_l)}(c)) - d_D^{(s_k)T}\mathcal{R}_N(p^{(s_l)}(c))\|^2,$$

where n_{src}^{obs} is the number of sources in the set of observations and n_{src}^{sim} is the number of sources in the acquisition used for the simulation. By doing this, we emphasize the fact that the two sets of sources (for the observation and the simulation) could be totally uncorrelated. This is impossible for the least squares misfit function where the same acquisition must be used.

By differentiating the set of sources, we are free from some important information usually required on the sources and which can be difficult to apprehend. For instance, the sources positions for the observations can be not precisely known, and can easily lead to the failure of the algorithm. Similarly, the source wavelet is not initially well known and requires to be approximated (see Subsection 4.6.4). Using $\mathcal{J}_G$ for the cost function implies that those two aspects have no importance anymore, we only need to know the position of the receivers. Furthermore, we can arbitrarily choose the number of sources (and their positions) for the simulations, and increase or decrease the quantity compared to the seismic acquisition used for the observations. Many possibilities follow.

7.3.1 Changing the acquisition set

In order to illustrate the promising potential of the method, we reproduce the multi-frequency experiment with a different acquisition for the simulations and the observations set. Namely, we halve the number of sources compared to the observation set, their positions are changed and we choose a different source wavelet. Those changes are detailed in Table 7.2. The progression in frequency and scale is prescribed in Table 7.1, and we show

the reconstruction after 15 Hz iterations in Figures 7.13 and 7.14. In the latest we employ a smoothing filtering of the final reconstruction, see Remark 7.3.

	acquisition for the observation	**acquisition for the simulation**
n_{src}	160	80
sources depth	100 m	80 m

Table 7.2: Comparison of acquisition setups employed for the observations and simulations. The source wavelet also differs.

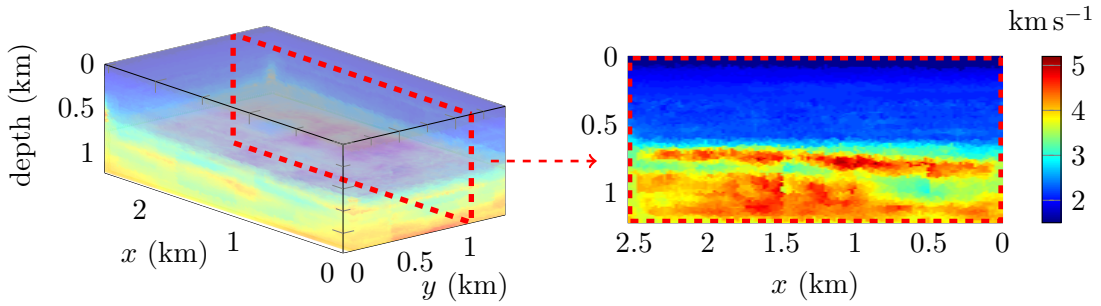


Figure 7.13: Reconstruction of the acoustic 3D Statoil wave speed from Cauchy data using frequencies from 3 to 15 Hz, and the cost function $\mathcal{J}_G$ (Equation (7.4)). The set of sources is different between the observations and the simulations, see Table 7.2. The initial guess is pictured in the Figure 7.9. The reconstruction uses simultaneous progression in frequency and scale, prescribed in Table 7.1.

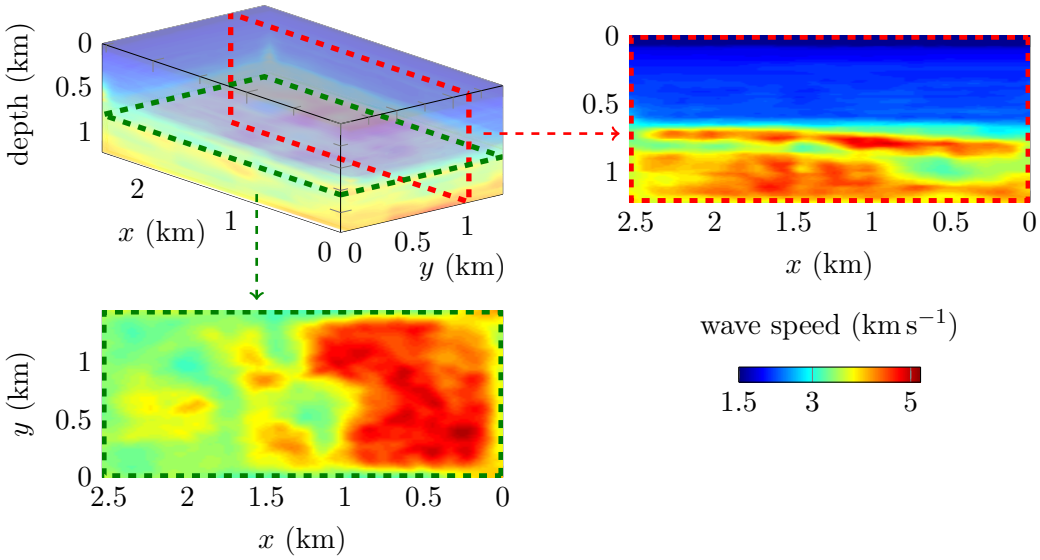


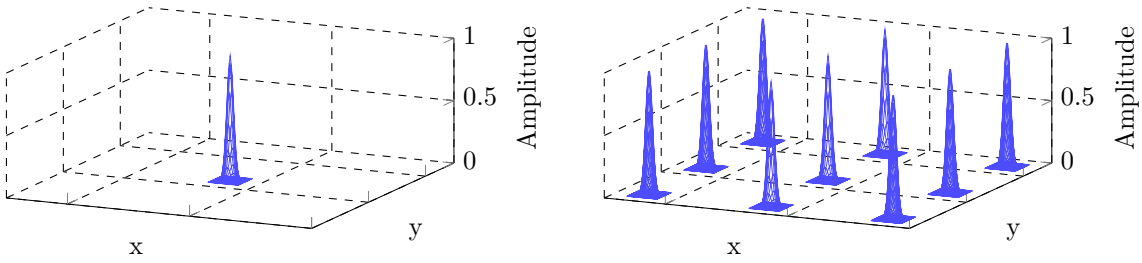
Figure 7.14: Application of a Gaussian smoothing filter (see Remark 7.3) onto the reconstruction of the acoustic 3D Statoil wave speed from Cauchy data using frequencies from 3 to 15 Hz, and the cost function $\mathcal{J}_G$ (Equation (7.4)). The set of sources is different between the observations and the simulations, see Table 7.2.

The reconstruction captures the main variations and accurately recovers the wave speed. We have halved the number of sources in the simulation compared to the ob-

servation, reducing the numerical cost accordingly. Yet we make full use of the observed data, benefiting from the cost function defined for the Cauchy data. Furthermore, we do need to know the position of the sources employed for observation, nor the source wavelet. Eventually, we observe similar accuracy as in the original experiment. This results is inconceivable from the standard least squares where the sources positions and wavelet must be known (or approximated) to expect any recovery.

7.3.2 Artificial simultaneous point sources

We pursue the illustration of the perspectives with another experience where the acquisition for the simulation is further modified. We reduce the number of sources to five. Yet, instead of taking a single point to characterize every source, we consider several points, as if there were simultaneous shots. Every shot consists in sixteen point sources that are simultaneously propagated. Hence we have kept 80 point sources but separated in five groups only. In Figure 7.15, we illustrate the source generated for the observed data (single point-source) and the multi-point source employ for the simulation. The use of this kind of simulations does not require any modification of the data set, and emphasizes the benefit of the cost function adapted to the Cauchy data.



(a) One source consisting of a single shot.

(b) One source consisting of a nine simultaneous shots.

Figure 7.15: Illustration of single or simultaneous point sources. In our experiment, the data are generated from single point source and the simulation employ multi-point source where each source is composed of sixteen shots.

The reconstruction is conducted similarly to in the previous experiment, with frequency from 3 to 15 Hz and with the progression in scale prescribed in Table 7.1. In Figure 7.16, we show the final reconstruction after the use of the 15 Hz data. Figure 7.17 applies a Gaussian smoothing filter for a better visualization, see Remark 7.3.

The drastic reduction of the number of sources is perfectly compensated by the incorporation of simultaneous shots. The final reconstruction shows some limitation on the side, see Figure 7.17, but otherwise it carries the same accuracy as for the previous test.

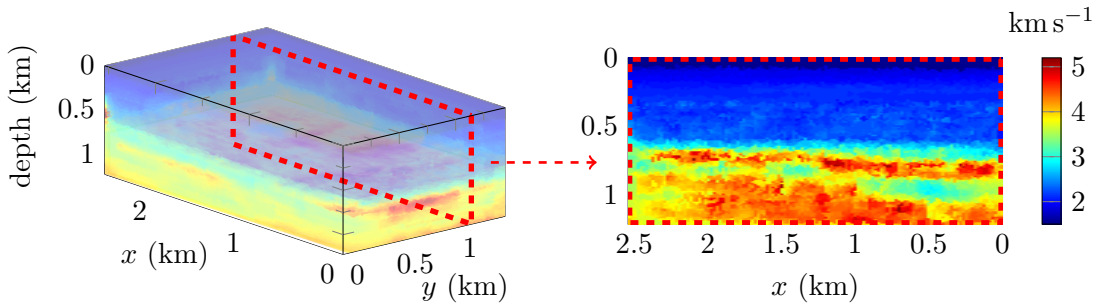


Figure 7.16: Reconstruction of the acoustic 3D Statoil wave speed from Cauchy data using frequencies from 3 to 15 Hz, and the cost function $\mathcal{J}_G$ (Equation (7.4)). The set of sources is different between the observations and the simulations. The simulations use only 5 sources composed of 16 simultaneous shots for every source. The initial guess is pictured in Figure 7.9. The reconstruction uses simultaneous progression in frequency and scale, prescribed in Table 7.1.

