Development of numerical methods for seismic imaging using the full waveform

Ludovic Métivier

UNIVERSITÉ DE GRENOBLE

MÉMOIRE D'HABILITATION

Pour l'obtention de l'

HABILITATION À DIRIGER DES RECHERCHES DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Mathématiques appliquées**

Arrêté ministériel : 23 Novembre 1988

Présenté par

Ludovic Métivier

préparé au sein du **Laboratoire Jean Kuntzmann**
et de **l'école doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique**

Development of numerical methods for seismic imaging using the full waveform

15 Septembre 2017 ,
devant le jury composé de :

William W. Symes
Professeur de mathématiques appliquées, chaire "Noah Harding", et Professeur en sciences de la Terre, Rice University, Houston, USA, Rapporteur

Yann Brenier
Directeur de recherches au CNRS, Centre de Mathématiques Laurent Schwartz, École Polytechnique, Palaiseau, France, Rapporteur

R.E Plessix
René-Edouard Plessix, Chercheur expert à Shell Global Solution International, Pays-Bas, Rapporteur

Jan Hesthaven
Professeur de mathématiques appliquées et des sciences de la simulation, École Polytechnique Fédérale de Lausanne, Lausanne, Suisse, Examinateur

Jeannot Trampert
Professeur de géophysique à l'Université d'Utrecht, Utrecht, Pays-Bas, Examinateur

Michel Campillo
Professeur à l'Institut des Sciences de la Terre, Université Grenoble Alpes, Grenoble, France, Examinateur

Emmanuel Maitre
Professeur de mathématiques appliquées, ENSIMAG, INP Grenoble, Grenoble, France, Président

Table des matières

Présentation générale

Ce mémoire résume l'activité de recherche que j'ai entreprise depuis mon recrutement comme chargé de recherche au CNRS (en octobre 2012) et mon affectation au laboratoire Jean Kuntzmann de l'université Grenoble Alpes. Il contient également un résumé de mon projet de recherche pour les prochaines années.

Mon activité de recherche est principalement concentrée sur le développement et l'analyse de méthodes numériques pour l'imagerie sismique basée sur l'inversion des formes d'ondes. Le développement de cette méthode d'imagerie sismique date des années 80, cependant son application aux données d'exploration dans l'industrie pétrolière, et aux données sismologiques dans l'académie, est plus récente : elle a commencé dans les années 2000 et s'est généralisée ces dix dernières années. Elle a atteint aujourd'hui le statut de méthode état-de-l'art pour ces applications.

Cependant, cette stratégie fait encore face à de nombreux verrous méthodologiques pour la généralisation de son approche à des modélisations plus réalistes et plus complexes (élasticité, viscosité, anisotropie), pour mieux extraire l'information des données, et pour automatiser son application sans nécessiter d'importants pré-traitements dépendants de l'expertise de spécialistes des applications visées. Ces verrous empêchent aujourd'hui la systématisation de l'approche, qui permettrait son utilisation à des échelles plus variées (des expériences de laboratoire sur des échantillons de roches jusqu'à la sismologie globale). Proposer des solutions pour dépasser ces verrous constitue le cœur de mon activité de recherche.

L'inversion des formes d'ondes se formule comme un problème inverse, basé sur la minimisation de l'écart entre des données simulées et des données observées. Les contributions scientifiques que j'ai pu apporter concernent à la fois les méthodes numériques utilisées pour la modélisation de la propagation des ondes et la solution du problème inverse lui-même.

Dans ce mémoire, j'ai choisi de présenter en détail trois de ces contributions. La première est une stratégie de couches absorbantes qui reste stable dans l'approximation élasto-dynamique anisotrope de la propagation des ondes, à la différence de la méthode de référence pour ce type de problème (couches parfaitement adaptées PML).

La seconde de ces contributions consiste en l'analyse et le développement d'une méthode d'optimisation dite de Newton tronquée, adaptée aux grandes tailles de problème associées aux applications de l'inversion de formes d'ondes. Cette méthode permet une meilleure approximation de la courbure locale de la fonction coût, amenant des reconstructions plus stables des propriétés mécaniques du sous-sol, en particulier quand plusieurs classes de paramètres sont reconstruites simultanément.

La troisième contribution présentée en détail dans ce manuscrit concerne la définition de la fonction coût elle-même. En particulier, je me suis intéressé à l'utilisation de distances de type transport optimal pour redéfinir la notion d'écart entre données observées et simulées. J'ai pu montrer que la convexité de ces distances par rapport à des décalages (temporel ou spatio-temporel) est une caractéristique fondamentale qui pourrait apporter une solution à un problème posé de longue date par l'inversion de formes d'ondes. La fonctionnelle aux moindre carrés utilisée couramment est en effet non-convexe et la possibilité de converger vers un minium local est toujours présente dans les applications, ce qui nécessite, pour les mener à bien, un savoir faire et un traitement des données parfois très lourd, sans garantie de succès.

La structure du mémoire est la suivante. Dans le premier chapitre, un Curriculum Vitae détaillé est donné. Il regroupe les informations sur mon parcours académique et professionnel. Le second chapitre présente mon activité de recherche. Après avoir introduit les concepts clefs de l'inversion de formes

d'ondes, les trois contributions mentionnées ci-dessus sont détaillées. Ce chapitre est complété par une présentation plus succincte d'autres travaux que j'ai menés, toujours dans le contexte de l'imagerie sismique. Le troisième chapitre présente mon projet de recherche pour les prochaines années.

General overview

This manuscript summarizes my research activity since I was engaged by the CNRS in October 2012 as a permanent researcher in the Jean Kuntzmann Laboratory (LJK), the applied mathematics laboratory of University Grenoble Alpes. It also contains an overview of my research project for the coming years.

My research activity mainly focuses on the development and the analysis of numerical methods for seismic imaging using the full waveform. This seismic imaging technique, despite its introduction in the early 80s and the increase of its use both in the industry and at the academy level, still faces methodological issues: accounting for more realistic wave propagation physics (elasticity, viscosity, anisotropy), extracting more information from the data, designing more robust workflow, less dependent on human expertise. These issues still prevent today the application of this strategy at more various scales (from laboratory experiments on rock samples to global seismology) with the objective of reaching unprecedented high resolution estimates of the mechanical properties of a material with noninvasive techniques. Providing methods trying to overcome or mitigate these issues is the main motivation of my research activity.

Full waveform inversion is an inverse problem, formulated as the minimization of the distance between simulated data and observed data. The contributions I have been able to bring so far are related both to the design of numerical strategies for the solution of the forward problem, *i.e.* the computation of the solution to the wave equation in complex media, and to the solution of the inverse problem itself.

In this manuscript I have chosen to focus on three of these contributions. The first is a strategy of absorbing layers which is stable when considering the solution of the elastodynamics equation in the presence of anisotropy, contrary to the state-of-the-art method for this category of computational problems (perfectly matched layers method).

The second is the design and the analysis of a truncated Newton strategy adapted to the large-scale aspect of the problem. Using this method it is possible to better estimate the local curvature of the misfit function, which yields more stable reconstruction of the subsurface mechanical properties, in particular for multi-parameter problems.

The third is related to the definition of the misfit function itself. In particular, I have been interested in the use of optimal transport distances to define the misfit between simulated and observed data. I have shown that the convexity of theses distances with respect to shifts between compared measures is a fundamental feature which might be a solution to a long standing problem in the seismic imaging community regarding the non-convexity of the conventional L^2 misfit function, and the convergence towards spurious local minima. Avoiding these local minima requires a human expertise and a preprocessing of the data which is often cumbersome and difficult to set up, with no guarantee of success.