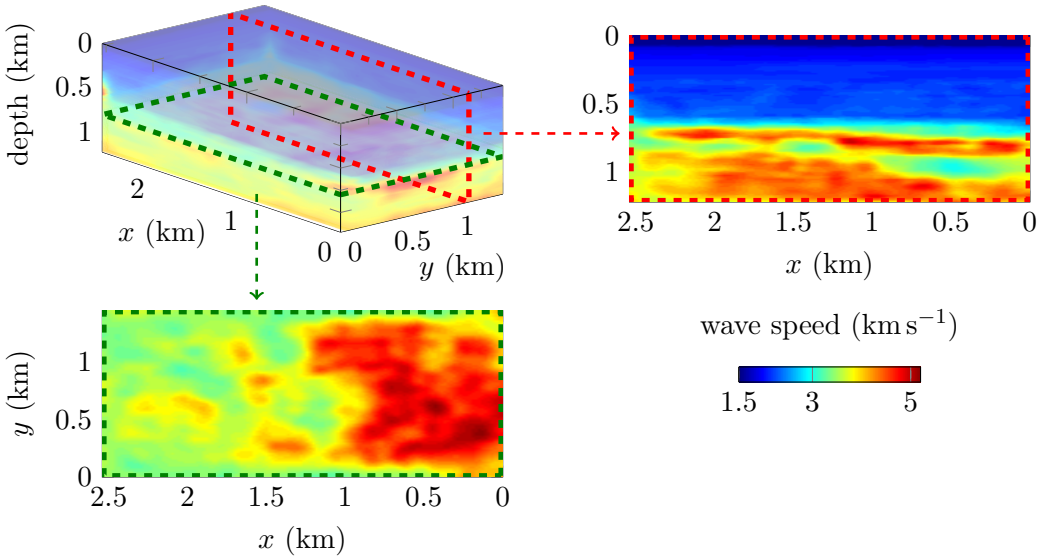


Figure 7.17: Application of a Gaussian smoothing filter (see Remark 7.3) onto the reconstruction of the acoustic 3D Statoil wave speed where the acquisition for the simulations use 5 sources composed of 16 simultaneous shots for every source.

The layer where the velocity decreases is well retrieved and the speed values are accurate. The use of Cauchy data with the cost function $\mathcal{J}_G$ allows a minimal acquisition for the simulation. It also naturally has a large impact onto the computational time. These experiments have all been done with 8 cores, it took about 22 h when all sources are used (Subsection 7.2.3, see Figure 7.10) and 15 h for the hereby test, that uses 5 sources only. Hence, we have, for this simple experiment, a gain of more than 30% in time.

The perspective of differentiating the observations and simulations acquisition sets is a promising extension of this work, and appears consistent with the results of this preliminary experiment. It shall allow less prior on the observational environment, without impacting on the reconstruction.

7.4 Conclusion

We have initiated the reconstruction from Cauchy data, motivated by the dual sensor devices developed in marine seismic acquisitions. Following the conditional stability result, the multi-level algorithm is conducted with piecewise linear model representations. It also provides a new cost function based on the Green's identity. We have performed numerical reconstruction for a three-dimensional acoustic domain starting from time domain data. We have been able to show the recovery from single frequency data and with a limited amount of coefficients to represent the model.

With the new cost function, we have seen that it is possible to differentiate the set of sources employed for the observations and the simulations. We also do not need any information on the source wavelet used for the observation. Those facets are impossible to obtain with the traditional least squares cost function, nor with any alternative relying on the direct difference between observation and simulation. Those perspectives offer a promising road ahead. The selection of the number and position of the sources to employ for the simulation should be investigated, in particular, in association with the multi-point sources acquisition we use.

Then, we would need to confront the behavior in more realistic applications, working with the observed vertical velocity and reduced accuracy data. Moreover, it is not clear if this can be extended for elastic prospecting. Yet it may be of interest for (rotational) seismology, where the source can be complicated to precisely identify; with our method, we can overcome this issue, if we access the two types of data.

Conclusion

In this project, we have developed methodologies to solve the seismic inverse problem for harmonic waves. The first cornerstone is a robust and efficient method to solve the forward problem, because it is intensively used during the process of reconstruction. For this purpose, we have used both finite differences, finite elements, and (hybridizable) discontinuous Galerkin. The latest family certainly provides the more flexibility, and can easily account for the topography, or incorporate knowledge on the geometry. However, large scale harmonic wave problems suffer from the numerical cost involved to factorize the matrix, and there is a need for new techniques.

The inverse problem for the reconstruction of subsurface properties is nonlinear and ill-posed. It is important to remember this fact when addressing its resolution, in order to carefully perform the method. Therefore, the essence of this project is to work hand in hand with the theory and numerical applications. We have studied the stability and convergence of the problem, and emphasized the mechanism of simultaneous progression in frequency and scale. This connection between the two must be extended for a precise quantification of the evolution. Namely, assuming the cost function has been reduced by $\delta\%$, it should be able to provide the next frequency and scale to employ, while maintaining the stability and convergence properties. This is an ongoing investigation. The machinery we have developed relies on piecewise constant (or linear) model representation, motivated by the analytical results. The numerical procedure should be extended to the use of more general wavelets. It can allow more flexibility in the reconstruction.

The inverse problem is solved via an iterative minimization algorithm. Intrinsically, many methods are available to conduct such schemes. The large scale imposes some restrictions, such as the limited information on the Hessian, and we have given examples of reconstructions with and without its utilization. Several perspectives regarding the evolution of the software have been pointed out, and the performance of such techniques should be evaluated. In particular, we can mention: a more precise line search, regularization terms in the cost function, the choice of preconditioner for the Hessian, and the use of level-sets

methods to identify subsurface objects. In addition, the methods must be adapted to the large scale problem and computationally effective. During this project, a toolbox to solve the seismic inverse problem has been developed. Because efforts have been made to allow a generic code, we expect for its evolution and growth in a near future. Eventually, we imagine to extend the seismic oriented algorithm we have developed to other reconstructions, where similar procedure holds. We have in mind the helioseismology, which has its own particularities but where the expertise of seismic can be valuable. The connection and specificities between the two topics are emphasized in Cobden et al. (2015).

The reconstruction from Cauchy data is a promising perspective and illustrate perfectly how analytical understanding and numerical applications can interact. In the context where both the pressure and vertical velocity are acquired, we have provided a novel procedure involving a distinct cost function, arising from the stability analysis. It allows the recovery of the subsurface medium with minimal information on the observation set, and we have initiated the possibilities of using multi-point sources for the simulation, in order to reduce the computational time. The investigation must now be continued, in particular, to confront the method to more realistic data sets.

Conclusion

Dans ce projet, nous avons développé une méthodologie pour résoudre le problème inverse en sismique avec des ondes harmoniques. La première étape est d'avoir une méthode robuste et efficace pour résoudre le problème direct, qui est intensivement utilisée lors de la procédure de reconstruction. Ainsi, nous avons présenté les méthodes utilisées au cours de notre étude: les différences finies, les méthodes de Galerkin continues et discontinues. Cette dernière famille est certainement la plus flexible et permet la considération précise de la topographie et de la géométrie des milieux. Cependant, la taille des matrices créées, dans le domaine fréquentiel, et leur factorisation, peuvent être rédhibitoires en termes de coûts de calcul pour les cas 3D élastiques. Ceci donne d'ailleurs lieu à des recherches actives dans les domaines de la discrétisation des EDPs et de l'algèbre linéaire.

Il est important de rappeler que le problème inverse est non-linéaire et mal posé, afin de mettre en avant les précautions à prendre au cours de sa résolution. L'essence de ce projet est de travailler à la fois sur des notions théoriques et numériques, pour obtenir de la robustesse. Pour ce faire, nous avons étudié la stabilité et la convergence du problème, et en avons déduit une progression simultanée en fréquence et en échelle. Cette connexion naturelle doit maintenant être précisément caractérisée, dans l'optique d'obtenir une méthode où la suite de fréquences et d'échelles est automatique. La représentation du modèle se base sur des fonctions constantes ou linéaires par morceaux, et il serait intéressant d'étendre la mise en place numérique à d'autres types d'ondelettes, pour apporter de possibles améliorations.

Le problème inverse est résolu via une méthode de minimisation itérative. La taille du problème restreint les possibilités, notamment en ce qui concerne la Hessienne, et nous avons illustré son utilisation. Il reste encore de nombreuses possibilités quant à la conduction de l'algorithme. La flexibilité du code développé au cours de cette étude doit permettre son évolution dans un futur proche.

La reconstruction avec les données de Cauchy illustre parfaitement les bénéfices mutuels

entre théorie et application. Nous avons employé une nouvelle fonction coût, qui découle de l'analyse de stabilité. Cela permet en particulier de limiter considérablement les informations nécessaires sur l'acquisition, et de prévoir une réduction des coûts de calcul avec des sources multi-points pour les simulations. Il faut maintenant poursuivre les tests, en particulier sur des données réalistes.

Appendix B

FWI toolbox code developed in `Fortran 90`

During this project, a computational toolbox has been developed to perform seismic imaging in the frequency domain. In this appendix, the different options implemented in the code are reviewed. This code has been developed in `Fortran 90`. It is capable of solving both the direct (modeling) and inverse seismic problems, designed following the Full Waveform Inversion algorithm. Yet the author has not taken action in the development of the discretization of the wave equation methods. Four discretizations techniques are currently operational and have been thoroughly validated in the context of the forward problem: the Finite Differences (FD) scheme has been developed with the work of Wang et al. (2010, 2011, 2012), the Discontinuous Galerkin (DG) and Finite Element (FE) approaches are due to the work of Chaumont-Frelet (2015) and the Hybridizable Discontinuous Galerkin (HDG) approach is the work of Bonnasse-Gahot (2015). Each of the discretization methods requires some code specifications, for example an appropriate adjoint state problem for the computation of the gradient. Here we present the main options that have been developed, without giving the details on those inner routine specificities.

The main difference between the discretizations is that they are not all able to solve any wave equations, we have the following situations:

- FD solve acoustic and elastic isotropic media, in two and three dimensions,
- DG solve acoustic and elastic isotropic media, in two dimensions,
- FE solve acoustic and elastic isotropic media, in three dimensions,
- HDG solve elastic and anisotropic media, in two and three dimensions.

Hence, the HDG discretization is currently the only method to handle a full stiffness tensor.

We note that the purpose of the toolbox is to generate generic developments so that all of the options developed for the full waveform inversion algorithm are available for all discretization methods, and an additional one could easily be incorporated as well.

B.1 Input parameter file

The software works with input parameter files, following the types of architecture developed in Total. The input parameter file is established with keywords, followed by the sign = and the value. Depending on the keywords the argument may be: numbers, strings or logical (`true` or `false`). We can also indicate a list of arguments, where every value is separated by a comma: `,`. For example, the list of frequency is given in the parameter file by:

```
frequency_list=1,2,3,4,5   # the frequency varies from 1 to 5 Hz.
directory_out=./output_directory   # directory for the results.
```

In the parameter file, the # indicates a comment and is not taken into account.

This structure of parameter files has greatly beneficiated from the architecture of Total and allows very clear setup for test cases. We now detail the different options that have been implemented in the code. We review the general options regarding the domain in Section B.2. The options to conduct the inverse scheme are detailed in Section B.3. At the exception of the discretization and choice of solver for the linear system, all of the options that are presented below are selected in the input parameter file and do not require a new compilation of the code. Indeed the objective is to provide a generic toolbox which must be user friendly.

B.2 Acquisition and propagators

B.2.1 Domain of interest

The domain of interest has a cubical shape, it is a rectangle in 2D and a rectangular cuboid in 3D. The domain size is read following a velocity header file where all dimensions are specified. Alternatively one can give the coordinates of the different bounding box. Depending on the discretization, the mesh file must also be given. The physical coefficients are read from different binary files (for the density, the wave speed, etc) yet one can instead specify a constant in the input parameter file to automatically generate homogeneous coefficients. The binary files represent the geophysical coefficients onto a Cartesian grid and are then projected on the mesh in the code. All the different informations concerning the seismic

acquisition (receivers and sources) are retrieved from external files. Then from the input parameter file, one can select specific sources only and discard receivers according to offsets.

The complex frequency (damping parameter and Fourier frequency) are entered as a list, as illustrated above.

B.2.2 Linear algebra

Once the appropriate wave equation has been taken into account (see the options for FD, DG and FE above), one can decide on the resolution of the underlying linear system. We employ direct solvers for the resolution because they are able to obtain the solution for several right-hand sides at the same time. In particular, in seismic several sources are typically considered in the acquisition (namely several hundreds or even thousands). Hence there will be as many right-hand sides created and as many requirement for linear system resolution. The use of direct solvers allows the solution of several right hand sides in parallel, which is more complicated with iterative solvers. During our project we have implemented and tested three solvers:

- *Hsolver*: developed in Wang et al. (2012) and totally adapted to structured discretization (i.e. finite differences). It is in particular completely optimized for frequency domain wave equation. It also incorporates Hierarchically Semiseparable (HSS) methods in order to reduce the amount of memory required for the factorization of the system. It also allows a block low rank approximation to reduce the computational time (at the price of accuracy).
- *GMIG solver*: developed by the Geo-Mathematical Imaging Group at Rice University, it is the extension of the Hsolver but for general geometry. It is still at an early stage but we have seen some promising results.
- *Mumps*: linear solver for sparse matrix, see Amestoy et al. (2001, 2006). It greatly benefits from constant improvements and recently provides block low rank approximation.

We leave the possibility to the user to choose how many right-hand sides (sources) are solved in parallel. It is interesting to select as many as possibly, but because of the memory limitation and the large number of sources involved in seismic, it is not always possible to deal with all sources at the same time.

B.2.3 Modeling

The user can choose to save the results (propagation of wave in the medium) following different formats:

- structured wavefield on a Cartesian grid, it is natural for FD discretization; for FE or DG we project the solution at the degrees of freedom onto a structured representation.
- unstructured wavefield for FE and DG discretizations; the solution is kept on the given mesh and saved in VTK format (Will et al. (2006)).
- The wavefield can also be saved only at the receivers positions. This is in particular useful to generate dataset that can later be used for inversion.

The quantities available to save are the following:

- pressure field for acoustic media,
- displacement or velocity field for elastic media.
- For acoustic and elastic media using DG or FE discretizations, one can also retrieve the derivative with respect to z or the quantities. (Neumann information would require a proper normal derivative, yet if the receiver position is at the boundary or in a constant layer of water for example, this simplification can hold).

In the Listing B.1 we show an example of input parameter file used for the modeling of wave propagation.

```
# 1) Dimension
dimension=3d

# 2) Acquisition information and decision
#   a) acquisition file for src and rcv
index_file=../acquisition/decision.index
#   b) depth of the sources and receivers
src_z=0
rcv_z=20
#   c) shots to be treated
shot_first=1
shot_last=10
shot_step=2
#   d) velocity files
file_vp=my_vp.header
file_vs=my_vs.header
```

```
#    e) density
constant_rho=1.0
#    f) discretization
mesh_file=my_mesh
poly_deg=4          # order of approximation
#    g) frequency list
damping_list=0,1
frequency_list=1,2,3,4,5

# 3) solver information
nb_rhs=5  # number of rhs to be solved in paralel

# 4) results
#    a) out main directory
directory_out=./my_directory_outputs/
#    b) structured grid projection size
dx=10
dy=10
dz=10
#    c) write onto disk
save_wavefield=true    # structured projection
save_receivers=true    # information at the receivers
save_vtk=true          # save wavefield on mesh
```

Listing B.1: Illustration of an input parameter file for the modeling of wave propagation. In this example we deal with a three-dimensional domain. From the input acquisition file, we select sources from 1 to 10, every two shots so that there is a total of five sources: the indexes {1,3,5,7,9}. The frequency varies from 1 to 5 Hz with damping parameter of 0 and 1, i.e. there is a total of 10 complex frequencies.

B.3 Inversion algorithm

Here we review the main options to solve the seismic inverse problem using FWI. The different methods have been presented in Section 4.6 and we only indicate how the toolbox works.

B.3.1 Initial information

Most of the options specified in Section B.2 are also required to process the iterative minimization algorithm for the reconstruction of subsurface properties. In particular, the user needs to specify the domain size and give the initial geophysical quantities following the

same keywords. Similarly the list of frequencies and acquisition setup have to be given. For additional requirement, the user must give the folder where the observed data are recorded.

Regarding the frequency, the user can choose to group frequencies or to follow a sequential progression.

B.3.2 Cost functions

The choice of the misfit functional follows two possibilities:

- standard least squares functional based on $L2$ minimization, Equation (4.8),
- logarithmic functional, Equation (4.7).

Furthermore, one can choose to only work with the real or imaginary part of the data. This is due to the frequency domain data, which are complex signals. Hence there is a total of six different functionals.

Similar options are given for the reconstruction of the source. The method was given for the least squares minimization in Subsection 4.6.4 but one can also specify the update for a logarithmic type functional. Hence we also define six functionals for the source reconstruction, based on the real and imaginary parts of the data as well.

B.3.3 Parametrization

The choice of parameter to be inverted is capital for the proper reconstruction and we have studied their effect in Chapter 5. Because it is yet unclear if an universal parametrization can be employed, we leave many possibilities in the toolbox. The selection is composed of the possibilities invoked in Chapter 5.

B.3.3.1 Acoustic parameterization

Acoustic media involve the reconstruction of the bulk modulus and the density. We have also identified some alternative quantities with the impedance and velocity. Basically, any combination of two parameters between c, κ, I and ρ is possible. In addition two functions can be applied to the selected parameter: the inverse and the inverse squared for the impedance and velocity; the inverse and the inverse of the square root for the bulk modulus

and density. It actually gives many possibilities but we have seen that it is more interesting to leave the density separated so that 27 remains. Furthermore, one can decide not to reconstruct the density to save time.

B.3.3.2 Elastic parameterization

Isotropic elastic wave equations are characterized by three geophysical coefficients: the density ρ and the Lamé parameters λ and μ. Several alternatives have been defined in Chapter 5, Section 5.6 which are:

- Lamé parameters and density (λ, μ, ρ),
- velocities and density (c_p, c_s, ρ),
- Poisson ratio, bulk modulus and density (κ, ν, ρ),
- impedances and density (I_p, I_s, ρ),

For all of the four possibilities, we also consider functions of the parameters: the inverse and the logarithm, for a total of 12 possibilities. Similarly to the acoustic medium, one can also choose not to reconstruct the density, in order to reduce the computational time.

B.3.3.3 TTI parameterization

For the inversion of TTI parameters, we have implemented parametrization based on the stiffness tensor or the Thomsen's parameters. The user can choose between

- stiffness coefficients inversion $(C_{11}, C_{22}, \ldots)$ or their inverse representation $(1/C_{11}, 1/C_{22}, \ldots)$.
- Lamé and Thomsen's parameters $(\lambda, \mu, \epsilon, \delta, \gamma, \theta, \phi, \rho)$ or their inverse.

Any of the parameters listed can be reconstructed or left aside.

B.3.4 Update methods

The choice of the method to conduct the iterative minimization algorithm follows the ones depicted in Chapter 4. In Section 4.6 we have detailed the final algorithm. Let us review those choices:

- gradient method (see Subsection 4.6.1): choice between gradient descent, nonlinear conjugate gradient and L-BFGS.
- Hessian methods (see Subsection 4.6.2): Gauss-Newton approximation or full Hessian method; possibility for regularization parameter on the diagonal.

Finally, the line search is conducted via a simple backtracking algorithm for which only the initial step needs to be given. The user can also decide to have a fixed step all along the iterative minimization.

B.3.5 Other options

The following options can also be defined by the user:

- the offset around the source to limit the receivers data to be included (e.g., to ignore receivers that are too close or too far from the source);
- the maximum and minimum number of iterations per frequency.