The structure of this manuscript is the following. In the first Chapter, a detailed Curriculum Vitae presents my professional activity and education background. The second Chapter is dedicated to a summary of my research activity, with a focus on the aforementioned items, after an overview of the full waveform inversion method and its history is given. This Chapter ends with a shorter description of other contributions I have proposed, still in the field of seismic imaging using the full waveform. The third Chapter is dedicated to a presentation of my research project for the coming years.

Chapter 1

Curriculum Vitae

Contents

1 General information

Ludovic Métivier
CNRS researcher at LJK, University Grenoble Alpes

Professional address:

Ludovic Métivier
Batiment IMAG
Université Grenoble Alpes
700 Avenue Centrale, 38401 Saint Martin d'Hères

Phone: + 33 4 57 42 17 67
E-mail: ludovic.metivier@univ-grenoble-alpes.fr

Websites:
`https://www-ljk.imag.fr/membres/Ludovic.Metivier/`
`https://seiscope2.osug.fr/`

2 Education and professional experience

Education

- **2006-2009.** *PhD in applied mathematics, Univ. Paris XIII, France*

  PhD in applied mathematics under the supervision of L. Halpern (Univ. Paris XIII), P. Lailly (IFPEN, Rueil-Malmaison) and F. Delprat-Jannaud (IFPEN, Rueil-Malmaison).

- **2005-2006.** *Master 2 degree, Univ. Paul Sabatier Toulouse, France*

  Master 2 degree in applied mathematics.

- **2003-2006.** *Engineer degree, ENSEEIHT, Toulouse, France*

  Engineer degree in informatics and applied mathematics.

Professional experience

- **2012-present.** *CNRS Researcher, LJK, Univ. Grenoble Alpes, France*

  Mathematics and interactions. Research project based on methodological developments for seismic imaging. Modeling of seismic wave: accounting for anisotropy, elasticity, development of robust absorbing boundary conditions. Seismic imaging using the full waveform. Approximation strategies for the local curvature of the misfit function. Application to multi-parameter inversion, uncertainty estimation. Development of a large scale numerical optimization strategy for smooth nonlinear function problems. Development of misfit functions based on optimal transport distances: application to full waveform inversion.

- **2011-2012.** *Post-doctoral position, Institut des Sciences de la Terre (ISTerre), Univ. Grenoble Alpes, France*

  Study, analysis, and implementation of new numerical strategies for seismic imaging using the full waveform. In particular, analysis and implementation of the truncated Newton strategy. Application to multi-parameter inversion.

- **2010-2011.** *Post-doctoral position, CEA Saclay, France*

  Modeling and numerical solution of a transport/chemical problem for liquid/liquid extraction. Implementation of a prototype within a software platform dedicated to the modeling of the back end of the nuclear fuel cycle.

- **2006-2009.** *PhD in applied mathematics, Univ. Paris XIII, France*

  PhD in mathematics applied to geophysics. Nonlinear inversion strategy of walkaway seismic data for improving the local resolution. PhD defended under the supervision of Laurence Halpern (Paris XIII, LAGA) and Patrick Lailly (IFPEN, Rueil-Malmaison), Florence Delprat-Jannaud (IFPEN, Rueil-Malmaison).

3 Projects and fundings

- **2016-*.** *SEISCOPE project*

  co-PI with R. Brossier (ISTerre-UGA) of the research project SEISCOPE. The project is built as a consortium of industrial partners which fund PhD projects, post-doc and engineer positions at the University. The project is organized following 3 years periods renewable once (6 years in total). A 3 years period has started in 2016 with 9 industrial partners: CGG, Chevron, Exxon-Mobil, JGI, Sinopec, Shell, Statoil, TOTAL, Woodside. Website: `https://seiscope2.osug.fr/`

- **2016-2021.** *ANR project "Défi des Autres savoirs" HIWAI*

  Local coordinator of the ANR project HIWAI "Homogenized seismic full Waveform Inversion and downscaling". The PI of this project is Y. Capdeville (Univ. Nantes, France, LPG)

- **2015-2016** *INSU-INSMI project*

  PI of the INSU-INSMI project "Implicit parallel in time integration strategy through Spectral Deferred Correction method for magneto-hydro-dynamic equations" (8kE)

- **2015-2016.** *Projet local AGIR*

  PI of the local Univ. Grenoble Alpes project "Asymptotic approach for preconditioning a multi-parameter inverse problem in seismic imaging" (10 kE)

- **2013-2014.** *Projet local AGIR*

  PI of the local Univ. Grenoble Alpes project "Wasserstein Distance for Full Waveform Inversion" (10 kE)

4 Research supervision

Unofficial PhD supervision

1. Clara Castellanos

- **Funding:** SEISCOPE
- **Period:** November 2010 - November 2013
- **Advisors:** Stéphane Operto, Géoazur (75 %), Stéphane Gaffet, CNRS, Géoazur (25 %)
- **Summary:** Source encoding techniques and non-smooth regularization strategies for full waveform inversion.

2. Francois Lavoué

- **Funding:** École normale supérieure de Lyon.
- **Period :** December 2010 - December 2013
- **Advisors** Stéphane Garambois, ISterre (50 %), Jean Virieux, ISTerre (50 %)
- **Summary:** Full waveform inversion applied to ground penetrating radar surface data.

Defended PhD

1. Yang Li

- **Funding:** exchange PhD student for 2 years from Harbin Institute (China), funded by the Chinese national program
- **Period:** September 2012- September 2014
- **Advisors** Ludovic Métivier, LJK (75 %), Jean Virieux, ISTerre (25 %)
- **Summary:** Implementation of the iterative solver CARP-CG. Application to 2D and 3D visco-elastic frequency-domain wave propagation modeling.

2. Okba Hamitou

- **Funding:** SEISCOPE
- **Period:** November 2013 - November 2016
- **Advisors:** Ludovic Métivier, LJK (50 %), Stéphane Labbé, LJK (25 %), Jean Virieux, ISTerre (25 %)
- **Summary:** Development of preconditioning techniques for the iterative solver CARP-CG. Application to 2D and 3D visco-elastic frequency-domain wave propagation modeling.

Ongoing PhD

1. Arnaud Pladys

- **Funding:** SEISCOPE
- **Start:** January 2017
- **Advisors:** Romain Brossier, ISTerre (50 %), Ludovic Métivier, LJK (50 %)
- **Summary:** Comparison and characterization of misfit functions for full waveform inversion.

2. Aude Allain

- **Funding:** National grant
- **Start:** December 2016
- **Advisors:** Ludovic Métivier, LJK (50 %), Edouard Oudet, LJK (50 %)
- **Summary:** Optimal transport distances for the comparison of signed measures. Application to seismic imaging.

3. Julien Thurin

- **Funding:** National grant
- **Start:** December 2016
- **Advisors:** Romain Brossier, ISTere (50 %), Ludovic Métivier, LJK (50 %)
- **Summary:** Uncertainty quantification through extended Kalman filter in seismic imaging.

4. Phong-Thu Trinh

- **Funding:** SEISCOPE
- **Start:** November 2015
- **Advisors:** Romain Brossier, ISTerre (33 %), Ludovic Métivier, LJK (33 %), Jean Virieux, ISTerre (33 %)
- **Summary:** 3D visco-elastic time-domain full waveform inversion using a spectral element modeling engine. Application to land data with complex topography

5. Philippe Le Bouteiller

- **Funding:** SEISCOPE
- **Start:** December 2015
- **Advisors:** Jean Virieux, ISTerre (50 %), Ludovic Métivier, LJK (50 %)
- **Summary:** Travel-time computation in complex media in the asymptotic approximation through Hamilton-Jacobi equations and a discontinuous Galerkin solver.

6. Hugo Sanchez

- **Funding:** CNRS handicap grant
- **Start:** November 2015
- **Advisors:** Jean Virieux, ISTerre (33 %), Ludovic Métivier, LJK (33 %), Josué Tago, UNAM (Mexico) (33 %)
- **Summary:** Kinematic inversion of seismic source through full waveform inversion. Application to real earthquake data.

7. Hugo Pinard

- **Funding:** SEISCOPE
- **Start:** November 2013
- **Advisors:** Stéphane Garambois, ISTerre (33 %), Ludovic Métivier, LJK (33 %), Michel Dietrich, ISTerre (33 %)
- **Summary:** Multi-parameter full waveform inversion for near surface imaging using Ground Penetrating Radar data. Application to cross-hole data from the LSBB site (Rustrell).

5 Scientific animation and teaching

- **2016.** *Master 2 module "Frontiers in Geophysics" in Earth Sciences Master (Univ. Grenoble Alpes)*

  Three days module on full waveform inversion for the master 2 degree in Earth Sciences, together with R. Brossier (ISTerre, Univ. Grenoble Alpes).

- **2016.** *EAGE workshop on seismic wave modeling (Vienna)*

  Organization of a one day workshop on seismic wave modeling together with R. Brossier (ISTerre, Univ. Grenoble Alpes) and M. Huiskes (Shell Global Solutions International) within the international applied geophysics conference EAGE in Vienna.

- **2016.** *Doctoral school training on full waveform inversion (Univ. Grenoble Alpes)*

  Three days training on full waveform inversion for PhD students in Earth Sciences and Applied Mathematics, together with R. Brossier (ISTerre, Univ. Grenoble Alpes).

- **2015.** *Mini-symposium at SIAM Conference on Geosciences (Stanford)*

  Organization of a mini-symposium on elastic imaging together with S. Chaillat (ENSTA, POEMS, CNRS) at the SIAM Conference on Geosciences in Stanford.

- **2015.** *Full Waveform Inversion training at CGG and TOTAL*

  Three days training at TOTAL (Pau) and CGG (Massy) on full waveform inversion, together with R. Brossier (ISTerre, Univ. Grenoble Alpes) and J. Virieux (ISTerre, Univ. Grenoble Alpes).

- **2014-*.** *SEISCOPE annual meeting (Grenoble)*

  Organization of the SEISCOPE annual meeting together with R. Brossier (ISTerre, Univ. Grenoble Alpes) and J. Virieux (ISTerre, Univ. Grenoble Alpes). Two days of scientific presentations on the SEISCOPE consortium research activities and one day training on our computation code. The meeting is organized for the sponsors of SEISCOPE but is opened to colleagues from other laboratories of the university.

- **2012-*.** *EDP team seminar at LJK (Grenoble)*

  Organization of the EDP team seminar EDP at LJK together with C. Jourdana (LJK, Univ. Grenoble Alpes).

6 Scientific production

Summary

- h-factor : 11 / i10 factor : 17 (Google Scholar data `https://scholar.google.fr/citations?user=S2-_b7oAAAAJ&hl=fr`)
- 32 articles in peer-reviewed international journals

- 1 patent
- 34 extended peer-reviewed international conference abstracts

Complete list of publications in peer-reviewed international journals

- 2017

1. Trinh, P. T., Brossier, R., Métivier, L., Virieux, J., & Wellington, P., 2017b. Bessel smoothing filter for spectral element mesh, *Geophysical Journal International*, **209**(3), 1489–1512
2. Métivier, L., Brossier, R., Operto, S., & J., V., 2017. Full waveform inversion and the truncated Newton method, *SIAM Review*, **59**(1), 153–195
3. Virieux, J., Asnaashari, A., Brossier, R., Métivier, L., Ribodetti, A., & Zhou, W., 2017. An introduction to Full Waveform Inversion, in *Encyclopedia of Exploration Geophysics*, pp. R1–1–R1–40, eds Grechka, V. & Wapenaar, K., Society of Exploration Geophysics

- 2016

4. Amestoy, P., Brossier, R., Buttari, A., L'Excellent, J.-Y., Mary, T., Métivier, L., Miniussi, A., & Operto, S., 2016a. Fast 3d frequency-domain full waveform inversion with a parallel Block Low-Rank multifrontal direct solver: application to OBC data from the North Sea, *Geophysics*, **81**(6), R363 – R383
5. Dupuy, B., Asnaashari, A., Brossier, R., Garambois, S., Métivier, L., Ribodetti, A., & Virieux, J., 2016. A downscaling strategy from FWI to microscale reservoir properties from high-resolution images, *The Leading Edge*, **35**, 1146–150
6. Li, Y., Han, B., Métivier, L., & Brossier, R., 2016. Optimal fourth-order staggered-grid finite-difference scheme for 3D frequency-domain viscoelastic wave modeling, *Journal of Computational Physics*, **321**, 1055–1078
7. Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J., 2016c. Increasing the robustness and applicability of full waveform inversion: an optimal transport distance strategy, *The Leading Edge*, **35**(12), 1060–1067
8. Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J., 2016b. An optimal transport approach for seismic tomography: Application to 3D full waveform inversion, *Inverse Problems*, **32**(11), 115008
9. Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J., 2016d. Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion, *Geophysical Journal International*, **205**, 345–377
10. Virieux, J., Brossier, R., Métivier, L., Operto, S., & Ribodetti, A., 2016. Direct and indirect inversions, *Journal of Seismology*
11. Yang, P., Brossier, R., Métivier, L., & Virieux, J., 2016a. Wavefield reconstruction in attenuating media: A checkpointing-assisted reverse-forward simulation method, *Geophysics*, **81**(6), R349–R362
12. Yang, P., Brossier, R., Métivier, L., & Virieux, J., 2016c. A review on the systematic formulation of 3D multiparameter full waveform inversion in viscoelastic medium, *Geophysical Journal International*, **207**(1), 129–149

- 2015

13. Castellanos, C., Métivier, L., Operto, S., Brossier, R., & Virieux, J., 2015. Fast full waveform inversion with source encoding and second-order optimization methods, *Geophysical Journal International*, **200(2)**, 720–744

14. Li, Y., Métivier, L., Brossier, R., Han, B., & Virieux, J., 2015b. 2D and 3D frequency-domain elastic wave modeling in complex media with a parallel iterative solver, *Geophysics*, **80(3)**, T101–T118
15. Métivier, L., Brossier, R., Labbé, S., Operto, S., & Virieux, J., 2014c. Smart: dissipative absorbing layer technique for general elastodynamics equations. application as s-waves filter in acoustic ti media., *Seismic Technology*, **11**(4), 14
16. Métivier, L. & Brossier, R., 2016b. The SEISCOPE optimization toolbox: A large-scale nonlinear optimization library based on reverse communication, *Geophysics*, **81**(2), F11–F25
17. Métivier, L., Brossier, R., & Virieux, J., 2015c. Combining asymptotic linearized inversion and full waveform inversion, *Geophysical Journal International*, **201**(3), 1682–1703
18. Operto, S., Miniussi, A., Brossier, R., Combe, L., Métivier, L., Monteiller, V., Ribodetti, A., & Virieux, J., 2015b. Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic approximation, *Geophysical Journal International*, **202**(2), 1362–1391

- 2014

19. Métivier, L., Brossier, R., Labbé, S., Operto, S., & Virieux, J., 2014b. A robust absorbing layer for anisotropic seismic wave modeling, *Journal of Computational Physics*, **279**, 218–240
20. Métivier, L., Bretaudeau, F., Brossier, R., Operto, S., & Virieux, J., 2014a. Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures, *Geophysical Prospecting*, **62**, 1353–1375
21. Operto, S., Brossier, R., Combe, L., Métivier, L., Ribodetti, A., & Virieux, J., 2014b. Computationally-efficient three-dimensional visco-acoustic finite-difference frequency-domain seismic modeling in vertical transversely isotropic media with sparse direct solver, *Geophysics*, **79(5)**, T257–T275
22. Tago, J., Métivier, L., & Virieux, J., 2014. SMART layers: a simple and robust alternative to PML approaches for elastodynamics, *Geophysical Journal International*, **199**(2), 700–706
23. Lavoué, F., Brossier, R., Métivier, L., Garambois, S., & Virieux, J., 2014. Two-dimensional permittivity and conductivity imaging by full waveform inversion of multioffset GPR data: a frequency-domain quasi-Newton approach, *Geophysical Journal International*, **197**(1), 248–268