After the minimum number of iterations per frequency has been performed, the algorithm compares the benefit of the cost function between the current iteration i and the iteration $i-5$. If the update has been less than 5% (this tolerance can be adjusted) then the frequency is updated as it assumes a stagnation of the convergence.

We also mention the possibility of model reduction through compression, illustrated in Chapter 3. Two aspects can be chosen:

- for the domain decomposition (we can force a structured partition or an automatic one according to neighboring values),
- the choice of model representation: piecewise constant or piecewise linear over each of the resulting subdomain of the partition.

B.4 Concluding remark

One of the objective of this software is to have generic routines so that new techniques, discretization methods or solvers can easily be implemented. We have mentioned the current options, and the goal is to keep this platform evolving in the future, and we have mentioned several techniques of interest regarding the resolution of the inverse problem via iterative minimization.

Bibliography

Ainsworth, M., Monk, P., and Muniz, W. (2006). Dispersive and dissipative properties of discontinuous galerkin finite element methods for the second-order wave equation. *Journal of Scientific Computing*, 27(1):5–40. [p. 53]

Akcelik, V., Biros, G., and Ghattas, O. (2002). Parallel multiscale gauss-newton-krylov methods for inverse wave propagation. In *Supercomputing, ACM/IEEE 2002 Conference*. [p. 123]

Aki, K. and Richards, P. G. (2002). *Quantitative seismology*, volume 1. [p. 16], [p. 17], [p. 20], [p. 21], [p. 43]

Alessandrini, G. (1988). Stable determination of conductivity by boundary measurements. *Appl. Anal.*, 27(1-3):153–172. [p. 80], [p. 94], [p. 105]

Alessandrini, G., de Hoop, M. V., Faucher, F., Gaburro, R., and Sincich, E. (2017a). Acoustic Full Waveform Inversion from Cauchy data via conditional well-posedness driven iterative regularization. In *Proceedings of the Project Review, Geo-Mathematical Imaging Group*, volume 1, Houston, United States. [p. 303]

Alessandrini, G., de Hoop, M. V., Gaburro, R., and Sincich, E. (2017b). Lipschitz stability for a piecewise linear Schrödinger potential from local Cauchy data. In *Proceedings of the Project Review, Geo-Mathematical Imaging Group*, volume 1, Houston, United States. [p. 303], [p. 305], [p. 308]

Alessandrini, G., de Hoop, M. V., Gaburro, R., and Sincich, E. (2017c). Lipschitz stability for a piecewise linear Schrödinger potential from local cauchy data. *arXiv preprint arXiv:1702.04222*. [p. 305], [p. 308]

Alessandrini, G. and Vessella, S. (2005). Lipschitz stability for the inverse conductivity problem. *Adv. in Appl. Math.*, 35(2):207–241. [p. 80], [p. 81], [p. 94]

Alkhalifah, T. and Plessix, R.-É. (2014). A recipe for practical full-waveform inversion in anisotropic media: An analytical parameter resolution study. *Geophysics*, 79(3):R91–R101. [p. 37], [p. 124]

Allaire, G. (2012). *Analyse numérique et optimisation: une introduction à la modélisation mathématique et à la simulation numérique*. Editions Ecole Polytechnique, second edition. Other version: 2005. [p. 53], [p. 125], [p. 142], [p. 143]

Amestoy, P. R., Duff, I. S., L'Excellent, J.-Y., and Koster, J. (2001). A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM Journal on Matrix Analysis and Applications*, 23(1):15–41. [p. 69], [p. 331]

Amestoy, P. R., Guermouche, A., L'Excellent, J.-Y., and Pralet, S. (2006). Hybrid scheduling for the parallel solution of linear systems. *Parallel computing*, 32(2):136–156. [p. 69], [p. 331]

Aminzadeh, F., Burkhard, N., Nicoletis, L., Rocca, F., and Wyatt, K. (1994). Seg/eaeg 3-d modeling project: 2nd update. *The Leading Edge*, 13(9):949–952. [p. 174], [p. 251]

Armijo, L. (1966). Minimization of functions having lipschitz continuous first partial derivatives. *Pacific Journal of mathematics*, 16(1):1–3. [p. 162]

Arnold, D. N., Brezzi, F., Cockburn, B., and Marini, L. D. (2002). Unified analysis of discontinuous galerkin methods for elliptic problems. *SIAM journal on numerical analysis*, 39(5):1749–1779. [p. 53]

Askan, A., Akcelik, V., Bielak, J., and Ghattas, O. (2007). Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures. *Bulletin of the Seismological Society of America*, 97(6):1990–2008. [p. 124]

Baeten, G. J. M. (1989). *Theoretical and practical aspects of the Vibroseis method*. PhD thesis, Technische Universiteit Delft. [p. 46], [p. 131]

Bal, G. (2012). Introduction to inverse problems. Lecture notes, Columbia University, New York. [p. 75], [p. 77], [p. 80]

Bamberger, A., Chavent, G., and Lailly, P. (1977). Une application de la théorie du contrôle à un problème inverse de sismique. *Annales de Géophysique*, 33:183–200. [p. 122]

Bamberger, A., Chavent, G., and Lailly, P. (1979). About the stability of the inverse problem in the 1-d wave equation. *Journal of Applied Mathematics and Optimisation*, 5:1–47. [p. 122]

Bao, G. and Li, P. (2005a). Inverse medium scattering for the helmholtz equation at fixed frequency. *Inverse Problems*, 21(5):1621. [p. 124]

Bao, G. and Li, P. (2005b). Inverse medium scattering problems for electromagnetic waves. *SIAM J. Appl. Math.*, 65(6):2049–2066 (electronic). [p. 94]

Bao, G. and Li, P. (2009). Numerical solution of an inverse medium scattering problem for Maxwell's equations at fixed frequency. *Journal of Computational Physics*, 228(12):4638–4648. [p. 124]

Bao, G. and Triki, F. (2010). Error estimates for the recursive linearization of inverse medium problems. *J. Comput. Math.*, 28(6):725–744. [p. 94]

Barucq, H., Calandra, H., Pham, H., and Tordeux, S. (2017). A study of the Numerical Dispersion for the Continuous Galerkin discretization of the one-dimensional Helmholtz equation. Research Report RR-9075, Inria Bordeaux Sud-Ouest ; Magique 3D. [p. 250]

Barucq, H., Chabassier, J., Pham, H., and Tordeux, S. (2016). A study of the numerical robustness of single-layer method with Fourier basis for multiple obstacle scattering in homogeneous media. Research Report RR-8988, Inria Bordeaux Sud-Ouest. [p. 53]

Bednar, J. B., Shin, C., and Pyun, S. (2007). Comparison of waveform inversion, part 2: phase approach. *Geophysical Prospecting*, 55(4):465–475. [p. 133]

Ben-Hadj-Ali, H., Operto, S., and Virieux, J. (2008). Velocity model-building by 3d frequency-domain, full-waveform inversion of wide-aperture seismic data. *Geophysics*, 73. [p. 94]

Bérenger, J.-P. (1994). A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114(2):185 – 200. [p. 67]

Beretta, E., de Hoop, M. V., Faucher, F., and Scherzer, O. (2016). Inverse boundary value problem for the helmholtz equation: quantitative conditional lipschitz stability estimates. *SIAM Journal on Mathematical Analysis*, 48(6):3962–3983. [p. 76]

Beretta, E., de Hoop, M. V., Francini, E., Vessella, S., and Zhai, J. (2014). Uniqueness and Lipschitz stability of an inverse boundary value problem for time-harmonic elastic waves. *arXiv preprint arXiv:1412.3465*. [p. 77], [p. 85], [p. 210]

Beretta, E., de Hoop, M. V., and Qiu, L. (2013a). Lipschitz stability of an inverse boundary value problem for a Schrödinger-type equation. *SIAM J. Math. Anal.*, 45(2):679–699. [p. 94], [p. 109], [p. 110]

Beretta, E. and Francini, E. (2011). Lipschitz stability for the electrical impedance tomography problem: the complex case. *Comm. Partial Differential Equations*, 36(10):1723–1749. [p. 80], [p. 94]

Beretta, E., Francini, E., and Vessella, S. (2013b). Uniqueness and lipschitz stability for the identification of Lamé parameters from boundary measurements. *arXiv preprint arXiv:1303.2443*. [p. 77], [p. 85]

Biondi, B. L. (2006). *3D seismic imaging*. Society of Exploration Geophysicists. [p. 122]

Blazek, K. D., Stolk, C., and Symes, W. W. (2013). A mathematical framework for inverse wave problems in heterogeneous media. *Inverse Problems*, 29(6):065001. [p. 95]

Bond, W. L. (1943). The mathematics of the physical properties of crystals. *Bell Labs Technical Journal*, 22(1):1–72. [p. 30]

Bonnasse-Gahot, M. (2015). *Simulation de la propagation d'ondes élastiques en domaine fréquentiel par des méthodes Galerkine discontinues*. PhD thesis, Université Nice Sophia Antipolis. [p. 54], [p. 67], [p. 70], [p. 146], [p. 147], [p. 329]

Bos, L., Gibson, P., Kotchetov, M., and Slawinski, M. (2004). Classes of anisotropic media: a tutorial. *Studia Geophysica et Geodaetica*, 48(1):265–287. [p. 28]

Bourgeois, L. (2013). A remark on lipschitz stability for inverse problems. *Comptes Rendus Mathematique*, 351(5-6):187–190. [p. 80]

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. [p. 133]

Brandsberg-Dahl, S., Chemingui, N., Valenciano, A., Ramos-Martinez, J., and Qiu, L. (2017). Fwi for model updates in large-contrast media. *The Leading Edge*, 36(1):81–87. [p. 301]

Brandwood, D. (1983). A complex gradient operator and its application in adaptive array theory. In *IEE Proceedings F-Communications, Radar and Signal Processing*, volume 130, pages 11–16. IET. [p. 138], [p. 139]