- 2013

24. Métivier, L., Brossier, R., Virieux, J., & Operto, S., 2013. Full Waveform Inversion and the truncated Newton method, *SIAM Journal On Scientific Computing*, **35(2)**, B401–B437
25. Operto, S., Brossier, R., Gholami, Y., Métivier, L., Prieux, V., Ribodetti, A., & Virieux, J., 2013. A guided tour of multiparameter full waveform inversion for multicomponent data: from theory to practice, *The Leading Edge*, **Special section Full Waveform Inversion**(September), 1040–1054

- 2012

26. Asnaashari, A., Brossier, R., Castellanos, C., Dupuy, B., Etienne, V., Gholami, Y., Hu, G., Métivier, L., Operto, S., Pageot, D., Prieux, V., Ribodetti, A., Roques, A., & Virieux, J., 2012. Hierarchical approach of seismic full waveform inversion, *Numerical Analysis and Applications*, **5**(2), 99–108
27. Métivier, L. & Montarnal, P., 2012. Strategies for solving index one dae with non-negative constraints: Application to liquid–liquid extraction, *Journal of Computational Physics*, **231**(7), 2945–2962

28. Métivier, L. & Roussel, H., 2012. Accounting robustly for instantaneous chemical equilibriums in reactive transport: a numerical method and its application to liquid–liquid extraction modeling, *Computers & Chemical Engineering*, **45**, 50–61

- 2011

29. Langouët, H., Métivier, L., Sinoquet, D., & Tran, Q.-H., 2011. Engine calibration: multi-objective constrained optimization of engine maps, *Optimization and Engineering*, **12**(3), 407–424
30. Métivier, L., Lailly, P., Delprat-Jannaud, F., & Halpern, L., 2011. A 2D nonlinear inversion of well-seismic data, *Inverse Problems*, **27**(5), 055005
31. Métivier, L., 2011. Interlocked optimization and fast gradient algorithm for a seismic inverse problem, *Journal of Computational Physics*, **230**(19), 7502–7518

- 2009

32. Métivier, L., 2009. Utilisation des équations Euler-PML en milieu hétérogène borné pour la résolution d'un problème inverse en géophysique, *ESAIM: Proc.*, **27**, 156–170

Patents

- 2010

1. Delprat-jannaud, F., Lailly, P., & Metivier, L., 2010. Method for imaging a target area of the subsoil using walkaway data, EP Patent 2,253,970

Peer-reviewed extended abstracts in international conferences

- 2017

1. Bouteiller, P. L., Jemaa, M. B., Chauris, H., Métivier, L., Tavakoli, B., Noble, M., & Virieux, J., 2017. Discontinuous galerkin method for tti eikonal equation, in *Proceedings of the* 79^{th} *EAGE Conference & Exhibition*, EAGE
2. Thurin, J., Brossier, R., & Métivier, L., 2017a. Ensemble-based uncertainty estimation in full waveform inversion, in *79th EAGE Conference and Exhibition 2017, Paris*, p. Tu P1 07
3. Thurin, J., Brossier, R., & Métivier, L., 2017b. An ensemble-transform kalman filter - full waveform inversion scheme for uncertainty estimation, in *87th SEG Conference and Exhibition 2017, Houston*
4. Trinh, P. T., Brossier, R., Métivier, L., Virieux, J., & Wellington, P., 2017c. Structure-smoothing Bessel filter for finite element mesh: Application on 3D elastic FWI, in *79th EAGE Conference and Exhibition 2017, Paris*
5. Trinh, P. T., Brossier, R., Métivier, L., Tavard, L., & Virieux, J., 2017a. Efficient 3d elastic FWI using a spectral-element method, in *87th SEG Conference and Exhibition 2017, Houston*
6. Yang, P., Brossier, R., Métivier, L., Virieux, J., & Zhou, W., 2017a. A second-order adjoint truncated newton approach to time-domain multiparameter full waveform inversion in viscoacoustic medium, in *Proceedings of the* 79^{th} *EAGE Conference & Exhibition*, EAGE

- 2016

7. Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J., 2016a. Overcoming cycle skipping in FWI: An optimal transport approach, in *Expanded Abstracts,* 78^{th} *Annual EAGE Meeting (Vienna)*
8. Métivier, L., Brossier, R., Oudet, E., Mérigot, Q., & Virieux, J., 2016e. An optimal transport distance for full-waveform inversion: Application to the 2014 chevron benchmark data set, in *SEG Technical Program Expanded Abstracts 2016*, pp. 1278–1283

9. Pinard, H., Garambois, S., Métivier, L., Dietrich, M., Sénéchal, G., & Rousset, D., 2016. Full-waveform inversion of gpr data acquired between boreholes in rustrel carbonates, in *Inter-Disciplinary Underground Science & Technology (i-DUST 2016), Avignon (France)*, vol. 12, E3S Web of Conferences

10. Yang, P., Brossier, R., Métivier, L., & Virieux, J., 2016b. Checkpointing-assisted reverse forward simulation: an optimal recomputation method for fwi and rtm, in *SEG Technical Program Expanded Abstracts 2016*, pp. 1089–1093

- 2015

11. Amestoy, P., Brossier, R., Buttari, A., L'Excellent, J.-Y., Mary, T., Métivier, L., Miniussi, A., Operto, S., Ribodetti, A., Virieux, J., & Weisbecker, C., 2015b. Efficient 3D frequency-domain full-waveform inversion of ocean-bottom cable data with sparse block low-rank direct solver: a real data case study from the North Sea, in *Expanded Abstracts, 85th Annual SEG Meeting (New Orleans)*, pp. 1303–1308, SEG

12. Amestoy, P., , Brossier, R., Buttari, A., L'Excellent, J.-Y., Mary, T., Métivier, L., Miniussi, A., Operto, S., Virieux, J., & Weisbecker, C., 2015a. 3D frequency-domain seismic modeling with a parallel BLR multifrontal direct solver, in *Expanded Abstracts, 85th Annual SEG Meeting (New Orleans)*, pp. 3606–3611, SEG

13. Beller, S., Monteiller, V., Operto, S., Nolet, G., Combe, L., Métivier, L., Virieux, J., Nissen-Meyer, T., Paul, A., & Zhao, L., 2015. 3d elastic full waveform inversion of teleseismic data for high-resolution ospheric imaging, in *77th EAGE Conference and Exhibition 2015*

14. Hamitou, O., Métivier, L., Labbé, S., Brossier, R., & Virieux, J., 2015. Preconditioning and multiple-right hand sides strategies for the solution of the frequency-domain wave propagation problem using the CGMN method, in *SEG Technical Program Expanded Abstracts 2015*, pp. 3612–3616

15. Lavoué, F., Brossier, R., Métivier, L., Operto, S., Garambois, S., & Virieux, J., 2015. Frequency-domain modelling and inversion of electromagnetic data for 2d permittivity and conductivity imaging: An application to the institut fresnel experimental data, *Near Surface Geophysics*, **13**(3), 227–241

16. Métivier, L., Brossier, R., Operto, S., & Virieux, J., 2015a. Acoustic multi-parameter FWI for the reconstruction of P-wave velocity, density and attenuation: preconditioned truncated newton approach, in *Expanded Abstracts, 85th Annual Meeting*, pp. 1198–1203, SEG

17. Métivier, L., Brossier, R., Operto, S., & Virieux, J., 2015b. Acoustic multi-parameter FWI for the reconstruction of P-wave velocity, density and attenuation: preconditioned truncated Newton approach, in *Expanded Abstracts, 85th Annual SEG Meeting (New Orleans)*

18. Operto, S., Miniussi, A., Brossier, R., Combe, L., Haller, N., Kjos, E., Métivier, L., Milne, R., Ribodetti, A., Song, Z., Virieux, J., & Zheng, Y., 2015a. Efficient 3d frequency-domain full-waveform inversion of ocean-bottom cable data: application to valhall in the visco-acoustic vti approximation, in *Expanded Abstracts, 77th Annual EAGE Meeting (Madrid)*

19. Pinard, H., Dietrich, M., Garambois, S., Lavoué, F., Métivier, L., & Virieux, J., 2015. Simultaneous gpr reconstruction of electrical conductivity and permittivity, in *77th EAGE Conference and Exhibition-Workshops*

- 2014

20. Bretaudeau, F., Brossier, R., Métivier, L., & Virieux, J., 2014. First-arrival delayed tomography using 1st and 2nd order adjoint-state method, in *Expanded Abstracts*, pp. 4757–4762, Society of Exploration Geophysics

21. Brossier, R., Pajot, B., Combe, L., Operto, S., Métivier, L., & Virieux, J., 2014. Time and frequency-domain FWI implementations based on time solver: analysis of computational complexities, in *Expanded Abstracts, 76th Annual EAGE Meeting (Amsterdam)*

22. Li, Y., Métivier, L., Brossier, R., Han, B., & Virieux, J., 2014a. A robust parallel iterative solver for frequency-domain elastic wave modeling, in *Expanded Abstracts, 76th Annual EAGE Meeting (Amsterdam)*

23. Li, Y., Métivier, L., Brossier, R., Han, B., & Virieux, J., 2014b. CARP-CG: A robust parallel iterative solver for frequency-domain elastic wave modeling, application to the Marmousi2 model, in *Expanded Abstracts, 84th Annual SEG Meeting (Denver)*

24. Masoni, I., Zhou, W., Brossier, R., Métivier, L., Operto, S., & Virieux, J., 2014b. Near-surface full waveform inversion using surface waves and reflected waves, in *Expanded Abstracts, 76th Annual EAGE Meeting (Amsterdam)*

25. Métivier, L., Brossier, R., Operto, S., & Virieux, J., 2014e. Multi-parameter FWI - an illustration of the Hessian operator role for mitigating trade-offs between parameter classes, in *Expanded Abstracts, 6th EAGE St-Petersbourg International Conference & Exhibition*

26. Métivier, L., Brossier, R., Operto, S., & Virieux, J., 2014d. A robust absorbing layer method for seismic wave simulation in anisotropic media, in *Expanded Abstracts, 76th Annual EAGE Meeting (Amsterdam)*

27. Operto, S., Brossier, R., Combe, L., Métivier, L., Ribodetti, A., & Virieux, J., 2014a. A frequency-domain seismic modeling engine for 3d visco-acoustic vti full waveform inversion of fixed- spread data, in *Expanded Abstracts, 76th Annual EAGE Meeting (Amsterdam)*

28. Pajot, B., Li, Y., Berthoumieux, V., Weisbecker, C., Brossier, R., Métivier, L., Thierry, P., Operto, S., & Virieux, J., 2014. A review of recent forward problem developments used for frequency-domain fwi, in *Expanded Abstracts, 76th Annual EAGE Meeting (Amsterdam)*

- 2013

29. Castellanos-Lopez, C., Métivier, L., Operto, S., & Brossier, R., 2013. Fast full waveform inversion with source encoding and second order optimization methods, in *75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013*, p. We 11 10

30. Lavoué, F., Brossier, R., Métivier, L., Garambois, S., & Virieux, J., 2013. 2D full waveform inversion of GPR surface data: permittivity and conductivity imaging, in *7th International Workshop on Advanced Ground-Penetrating Radar (IWAGPR 2013), Nantes (France)*

- 2012

31. Métivier, L., Brossier, R., Virieux, J., & Operto, S., 2012b. Toward Gauss-Newton and exact Newton optimization for full waveform inversion, in *EAGE, 74th Conference and Exhibition*

32. Métivier, L., Brossier, R., Virieux, J., & Operto, S., 2012a. Optimization schemes in FWI: the truncated Newton method, in *2012 SEG Abstracts*

- 2009

33. Métivier, L., Delprat-Jannaud, F., Halpern, L., & Lailly, P., 2009. 2d nonlinear inversion of walkaway data, in *SEG Technical Program Expanded Abstracts 2009*, pp. 2342–2346, Society of Exploration Geophysicists

- 2008

34. Langouët, H., Métivier, L., Sinoquet, D., & Tran, Q.-H., 2008. Optimization for engine calibration, in *ENGOPT International conference on engineering optimization, Rio de Janeiro, Brazil*, pp. 1–5

7 Honors and awards

- **2017.** *SIAM award for "Full-waveform inversion and the truncated Newton method"*

  Publication of the article "Full-waveform inversion and the truncated Newton method" in the section SIAM SIGEST of SIAM REVIEW 2017

- **2015.** *Citation as excellent reviewer for Geophysical Journal International*

  Award for the work as a peer-reviewer for Geophysical Journal International

- **2015.** *Honorable mention in the category best paper in Geophysics*

  Award for the article: "Computationally efficient three-dimensional acoustic finite-difference frequency-domain seismic modeling in vertical transversely isotropic media with sparse direct solver", S. Operto, R. Brossier, L. Combe, L. Métivier, A. Ribodetti, J. Virieux, Geophysics, 2014, 79(5),T257–T275

- **2013.** *Best student paper award IWAGPR 2013*

  Award for the oral presentation at the conference IWAGPR 2013 "2D full waveform inversion of GPR surface data: permittivity and conductivity imaging", F. Lavoué, R. Brossier, L. Métivier, S. Garambois, J. Virieux

- **2011.** *Yves Chauvin award*

  Yves Chauvin award from IFPEN for the PhD thesis: "A nonlinear inverse method for high resolution seismic imaging"

- **2011.** *Honorific degree from Paris XIII University*

  Honorific degree for the PhD thesis: "A nonlinear inverse method for high resolution seismic imaging"

- **2010.** *Competence certificate in high performance computing from C3I*

  Competence certificate in high performance computing delivered from GENCI (Grand Equipement National pour le Calcul Intensif) for the PhD: "A nonlinear inverse method for high resolution seismic imaging"

- **2008.** *Poster award CANUM 2008*

  Poster award: "2D Inversion of well-seismic data", L. Métivier, F. Delprat-Jannaud, P. Lailly, L. Halpern at CANUM (French National congress on numerical analysis)

8 Editorial activities

- **Associate editor** for Geophysical Journal International since January 2016
- **Reviewer** for the journals
  - SIAM Journal on Scientific Computing
  - Inverse Problems
  - Journal on Scientific computing
  - Geophysics
  - Geophysical Journal International
  - Geophysical Prospecting
  - Computers and geosciences
  - Journal of Applied Geophysics
  - Engineering Analysis with Boundary Elements

9 Invitations

- **2017.** *ANR MAGA meeting (Grenoble)*

  Optimal transport for signed measures: application to seismic imaging.

- **2017.** *OILWS2 workshop: Full Waveform Inversion and Velocity Analysis, IPAM, UCLA (Los Angeles)*

  On the use of Wasserstein distances for full waveform inversion.

- **2016.** *ANR Optiform/Geometrya meeting, Ecole Normal Superieure (Paris)*

  Optimal transport distance for seismic imaging.

- **2016.** *Séminaire Maison de la Modélisation (Saclay)*

  An implementation of an optimal transport distance for full waveform inversion.