Brenders, A. J., Albertin, U., and Mika, J. (2012). Comparison of 3d time-and frequency-domain waveform inversion: benefits and insights of a broadband, discrete-frequency strategy. In *SEG Technical Program Expanded Abstracts 2012*, pages 1–5. Society of Exploration Geophysicists. [p. 198]

Brossier, R. (2009). *Imagerie sismique à deux dimensions des milieux viso-élastiques par inversion des formes d'ondes : développements méthodologiques et applications*. PhD thesis, Universit'e de Nice-Sophia Antipolis. [p. 233]

Brossier, R. (2011). Two-dimensional frequency-domain visco-elastic full waveform inversion: Parallel algorithms, optimization and performance. *Computers & Geosciences*, 37(4):444–455. [p. 210]

Brossier, R., Etienne, V., Operto, S., and Virieux, J. (2010a). Frequency-domain numerical modelling of visco-acoustic waves based on finite-difference and finite-element discontinuous galerkin methods. In *Acoustic waves*. InTech. [p. 53]

Brossier, R., Operto, S., and Virieux, J. (2010b). Which data residual norm for robust elastic frequency-domain full waveform inversion? *GEOPHYSICS*, 75(3):R37–R46. [p. 133]

Bunks, C., Saleck, F. M., Zaleski, S., and Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473. [p. 124], [p. 134], [p. 194], [p. 198]

Calderón, A. P. (1980). On an inverse boundary value problem. *Seminar on Numerical Analysis and its Applications to Continuum Physics*, pages 65–73. [p. 76], [p. 78]

Carlson, D., Söllner, W., Tabti, H., Brox, E., and Widmaier, M. (2007). Increased resolution of seismic data from a dual-sensor streamer cable. In *SEG Technical Program Expanded Abstracts 2007*, pages 994–998. Society of Exploration Geophysicists. [p. 130], [p. 305], [p. 311]

Cartan, H. (1971). *Differential calculus*, volume 1. Hermann. [p. 191]

Chaumont-Frelet, T. (2015). *Finite element approximation of Helmholtz problems with application to seismic wave propagation*. PhD thesis, INSA de Rouen. [p. 54], [p. 329]

Chavent, G. (1974). Identification of functional parameters in partial differential equations. In Goodson, R. E. and Polis, M., editors, *Identification of Parameters in Distributed Systems*, pages 31–48. ASME, New York. [p. 122], [p. 142]

Chavent, G. (2010). *Nonlinear least squares for inverse problems: theoretical foundations and step-by-step guide for applications*. Springer Science & Business Media. [p. 123], [p. 125], [p. 142], [p. 150], [p. 163], [p. 186], [p. 187], [p. 188], [p. 189]

Chavent, G. and Clément, F. (1992). *Waveform inversion through MBTT formulation*. INRIA. [p. 300]

Chavent, G. and Kunisch, K. (1996). On weakly nonlinear inverse problems. *SIAM Journal on Applied Mathematics*, 56(2):542–572. [p. 186], [p. 189]

Chen, P., Jordan, T. H., and Zhao, L. (2007). Full three-dimensional tomography: a comparison between the scattering-integral and adjoint-wavefield methods. *Geophysical Journal International*, 170(1):175–181. [p. 123]

Choi, Y., Min, D.-J., and Shin, C. (2008). Frequency-domain elastic full waveform inversion using the new pseudo-hessian matrix: Experience of elastic marmousi-2 synthetic data. *Bulletin of the Seismological Society of America*, 98(5):2402–2415. [p. 123], [p. 301]

Claerbout, J. F. (1971). Toward a unified theory of reflector mapping. *Geophysics*, 36(3):467–481. [p. 122]

Claerbout, J. F. (1985). Imaging the Earth's interior. [p. 122]

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. [p. 300]

Cobden, L. J., Fichtner, A., and Tong, V. (2015). Waveform tomography in geophysics and helioseismology. [p. 326]

Cohen, G. C. (2003). Higher-order numerical methods for transient wave equations. [p. 53]

Cohen, G. C. and Pernet, S. (2017). *Finite element and discontinuous Galerkin methods for transient wave equations*. Springer. [p. 53]

Colton, D. and Kress, R. (1998). *Inverse acoustic and electromagnetic scattering theory*. Applied mathematical sciences. Springer, New York, second edition. [p. 36], [p. 38], [p. 47], [p. 75]

Conn, A. R., Gould, N. I. M., and Toint, P. L. (2000). *Trust Region Methods*. Society for Industrial and Applied Mathematics. [p. 123]

Courant, R., Friedrichs, K., and Lewy, H. (1928). Über die partiellen differenzengleichungen der mathematischen physik. *Mathematische Annalen*, 100(1):32–74. [p. 53]

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. [p. 122]

Dahlen, F. A. and Tromp, J. (1998). *Theoretical global seismology*. Princeton, N.J. Princeton University Press. [p. 20]

Dai, Y.-H. and Yuan, Y. (1999). A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on optimization*, 10(1):177–182. [p. 160]

Datchev, K. and de Hoop, M. V. (2015). Iterative reconstruction of the wavespeed for the wave equation with bounded frequency boundary data. *arXiv preprint arXiv:1506.09014*. [p. 77], [p. 95]

Davies, E. B. (1995). *Spectral theory and differential operators*, volume 42 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge. [p. 98], [p. 99]

de Hoop, M. V., Qiu, L., and Scherzer, O. (2012). A convergence analysis of a multi-level projected steepest descent iteration for nonlinear inverse problems in banach spaces subject to stability constraints. *arXiv preprint arXiv:1206.3706*. [p. 123], [p. 160], [p. 163]

de Hoop, M. V., Spencer, C., and Burridge, R. (1999). The resolving power of seismic amplitude data; an anisotropic inversion/migration approach. *Geophysics*, 64(3):852–873. [p. 210]

Demanet, L. (2014). Waves and imaging. Lecture notes, Massachusetts Institute of Technology. [p. 36]

Diaz, J. (2005). *Approches analytiques et numériques de problèmes de transmission en propagation d'ondes en régime transitoire. Application au couplage fluide-structure et aux méthodes de couches parfaitement adaptées.* PhD thesis, ENSTA ParisTech. [p. 67]

Duveneck, E., Milcik, P., Bakker, P. M., and Perkins, C. (2008). Acoustic vti wave equations and their application for anisotropic reverse-time migration. In *SEG Technical Program Expanded Abstracts 2008*, pages 2186–2190. Society of Exploration Geophysicists. [p. 37]

Eisenstat, S. C. and Walker, H. F. (1994). Globally convergent inexact newton methods. *SIAM Journal on Optimization*, 4:393–422. [p. 123]

Engquist, B. and Majda, A. (1977). Absorbing boundary conditions for numerical simulation of waves. *Proceedings of the National Academy of Sciences*, 74(5):1765–1766. [p. 68], [p. 306]

Epanomeritakis, I., Akçelik, V., Ghattas, O., and Bielak, J. (2008). A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion. *Inverse Problems*, 24(3):034015. [p. 123]

Feldman, J., Salo, M., and Uhlmann, G. (2015). *The Calderón problem — An Introduction to Inverse Problems.* Unpublished edition. [p. 76], [p. 105], [p. 106]

Felício Fuck, R. and Tsvankin, I. (2009). Analysis of the symmetry of a stressed medium using nonlinear elasticity. *Geophysics*, 74(5):WB79–WB87. [p. 28]

Fichtner, A. and Trampert, J. (2011). Hessian kernels of seismic data functionals based upon adjoint techniques. *Geophysical Journal International*, 185(2):775–798. [p. 124], [p. 149]

Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y., and Villaseñor, A. (2013). Multiscale full waveform inversion. *Geophysical Journal International*, 194(1):534–556. [p. 125]

Fletcher, R. and Reeves, C. M. (1964). Function minimization by conjugate gradients. *The computer journal*, 7(2):149–154. [p. 160]

Futterman, W. (1962). Dispersive body waves. *Journal of Geophysical Research*, 67:5279–5291. [p. 43]

Gardner, G., Gardner, L., and Gregory, A. (1974). Formation velocity and density–The diagnostic basics for stratigraphic traps. *Geophysics*, 39(6):770–780. [p. 225]

Gauthier, O., Virieux, J., and Tarantola, A. (1986). Two-dimensional nonlinear inversion of seismic waveforms; numerical results. *Geophysics*, 51(7):1387–1403. [p. 122], [p. 123]

Gélis, C., Virieux, J., and Grandjean, G. (2007). Two-dimensional elastic full waveform inversion using Born and Rytov formulations in the frequency domain. *Geophysical Journal International*, 168(2):605–633. [p. 123], [p. 232], [p. 233], [p. 289]

Gilbarg, D. and Trudinger, N. S. (1983). *Elliptic partial differential equations of second order*, volume 224 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition. [p. 96]

Green, G. (1828). *An essay on the application of mathematical analysis to the theories of electricity and magnetism*. author. [p. 20]

Guitton, A. and Alkhalifah, T. (2017). A parameterization study for elastic vti full waveform inversion of hydrophone components: synthetic and north sea field data examples. *Geophysics*, 82(6):1–70. [p. 124], [p. 210]

Ha, W., Pyun, S., Yoo, J., and Shin, C. (2010). Acoustic full waveform inversion of synthetic land and marine data in the laplace domain. *Geophysical Prospecting*, 58(6):1033–1048. [p. 123], [p. 124], [p. 164]

Haar, A. (1911). Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 71:38–53. [p. 87]

Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin*, pages 49–52. [p. 47], [p. 75]

Hadamard, J. (1923). *Lecture's on Cauchy's problem in linear partial differential equations*. Yale University press. [p. 47], [p. 75]

Hähner, P. (1996). A periodic Faddeev-type solution operator. *J. Differential Equations*, 128(1):300–308. [p. 105]

Hähner, P. and Hohage, T. (2001). New stability estimates for the inverse acoustic inhomogeneous medium problem and applications. *SIAM Journal on Mathematical Analysis*, 33(3):670–685. [p. 94]