- **2015.** *4ièmes Journées Scientifiques Equip Meso : Sciences de l'Univers (Toulouse)*

  Full waveform inversion for high resolution seismic imaging: HPC issues on recent applications and ongoing research.

- **2015.** *Séminaire MATHIAS TOTAL (Paris)*

  High resolution seismic imaging through full waveform inversion: successful attempts and challenges for the next years.

- **2015.** *Platform for Advanced Science Computing (PASC) conference 2015, ETH (Zurich)*

  Multi-parameter Full Waveform Inversion: tentative simultaneous reconstruction of v_P, ρ, Q_p in the 2D acoustic approximation.

- **2015.** *Joint Inversion Summer School (Barcelonnette)*

  Numerical optimization and adjoint state methods for large-scale nonlinear least-squares problems.

- **2014.** *Journées Ondes Sud-Ouest, ONERA (Toulouse)*

  A robust absorbing layer for seismic wave propagation in anisotropic media.

- **2013.** *Colloquium Julich Forschungszentrum (Julich)*

  Second-order derivatives approximation for multi-parameter Full Waveform Inversion.

- **2013.** *Basel Mathematisches Institut (Bâle))*

  Full Waveform Inversion using the truncated Newton method.

- **2013.** *ENSTA-POEMS seminar (Saclay)*

  Utilisation d'une méthode de Newton tronqué préconditionnée pour l'imagerie de structures complexes par inversion de formes d'ondes complètes.

- **2012.** *DYSCO seminar (Allevard)*

  Full Waveform Inversion using the truncated Newton method : imaging complex subsurface structures.

- **2012.** *EDP Rhône-Alpes meeting (Chambéry)*

  Numerical methods for Full Waveform Inversion.

- **2012.** *LJK Team EDP seminar, Univ. Joseph Fourier (Grenoble)*

  Numerical challenges for full waveform inversion.

- **2011.** *Sergio Stecco department seminar at Florence University (Florence)*
  Stiff Differential Algebraic Equations solver for a reactive transport problem.
- **2011.** *LGIT seminar, Univ. Joseph Fourier (Grenoble)*
  A nonlinear inversion for high resolution seismic imaging.
- **2011.** *INRIA team ESTIME seminar, Rocquencourt (Paris)*
  A nonlinear inversion for high resolution seismic imaging.
- **2011.** *IHP seminar (Paris)*
  A nonlinear inversion for high resolution seismic imaging.
- **2010.** *LJK Team EDP seminar, Univ. Joseph Fourier (Grenoble)*
  A nonlinear inversion for high resolution seismic imaging.

Chapter 2

Research activity

Contents

1 Introduction: a state of the art of full waveform inversion

1.1 Generalities

Full waveform inversion is an imaging technique developed in the framework of reflection seismology in the beginning of the 80s (Lailly, 1983a; Tarantola, 1984). Local measurements of the displacement (or pressure) wavefield at the surface (and/or in boreholes) constitute the observed data. The method is formulated as the minimization of the distance between this observed data and data simulated through the solution of a system of partial differential equations modeling the wave propagation within the subsurface. Different levels of approximation can be used, from the acoustic approximation, to more realistic visco-elastic anisotropic wave propagation. The minimization is performed over one or several parameters of these equations, related to the mechanical properties of the subsurface. As such, FWI can be seen as a parameter identification problem, or a PDE constrained optimization problem (Nocedal & Wright, 2006). In general, the parameters mainly influencing the wave propagation within the subsurface are the compressional wave (P-wave) and shear wave (S-wave) velocities. In the acoustic approximation, the latter is neglected as only compressional waves are considered. Secondary parameters involve density, attenuation coefficient, and anisotropy parameters.

The FWI method is usually formulated as a least-squares minimization problem, *i.e.* the distance used to measure the discrepancy between the observed and synthetic data is the one associated with the L^2 norm. The wave equation is discretized through finite-differences (Virieux, 1986; Levander, 1988) (well adapted to the acoustic case for data acquired on sea) or finite-element strategies (Etienne et al., 2010; Brossier et al., 2010; Barucq et al., 2016; Komatitsch et al., 2002; Komatitsch & Tromp, 2002b) (well adapted to elastic propagation to include accurate approximation of the topography for data acquired on land). For realistic applications, the discrete inverse problem usually involves millions to billions unknown parameters. For this reason, the minimization problem is solved through local optimization solvers, based on quasi-Newton strategies (Nocedal & Wright, 2006). Global or semi-global optimization strategies (Monte-Carlo Markov Chain, genetic algorithms, simulated annealing) are not adapted to solve this problem unless a re-parameterization is used to reduce drastically the number of discrete unknowns down to few tenths to one hundred. However, such re-parameterization are difficult to define except for schematic case studies (gradient media, 1D stratified media).

The quasi-Newton strategies require the ability to compute efficiently the misfit function and its gradient. This is made possible through the use of adjoint techniques, originating from optimal control theory (Lions, 1968; Plessix, 2006). In the FWI context, the gradient of the misfit function through the adjoint technique amounts to the correlation of two wavefields: the incident wavefield originating from the source used to generate the data, and the adjoint wavefield, originating from the receivers used to acquire the data. These two wavefields are thus solution of the same wave propagation equation with different source terms. However, the incident wavefield is computed forward in time (from an initial condition) while the adjoint wavefield is computed backward in time, from a final condition. This adjoint technique is crucial for FWI as it enables a fast gradient computation, which does not require the computation of the sensitivity matrix of the wavefield with respect to the subsurface parameters (also called the Fréchet derivatives matrix). For realistic size applications, storing/computing this matrix would require prohibitively large storage facilities/computational resources.

Compared to standard imaging technique, such as travel-times tomography, based on the interpretation of the arrival times only, FWI aims at interpreting the whole signal, without a specific pre-processing implying any identification of seismic events such as transmitted waves, reflected waves, refracted waves, converted waves. Such identifications are at the heart of standard geophysical strategies to infer mechanical properties from the measurement of the wavefields. In turns, FWI should be able to take into account all these events, and produce high resolution estimates of the mechanical properties of the subsurface, in the limit of half the shortest wavelength of the propagated signal (Devaney, 1984).

At the time FWI was introduced, two important limitations were identified for its proper use on seismic reflection data. First, repeatedly solve 2D or 3D wave propagation problems requested a computational power not available at this time. Second, a sensitivity gap was identified between the low wavenumber information possibly reconstructed through arrival-time tomography strategies, and the

high wavenumber information carried out by reflection data (Gauthier et al., 1986; Jannane et al., 1989).

Because of this sensitivity gap, the method was first thought of as not suitable for seismic imaging. The work of Pratt & Worthington (1990); Pratt (1990) changed this viewpoint. Instead of considering reflection data, they considered cross-hole acquisition survey, for which sources and receivers are located in two boreholes. This acquisition configuration allows to record mainly transmitted energy instead of reflected energy. Transmitted data contains more information on the intermediated wavenumber which are lacking in the reflected data. In addition, Pratt & Worthington (1990) set up a multi-scale frequency-domain strategies. The data is decimated in few discrete frequencies to reduce the computational cost of the overall strategy and the volume of data to manage. The FWI strategy interprets the low frequency part of the data first. The reconstructed velocity model is then used as as starting model for the interpretation of higher frequency data. This multi-scale approach allows for the reconstruction of low to high resolution feature of the wave velocity. This very important contribution was a proof of concept for FWI emphasizing the importance of transmitted waves and the possibility to set-up a suitable computational framework through a frequency-domain strategy.

The joint development, in the beginning of the 2000s, of wide azimuth broadband acquisition systems, and high performance computing facilities, was the starting point for successful applications of FWI to seismic exploration data. To illustrate the interest of the FWI strategy and particularly the increase in resolution that can be achieved compared to travel-time tomography strategies, I present in the following the results obtained by Operto et al. (2015b) on a shallow water ocean bottom cable (OBC) data, acquired in the North Sea, near the Valhall oil and gas field.

1.2 An example of application at the seismic exploration scale

A map of the Valhall field and of the acquisition geometry is presented in Figure 1.1. The targeted area covers a surface of 145 km^2. A recording layout of 12 cables equipped with four-component receivers is located on the sea bottom. The shallow water environment (the water layer is no deeper than 60 m in this area) is favorable to the deployment of such devices. The nominal distance between cables is 300 m, with the two outer cables at 600 m, for a total of 2302 receivers. The data is acquired using airgun pressure sources towed along cables at 5 m below the sea surface, for a total of 49954 shots. The inline spacing between two consecutive shots is 50 m. The source–receiver reciprocity is used, the hydrophone components being considered as explosive sources while the shots are considered as hydrophones sensors. The purpose it to reduce the number of forward modeling during FWI by one order of magnitude (from 49954 to 2302).

An instance of seismic data associated with one seismic shot and a line of receivers is presented in Figure 1.2. The vertical axis corresponds to a time-axis, while the horizontal axis corresponds to the receiver spatial position. This a typical representation of data in seismic exploration. The black and white color scale indicates respectively positive and negative values. The complexity of the signal illustrates the complexity of the wave propagation beneath the subsurface due to its heterogeneous mechanical properties. Direct waves propagating in the water column can be identified, as well as diving waves traveling in the upper part of the target, and reflected waves, recorded at later times by the receivers.

The FWI problem is solved in the 3D visco-acoustic frequency-domain approximation, meaning that the wave propagation modeling engine is based on the following generalized Helmholtz equation

$$-\frac{\omega^2}{c^2}p - \mathrm{div}\left(\frac{1}{\rho}\nabla p\right) = s, \tag{1.1}$$

where p is the monochromatic pressure wavefield, ω is the angular frequency, $c(x)$ is the pressure wave velocity (P-wave velocity) and $\rho(x)$ is the density. A compact finite-difference scheme is used to discretize this equation (Hustedt et al., 2004; Operto et al., 2014b). The resulting linear system is solved with a direct solver. The matrix associated to this linear system is factorized as a product of lower and upper triangular matrices (LU decomposition). The parallel solver MUMPS (Amestoy et al., 2000; MUMPS team, 2015) is used to this purpose. Given the large number of right hand sides

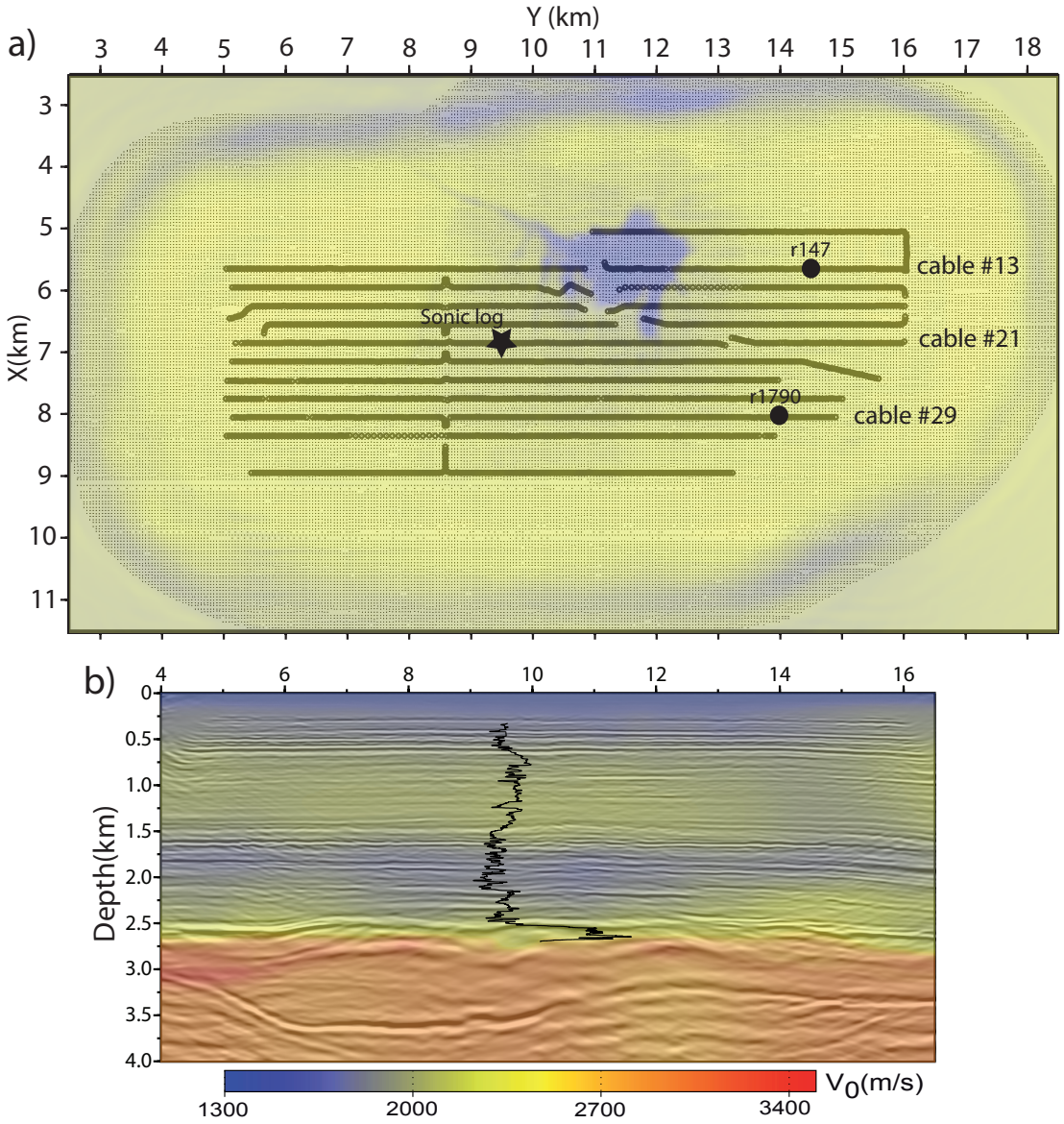

Figure 1.1 – (a) OBC acquisition layout. The ocean-bottom cables are denoted by red lines. The shot positions by black dots. Two receiver positions r_1 and r_2 at (X,Y)=(5.6 km,14.5 km) and (X,Y)=(8021 m,14108 m), respectively, are denoted by circles. An horizontal slice of the gas cloud at 1 km in depth, which was obtained by FWI, is shown in transparency to delineate its zone of influence. The star denotes the position of the well log. Positions of cables 13, 21 and 29 are labeled. (b) Reverse time migrated image with superimposed in transparency vertical wave velocities determined by reflection traveltime tomography.The sonic log is also shown by the black line (adapted from Prieux et al. (2011)). Figures extracted from Operto et al. (2015b).