Hanke, M. (2014). A note on the nonlinear Landweber iteration. *Numerical Functional Analysis and Optimization*, 35(11):1500–1510. [p. 160]

Henkin, G. and Novikov, R. (1988). A multidimensional inverse problem in quantum and acoustic scattering. *Inverse problems*, 4(1):103. [p. 76]

Hestenes, M. R. and Stiefel, E. (1952). *Methods of conjugate gradients for solving linear systems*, volume 49. NBS. [p. 160]

Hesthaven, J. S. and Warburton, T. (2007). *Nodal discontinuous Galerkin methods: algorithms, analysis, and applications*. Springer Science & Business Media. [p. 53]

Hohage, T. (1997). Logarithmic convergence rates of the iteratively regularized gauss - newton method for an inverse potential and an inverse scattering problem. *Inverse Problems*, 13(5):1279. [p. 94]

Hustedt, B., Operto, S., and Virieux, J. (2004). Mixed-grid and staggered grid finite difference methods for frequency-domain acoustic wave modelling. *Geophysical Journal International*, 157(3):1269–1296. Geophysical Journal International, v. 157, n. 3, p. 1269-1296, 2004. http://dx.doi.org/10.1111/j.1365-246X.2004.02289.x. [p. 59]

Ihlenburg, F. (2006). *Finite element analysis of acoustic scattering*, volume 132. Springer Science & Business Media. [p. 53]

Imanuvilov, O., Uhlmann, G., and Yamamoto, M. (2010). The Calderón problem with partial data in two dimensions. *Journal of the American Mathematical Society*, 23(3):655–691. [p. 76]

Isakov, V. (2006). *Inverse problems for partial differential equations*, volume 127. Springer. [p. 75], [p. 76]

Iversen, E. and Pšenčík, I. (2007). Ray tracing for continuously rotated local coordinates belonging to a specified anisotropy. *Studia Geophysica et geodaetica*, 51(1):37–58. [p. 30]

Jeong, W., Lee, H.-Y., and Min, D.-J. (2012). Full waveform inversion strategy for density in the frequency domain. *Geophysical Journal International*, 188:1221–1241. [p. 210], [p. 225]

Jo, C.-H., Shin, C., and Suh, J. H. (1996). An optimal 9-point, finite-difference, frequency-space, 2-d scalar wave extrapolator. *Geophysics*, 61(2):529–537. [p. 58], [p. 59]

Jun, H., Kim, Y., Shin, J., and Shin, C. (2013). 2d elastic time-laplace-fourier-domain hybrid full waveform inversion. In *SEG Technical Program Expanded Abstracts 2013*, pages 1008–1013. Society of Exploration Geophysicists. [p. 122]

Jun, H., Park, E., and Shin, C. (2015). Weighted pseudo-hessian for frequency-domain elastic full waveform inversion. *Journal of Applied Geophysics*, 123:1–17. [p. 301]

Kadu, A., Van Leeuwen, T., and Mulder, W. (2017). Parametric level-set full-waveform inversion in the presence of salt bodies. In *SEG Technical Program Expanded Abstracts 2017*, pages 1518–1522. Society of Exploration Geophysicists. [p. 300]

Kamath, N. and Tsvankin, I. (2016). Elastic full-waveform inversion for vti media: Methodology and sensitivity analysis. *Geophysics*, 81(2):C53–C68. [p. 210]

Keller, J. B. (1976). Inverse problems. *The American Mathematical Monthly*, 83(2):107–118. [p. 75]

Kenig, C. E., Sjöstrand, J., and Uhlmann, G. (2007). The Calderón problem with partial data. *Annals of mathematics*, pages 567–591. [p. 76]

Kennett, B., Sambridge, M., and Williamson, P. (1988). Subspace methods for large inverse problems with multiple parameter classes. *Geophysical Journal International*, 94(2):237–247. [p. 124]

Kern, M. (2016). *Numerical Methods for Inverse Problems*. John Wiley & Sons. [p. 125], [p. 137], [p. 142], [p. 150], [p. 151], [p. 171]

Kirsch, A. (1996). *An introduction to the mathematical theory of inverse problems*. Applied mathematical sciences. Springer, New York. [p. 36], [p. 47], [p. 75], [p. 76], [p. 160]

Kjartansson, E. (1979). Constant q-wave propagation and attenuation. *Journal of Geophysical Research: Solid Earth*, 84(B9):4737–4748. [p. 43]

Köhn, D., De Nil, D., Kurzmann, A., Przebindowska, A., and Bohlen, T. (2012). On the influence of model parametrization in elastic full waveform tomography. *Geophysical Journal International*, 191(1):325–345. [p. 210]

Kolsky, H. (1956). The propagation of stress pulses in viscoelastic solids. *Philosophical magazine*, 1:693–710. [p. 43]

Körnig, M. and Müller, G. (1989). Rheological models and interpretation of postglacial uplift. *Geophysical Journal International*, 98(2):243–253. [p. 44]

Kreutz-Delgado, K. (2009). The complex gradient operator and the cr-calculus. *arXiv preprint arXiv:0906.4835*. [p. 138]

Kupradze, V., Gegelia, T., Basheleishvili, M., and Burchuladze, T. V. (1976). Three-dimensional problems of the mathematical theory of elasticity and thermoelasticity. [p. 35], [p. 48]

Kupradze, V. D. (1963). *Progress in solid mechanics. 3. Dynamical problems in elasticity*. North-Holland Publishing Company. [p. 48]

Lailly, P. (1983). The seismic inverse problem as a sequence of before stack migrations. In Bednar, J. B., editor, *Conference on Inverse Scattering: Theory and Application*, pages 206–220. Society for Industrial and Applied Mathematics. [p. 122], [p. 142]

Landa, E. and Treitel, S. (2016). Seismic inversion: What it is, and what it is not. *The Leading Edge*. [p. 136], [p. 186]

Landweber, L. (1951). An iteration formula for fredholm integral equations of the first kind. *American Journal of Mathematics*, 73(3):615–624. [p. 160]

Lavaud, B., Kabir, N., and Chavent, G. (1999). Pushing avo inversion beyond linearized approximation. *Journal of Seismic Exploration*, 8(3):279–302. [p. 210]

Le Dimet, F.-X. and Shutyaev, V. (2000). On Newton methods in data assimilation. *Russian Journal of Numerical Analysis and Mathematical Modelling*, 15(5):419–434. [p. 157]

Lewis, W. and Vigh, D. (2016). 3d salt geometry inversion in full-waveform inversion using a level-set method. In *SEG Technical Program Expanded Abstracts 2016*, pages 1221–1226. Society of Exploration Geophysicists. [p. 300]

Li, H. and Adali, T. (2006). Optimization in the complex domain for nonlinear adaptive filtering. In *Signals, Systems and Computers, 2006. ACSSC'06. Fortieth Asilomar Conference on*, pages 263–267. IEEE. [p. 138]

Liao, Q. and McMechan, G. A. (1996). Multifrequency viscoacoustic modeling and inversion. *Geophysics*, 61(5):1371–1378. [p. 43]

Lin, Y., Abubakar, A., Habashy, T., et al. (2012). Seismic full-waveform inversion using truncated wavelet representations. In *2012 SEG Annual Meeting*. Society of Exploration Geophysicists. [p. 124]

Lions, J. L. and Mitter, S. K. (1971). *Optimal control of systems governed by partial differential equations*, volume 1200. Springer Berlin. [p. 142]

Loris, I., Douma, H., Nolet, G., Daubechies, I., and Regone, C. (2010). Nonlinear regularization techniques for seismic tomography. *Journal of Computational Physics*, 229:890–905. [p. 90], [p. 124]

Loris, I., Nolet, G., Daubechies, I., and Dahlen, F. (2007). Tomographic inversion using l1-norm regularization of wavelet coefficients. *Geophysical Journal International*, 170:359–370. [p. 90], [p. 124]

Ma, Y. and Hale, D. (2012). Quasi-Newton full-waveform inversion with a projected Hessian matrix. *Geophysics*, 77(5):R207–R216. [p. 123]

Madariaga, R. (1976). Dynamics of an expanding circular fault. *Bulletin of the Seismological Society of America*, 66(3):639–666. [p. 58]

Magnanini, R. and Papi, G. (1985). An inverse problem for the Helmholtz equation. *Inverse Problems*, 1(4):357–370. [p. 104]

Malinowski, M., Operto, S., and Ribodetti, A. (2011). High-resolution seismic attenuation imaging from wide-aperture onshore data by visco-acoustic frequency-domain full-waveform inversion. *Geophysical Journal International*, 186(3):1179–1204. [p. 44], [p. 124]

Mallat, S. (2008). *A wavelet tour of signal processing*. Academic press. [p. 90]

Mandache, N. (2001). Exponential instability in an inverse problem for the Schrödinger equation. *Inverse Problems*, 17(5):1435–1444. [p. 79], [p. 80], [p. 94]

Marfurt, K. J. (1984). Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations. *Geophysics*, 49(5):533–549. [p. 59]

Martin, G. S. (2004). *The Marmousi2 model, elastic synthetic data, and an analysis of imaging and AVO in a structurally complex environment*. PhD thesis, University of Houston. [p. 166], [p. 244]

Martin, G. S., Wiley, R., and Marfurt, K. J. (2006). Marmousi2: An elastic upgrade for marmousi. *The Leading Edge*, 25(2):156–166. [p. 166], [p. 177], [p. 286]

Martin, P. A. (2006). *Multiple scattering: interaction of time-harmonic waves with N obstacles*, volume 107. Cambridge University Press. [p. 47], [p. 53]

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J. (2016). Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, 205(1):345–377. [p. 301]

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. [p. 123], [p. 149], [p. 151], [p. 157]

Monk, P. (2003). *Finite element methods for Maxwell's equations*. Oxford University Press. [p. 53]

Mora, P. (1987). Nonlinear two-dimensional elastic inversion of multioffset seismic data. *Geophysics*, 52(9):1211–1228. [p. 122], [p. 123]

Muller, G. (1983). Rheological properties and velocity dispersion of a medium with power-law dependence of q on frequency. *Journal of Geophysics-Zeitschrift fur Geophysik*, 54(1):20–29. [p. 43], [p. 44]

Nachman, A. I. (1988). Reconstructions from boundary measurements. *Annals of Mathematics*, 128(3):531–576. [p. 76], [p. 79]

Nachman, A. I. (1996). Global uniqueness for a two-dimensional inverse boundary value problem. *Ann. of Math. (2)*, 143(1):71–96. [p. 76], [p. 79]

Nagayasu, S., Uhlmann, G., and Wang, J.-N. (2013). Increasing stability in an inverse problem for the acoustic equation. *Inverse Problems*, 29(2):025012. [p. 83], [p. 94]

Nakamura, G. and Uhlmann, G. (1993). Identification of lamé parameters by boundary measurements. *American Journal of Mathematics*, pages 1161–1187. [p. 76]

Nakamura, G. and Uhlmann, G. (1994). Global uniqueness for an inverse boundary problem arising in elasticity. *Inventiones mathematicae*, 118(1):457–474. [p. 76]

Nakamura, G. and Uhlmann, G. (2003). Global uniqueness for an inverse boundary value problem arising in elasticity. *Inventiones mathematicae*, 152(1):205–207. [p. 76]

Nataf, F. (2013). Absorbing boundary conditions and perfectly matched layers in wave propagation problems. [p. 67]

Newton, I. (1687). *Philosophiae naturalis principia mathematica*. J. Societatis Regiae ac Typis J. Streater. [p. 20], [p. 21]

Nocedal, J. (1980). Updating quasi-Newton matrices with limited storage. *Mathematics of computation*, 35(151):773–782. [p. 160]

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer Sries in Operations Research, 2 edition. [p. 123], [p. 125], [p. 135], [p. 149], [p. 157], [p. 159], [p. 163]

Novikov, R. G. (1986). Reconstruction of a two-dimensional Schrödinger operator from the scattering amplitude for fixed energy. *Functional Analysis and Its Applications*, 20(3):246–248. [p. 76]

Novikov, R. G. (1988). Multidimensional inverse spectral problem for the equation $-\Delta\psi + (v(x) - Eu(x))\psi = 0$. *Functional Analysis and Its Applications*, 22(4):263–272. [p. 76]

Novikov, R. G. (1994). The inverse scattering problem at fixed energy for the three-dimensional Schrödinger equation with an exponentially decreasing potential. *Communications in Mathematical Physics*, 161(3):569–595. [p. 76]

Novikov, R. G. (2004). Formulae and equations for finding scattering data from the dirichlet-to-neumann map with nonzero background potential. *Inverse Problems*, 21(1):257. [p. 76]

Novikov, R. G. (2010). New global stability estimates for the Gel'fand–Calderón inverse problem. *Inverse problems*, 27(1):015001. [p. 76]

Novikov, R. G. (2011). New global stability estimates for the Gel'fand-Calderon inverse problem. *Inverse Problems*, 27(1):015001. [p. 76], [p. 94]

Operto, S., Virieux, J., Amestoy, P., L'Excellent, J.-Y., Giraud, L., and Ben Hadj Ali, 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 à SM211. [p. 59]

Operto, S., Virieux, J., Ribodetti, A., and Anderson, J. (2009). Finite-difference frequency-domain modeling of viscoacoustic wave propagation in 2d titled transversely isotropic (tti) media. *Geophysics*, 74:75–95. [p. 59]

Pan, G. S., Phinney, R. A., and Odom, R. I. (1988). Full-waveform inversion of plane-wave seismograms in stratified acoustic media; theory and feasibility. *Geophysics*, 53(1):21–31. [p. 122]

Pekeris, C. L. (1934). An inverse boundary value problem in seismology. *Physics*, 5(10):307–316. [p. 75]

Petrov, P. V. and Newman, G. A. (2014). Three-dimensional inverse modelling of damped elastic wave propagation in the fourier domain. *Geophysical Journal International*, 198(3):1599–1617. [p. 164], [p. 198], [p. 234]

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. [p. 142]

Plessix, R.-E. and Cao, Q. (2011). A parametrization study for surface seismic full waveform inversion in an acoustic vertical transversely isotropic medium. *Geophysical Journal International*, 185(1):539–556. [p. 37], [p. 124]

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. [p. 160]

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. [p. 163]

Pratt, R. G. and Goulty, N. R. (1991). Combining wave-equation imaging with traveltime tomography to form high-resolution images from crosshole data. *Geophysics*, 56(2):208–224. [p. 122]

Pratt, R. G., Shin, C., and Hick, G. J. (1998). Gauss–newton and full newton methods in frequency–space seismic waveform inversion. *Geophysical Journal International*, 133(2):341–362. [p. 116], [p. 123], [p. 149], [p. 150], [p. 159], [p. 210], [p. 214]

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(2):323–340. [p. 122], [p. 194]

Pratt, R. G. and Worthington, M. H. (1990). Inverse theory applied to multi-source cross-hole tomography. *Geophysical Prospecting*, 38(3):287–310. [p. 122]

Prieux, V., Brossier, R., Operto, S., and Virieux, J. (2013a). Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the valhall field. part 1: Imaging compressional wave speed, density and attenuation. *Geophysical Journal International*, page ggt177. [p. 210]

Prieux, V., Brossier, R., Operto, S., and Virieux, J. (2013b). Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the valhall field. part 2: Imaging compressive-wave and shear-wave velocities. *Geophysical Journal International*, page ggt178. [p. 210]

Pyun, S., Shin, C., and Bednar, J. B. (2007). Comparison of waveform inversion, part 3: amplitude approach. *Geophysical Prospecting*, 55(4):477–485. [p. 133]

Qiu, L., Chemingui, N., Zou, Z., and Valenciano, A. (2016). Full-waveform inversion with steerable variation regularization. In *SEG Technical Program Expanded Abstracts 2016*, pages 1174–1178. Society of Exploration Geophysicists. [p. 301]

Qiu, L., Ramos-Martínez, J., Valenciano, A., Yang, Y., and Engquist, B. (2017). Full-waveform inversion with an exponentially encoded optimal-transport norm. In *SEG Technical Program Expanded Abstracts 2017*, pages 1286–1290. Society of Exploration Geophysicists. [p. 301]

Rao, S. S. (2007). *Vibration of Continuous Systems*. John Wiley & Sons, Inc. [p. 20]

Rayleigh, J. W. S. (1896). *The theory of sound*, volume 2. Macmillan. [p. 20]

Reed, W. H. and Hill, T. (1973). Triangular mesh methods for the neutron transport equation. Technical report, Los Alamos Scientific Lab., N. Mex.(USA). [p. 53]

Ribodetti, A. and Virieux, J. (1998). Asymptotic theory for imaging the attenuation factor Q. *Geophysics*, 63(5):1767–1778. [p. 44]

Riviere, B. (2008). *Discontinuous Galerkin methods for solving elliptic and parabolic equations: theory and implementation*. SIAM. [p. 53], [p. 66]

Rønholt, G., Lie, J., Korsmo, O., Danielsen, B., Brandsberg-Dah, S., Brown, S., Chemingui, N., Mavilio, A. V., Whitmore, D., and Martinez, R. (2015). Broadband velocity model building and imaging using reflections, refractions and multiples from dual-sensor streamer data. In *14th International Congress of the Brazilian Geophysical Society & EXPOGEF, Rio de Janeiro, Brazil, 3-6 August 2015*, pages 1006–1009. Brazilian Geophysical Society. [p. 305]

Rusmanugroho, H., Modrak, R., and Tromp, J. (2017). Anisotropic full-waveform inversion with tilt-angle recovery. *Geophysics*, 82(3):R135–R151. [p. 124]

Shi, J., Faucher, F., de Hoop, M. V., and Calandra, H. (2014). Multi-level elastic full waveform inversion in isotropic media via quantitative lipschitz stability estimates. In *Proceedings of the Project Review, Geo-Mathematical Imaging Group*, volume 1, Chicago, United States. [p. 159]

Shin, C. and Cha, Y. H. (2008). Waveform inversion in the laplace domain. *Geophysical Journal International*, 173(3):922–931. [p. 198], [p. 234]

Shin, C. and Cha, Y. H. (2009). Waveform inversion in the laplace fourier domain. *Geophysical Journal International*, 177(3):1067–1079. [p. 123], [p. 164]

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. [p. 123]

Shin, C., Koo, N.-H., Cha, Y. H., and Park, K.-P. (2010). Sequentially ordered single-frequency 2-d acoustic waveform inversion in the laplace–fourier domain. *Geophysical Journal International*, 181(2):935–950. [p. 164], [p. 198], [p. 234]

Shin, C. and Min, D.-J. (2006). Waveform inversion using a logarithmic wavefield. *Geophysics*, 71(3):R31–R42. [p. 123], [p. 133]

Shin, C., Pyun, S., and Bednar, J. B. (2007). Comparison of waveform inversion, part 1: conventional wavefield vs logarithmic wavefield. *Geophysical Prospecting*, 55(4):449–464. [p. 133]

Sirgue, L. and Pratt, R. G. (2004). Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248. [p. 124], [p. 134], [p. 194], [p. 234]

Slawinski, M. A. (2010). *Waves and rays in elastic continua*. World Scientific. [p. 21], [p. 23], [p. 24], [p. 27], [p. 28]

Stefanov, P. and Uhlmann, G. (1998). Stability estimates for the hyperbolic dirichlet to neumann map in anisotropic media. *Journal of functional analysis*, 154(2):330–358. [p. 77]

Stefanov, P. and Uhlmann, G. (2005). Stable determination of generic simple metrics from the hyperbolic dirichlet-to-neumann map. *International Mathematics Research Notices*, 2005(17):1047–1061. [p. 77]

Stefanov, P. and Uhlmann, G. (2009). Thermoacoustic tomography with variable sound speed. *Inverse Problems*, 25(7):075011. [p. 77]

Štekl, I. and Pratt, R. G. (1998). Accurate viscoelastic modeling by frequency-domain finite differences using rotated operators. *Geophysics*, 63(5):1779–1794. [p. 44], [p. 58]

Sylvester, J. and Uhlmann, G. (1987). A global uniqueness theorem for an inverse boundary value problem. *Ann. of Math. (2)*, 125(1):153–169. [p. 76], [p. 94], [p. 105]

Sylvester, J. and Uhlmann, G. (1990). The Dirichlet to Neumann map and applications. *Inverse problems in partial differential equations*, pages 101–139. [p. 76]

Symes, W. W. (2009). The seismic reflection inverse problem. *Inverse Problems*, 25(12):123008, 39. [p. 94]

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49:1259–1266. [p. 116], [p. 122], [p. 149]

Tarantola, A. (1986). A strategy for nonlinear elastic inversion of seismic reflection data. *Geophysics*, 51(10):1893–1903. [p. 146], [p. 186], [p. 210]

Tarantola, A. (1987a). *Inverse problem theory : methods for data fitting and model parameter estimation*. Elsevier, Amsterdam. Other Versions: 1988, 1994, 1998, 2002. [p. 123], [p. 133]

Tarantola, A. (1987b). Inversion of travel times and seismic waveforms. In *Seismic tomography*, pages 135–157. Springer. [p. 122], [p. 142]

Tarantola, A. (1988). Theoretical background for the inversion of seismic waveforms including elasticity and attenuation. *Pure and Applied Geophysics*, 128(1-2):365–399. [p. 142]

Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA. [p. 75], [p. 124], [p. 146], [p. 150], [p. 159], [p. 210]

Tarantola, A. and Valette, B. (1982). Generalized nonlinear inverse problem solved using the least squares criterion. *Reviews of Geophysics and Space Physics*, 20:219–232. [p. 75]

Tcheverda, V., Gadylshin, K., Chavent, G., Shin, C., and Shin, J. (2016). Reflection fwi in mbtt formulation: Macro velocity reconstruction. In *7th EAGE Saint Petersburg International Conference and Exhibition*. [p. 300]

Tenghamn, R., Vaage, S., and Borresen, C. (2007). A dual-sensor towed marine streamer: Its viable implementation and initial results. In *SEG Technical Program Expanded Abstracts 2007*, pages 989–993. Society of Exploration Geophysicists. [p. 130], [p. 305]

Thomée, V. (2001). From finite differences to finite elements: A short history of numerical analysis of partial differential equations. *Journal of Computational and Applied Mathematics*, 128(1):1–54. [p. 53]

Thomsen, L. (1986). Weak elastic anisotropy. *Geophysics*, 51:1954–1966. [p. 29], [p. 31]

Trampert, J., Fichtner, A., and Ritsema, J. (2012). Resolution tests revisited: the power of random numbers. *Geophysical Journal International*, 192(2):676–680. [p. 125]

Tromp, J., Tape, C., and Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophys. J. Int.*, 160:195–216. [p. 124], [p. 210]

Tsvankin, I. (1997). Anisotropic parameters and p-wave velocity for orthorhombic media. *Geophysics*, 62(4):1292–1309. [p. 28]

Tsvankin, I. (2012). *Seismic signatures and analysis of reflection data in anisotropic media*. Society of Exploration Geophysicists. [p. 28]

Tsvankin, I. and Grechka, V. (2011). *Seismology of azimuthally anisotropic media and seismic fracture characterization*. Society of Exploration Geophysicists. [p. 21], [p. 28]

Turkel, E. and Yefet, A. (1998). Absorbing pml boundary layers for wave-like equations. *Applied Numerical Mathematics*, 27(4):533–557. [p. 68]

Uhlmann, G. (2009). Electrical impedance tomography and calderón's problem. *Inverse problems*, 25(12):123011. [p. 76], [p. 78]

Uhlmann, G. and Wang, J. (2008). Complex geometrical optics solutions and reconstruction of discontinuities. *SIAM J. Appl. Math*, 68:1026–1044. [p. 76]

Ursin, B. and Toverud, T. (2002). Comparison of seismic dispersion and attenuation models. *Studia Geophysica et Geodaetica*, 46:293–320. [p. 43]

Virieux, J. (1984). Sh-wave propagation in heterogeneous media: velocity-stress finite-difference method. *Geophysics*, 49(11):1933–1942. [p. 58]

Virieux, J. (1986). P-sv wave propagation in heterogeneous media: Velocity-stress finite-difference method. *Geophysics*, 51(4):889–901. [p. 58]

Virieux, J. and Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26. [p. 125], [p. 163], [p. 179], [p. 210], [p. 225]

Wang, S., de Hoop, M. V., and Xia, J. (2010). Acoustic inverse scattering via helmholtz operator factorization and optimization. *Journal of Computational Physics*, 229(22):8445 – 8462. [p. 329]

Wang, S., de Hoop, M. V., and Xia, J. (2011). On 3d modeling of seismic wave propagation via a structured parallel multifrontal direct helmholtz solver. *Geophysical Prospecting*, 59(5):857–873. [p. 54], [p. 68], [p. 69], [p. 329]

Wang, S., de Hoop, M. V., Xia, J., and Li, X. S. (2012). Massively parallel structured multifrontal solver for time-harmonic elastic waves in 3-d anisotropic media. *Geophysical Journal International*, 191(1):346–366. [p. 54], [p. 69], [p. 329], [p. 331]

Wang, Z., Navon, I., Le Dimet, F., and Zou, X. (1992). The second order adjoint analysis: theory and applications. *Meteorology and atmospheric physics*, 50(1-3):3–20. [p. 123], [p. 149]

Whitmore, N., Valenciano, A., Sollner, W., and Lu, S. (2010). Imaging of primaries and multiples using a dual-sensor towed streamer. In *SEG Technical Program Expanded Abstracts 2010*, pages 3187–3192. Society of Exploration Geophysicists. [p. 305]

Will, S., Ken, M., and Bill, L. (2006). The visualization toolkit: an object-oriented approach to 3d graphics. kitware. [p. 332]

Wolfe, P. (1969). Convergence conditions for ascent methods. *SIAM review*, 11(2):226–235. [p. 162]

Wolfe, P. (1971). Convergence conditions for ascent methods. ii: Some corrections. *SIAM review*, 13(2):185–188. [p. 162]

Xia, J., Chandrasekaran, S., Gu, M., and Li, X. S. (2009). Superfast multifrontal method for large structured linear systems of equations. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1382–1411. [p. 70]

Xia, J., Chandrasekaran, S., Gu, M., and Li, X. S. (2010). Fast algorithms for hierarchically semiseparable matrices. *Numerical Linear Algebra with Applications*, 17(6):953–976. [p. 70]

Xin, Z., Xia, J., de Hoop, M. V., Cauley, S., and Balakrishnan, V. (2017). A distributed-memory randomized structured multifrontal method for sparse direct solutions. *SIAM Journal on Scientific Computing*, 39(4):C292–C318. [p. 70]

Xu, K. and McMechan, G. A. (2014). 2d frequency-domain elastic full-waveform inversion using time-domain modeling and a multistep-length gradient approach. *Geophysics*, 79(2):R41–R53. [p. 122]

Yuan, Y. O. and Simons, F. J. (2014). Multiscale adjoint waveform-difference tomography using wavelets. *Geophysics*, 79(3):WA79–WA95. [p. 90], [p. 124]

Yuan, Y. O., Simons, F. J., and Bozdağ, E. (2015). Multiscale adjoint waveform tomography for surface and body waves. *Geophysics*, 80(5):R281–R302. [p. 90], [p. 124], [p. 233], [p. 289]

Zhang, Y., Zhang, H., and Zhang, G. (2011). A stable tti reverse time migration and its implementation. *Geophysics*, 76(3):WA3–WA11. [p. 30]

List of Figures

List of Tables

Abstract

We investigate the recovery of subsurface Earth parameters. We consider the seismic imaging as a large scale iterative minimization problem, and deploy the Full Waveform Inversion (FWI) method, for which several aspects must be treated. The reconstruction is based on the wave equations because the characteristics of the measurements indicate the nature of the medium in which the waves propagate. First, the natural heterogeneity and anisotropy of the Earth require numerical methods that are adapted and efficient to solve the wave propagation problem. In this study, we have decided to work with the harmonic formulation, i.e., in the frequency domain. Therefore, we detail the mathematical equations involved and the numerical discretization used to solve the wave equations in large scale situations.

The inverse problem is then established in order to frame the seismic imaging. It is a nonlinear and ill-posed inverse problem by nature, due to the limited available data, and the complexity of the subsurface characterization. However, we obtain a conditional Lipschitz-type stability in the case of piecewise constant model representation. We derive the lower and upper bound for the underlying stability constant, which allows us to quantify the stability with frequency and scale. It is of great use for the underlying optimization algorithm involved to solve the seismic problem. We review the foundations of iterative optimization techniques and provide the different methods that we have used in this project. The Newton method, due to the numerical cost of inverting the Hessian, may not always be accessible. We propose some comparisons to identify the benefits of using the Hessian, in order to study what would be an appropriate procedure regarding the accuracy and time. We study the convergence of the iterative minimization method, depending on different aspects such as the geometry of the subsurface, the frequency, and the parametrization. In particular, we quantify the frequency progression, from the point of view of optimization, by showing how the size of the basin of attraction evolves with frequency.

Following the convergence and stability analysis of the problem, the iterative minimization algorithm is conducted via a multi-level scheme where frequency and scale progress simultaneously. We perform a collection of experiments, including acoustic and elastic media, in two and three dimensions. The perspectives of attenuation and anisotropic reconstructions are also introduced. Finally, we study the case of Cauchy data, motivated by the dual sensors devices that are developed in the geophysical industry. We derive a novel cost function, which arises from the stability analysis of the problem. It allows elegant perspectives where no prior information on the acquisition set is required.

Key words: Inverse Problem – Seismic Imaging – Time-Harmonic Wave Propagation – Stability and Convergence Analysis – Full Waveform Inversion.