(associated with the number of sources) which have to be considered at each iteration of the inversion, this strategy is of particular interest as once the matrix is factorized, the solution of the linear system

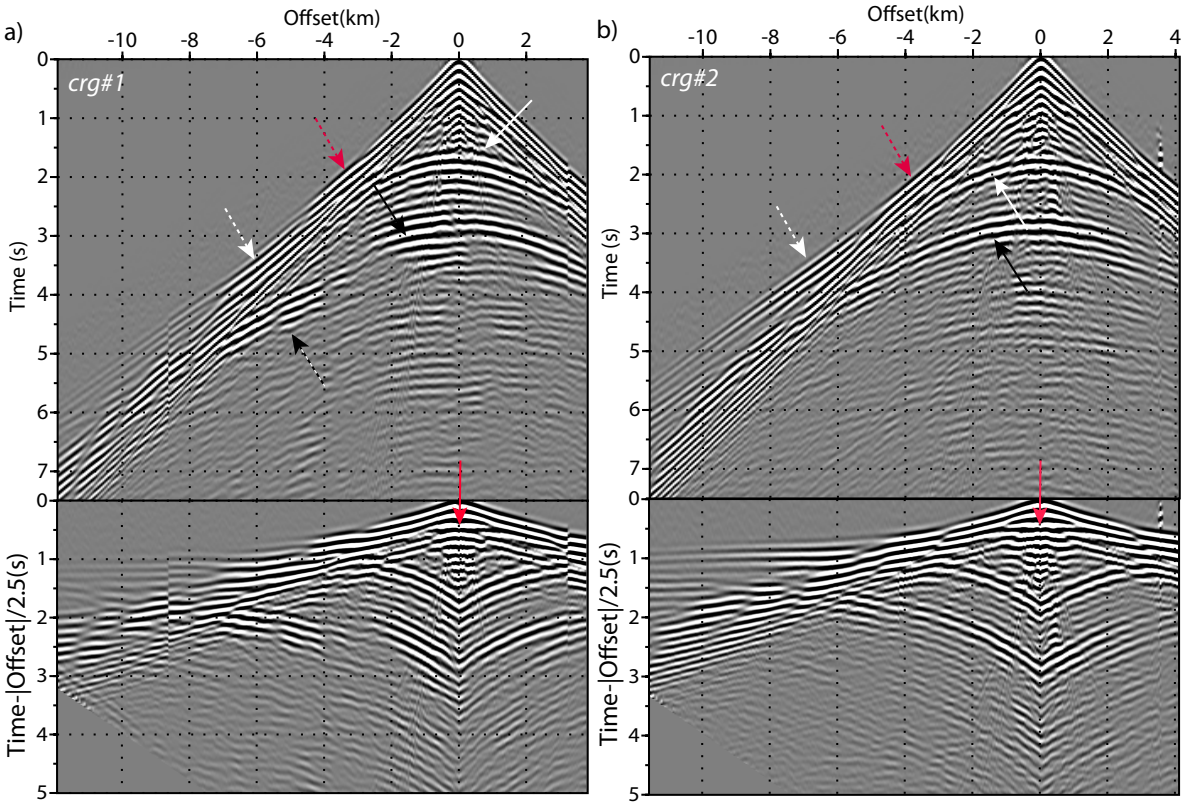

Figure 1.2 – Time-domain common-receiver gathers for inline (Y) shot profile running across the receiver position. (a) Receiver r_1. (b) Receiver r_2. The data are plotted with a reduction velocity of 2.5 km.s^{-1} on the bottom panel. The red, white, black arrows point on the reflection from a shallow reflector, the top of the low-velocity zone and the top of the reservoir, respectively. The solid arrow points on the pre-critical reflections, while the dashed ones points on the post-critical reflections. The critical distance is difficult to identify for the reflection from the top of reservoir because of the interference with multi-refracted waves at the sea bottom. Figure extracted from (Operto et al., 2015b).

for a given right-hand-side can be computed efficiently through forward/backward substitutions. This however requires high performance computing devices with sufficient in-core memory as even if the finite-difference matrix is sparse, the L and U factors are dense over the bandwidth of the initial matrix. Using today architectures, this strategy is feasible for 3D wave propagation modeling over distances up to 100 wavelengths in each direction.

The minimization problem amounts to find the P-wave velocity $c(x)$ in equation (1.1) which minimizes the L^2 distance between observed and synthetic data. It is solved through a l-BFGS quasi-Newton algorithm (Nocedal, 1980; Nocedal & Wright, 2006). A frequency continuation/multiscale strategy is implemented. As set of 11 frequencies is considered from 3.5 Hz to 10 Hz. For each frequency, a FWI problem is solved, the final result of one inversion being used as an initial guess for the next inversion. The results are presented in Figures 1.3 (3D view) and 1.4 (vertical 2D slices). Starting from a low resolution estimation of the P-wave velocity obtained through conventional first-arrival traveltime tomography (Nolet, 2008), FWI yields a high resolution quantitative estimation of the P-wave velocity model. In particular, the gas cloud (low velocity anomaly) in the center of the model is precisely

delineated, as well as the high velocity horizontal reflector, the cap rock of the reservoir located just below. These results clearly illustrate the resolution potential of FWI compared to conventional imaging methods.

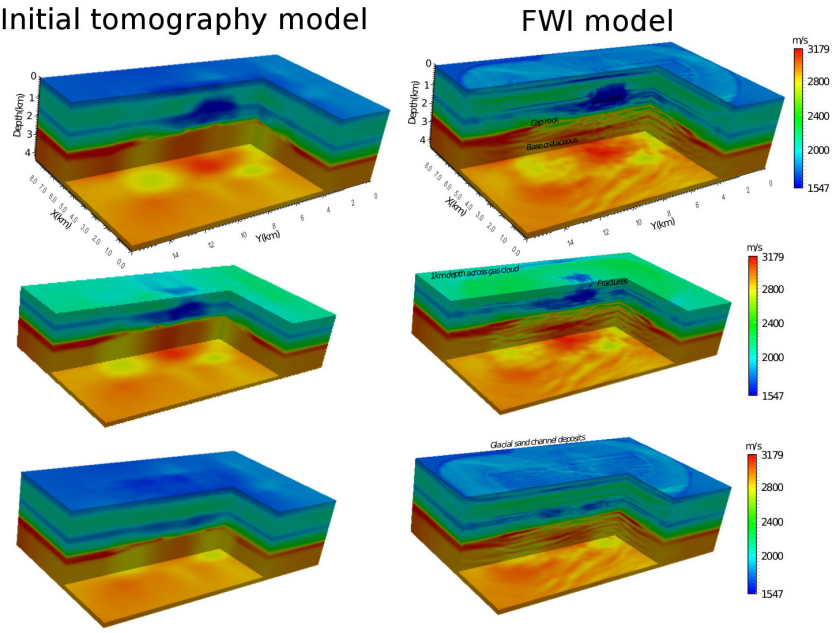

Figure 1.3 – Example of FWI application in the 3D acoustic approximation. The Pressure wave (P-wave) velocity is inverted from Ocean Bottom Cable (OBC) seismic data recorded near the Valhall oil field in North sea. From a low resolution estimation computed through tomography (left), FWI builds a high resolution estimation of the P-wave velocity, which helps in the geological interpretation of the target. In particular, the low velocity anomaly in blue, which corresponds to the presence of a gas pocket, is clearly delineated compared to the tomography estimation. Figure extracted from (Operto et al., 2015b).

1.3 Other application fields

While FWI has been mainly developed in the framework of seismic exploration, its fairly general formalism enables to use it at very different scales. We review here other fields where FWI is used (or starts being used).

1.3.1 Seismology

FWI is used in seismology, for regional to global subsurface imaging. The main difference with seismic exploration is related to the type of data which is used. While seismic exploration relies on controlled source experiments, where dense arrays of sources and sensors are used to image a localized target, in seismology the data is generated by earthquakes, and recorded by few seismometers. As a result, the coverage is drastically sparser, and unevenly located. Compared to seismic exploration however the recorded signal is less complex, as the data emitted by earthquakes travel at lower frequencies. This makes easier data pre-processing for the identification of particular seismic events (phases). Another difference with seismic exploration also relies on the necessity to use elastic wave propagation modeling to correctly model surface waves resulting from the interaction between body waves and the free surface. Their role cannot be neglected in the inversion, while such waves do not exist for marine exploration data (such as the Valhall case study). For this reason, FWI at global and regional scales relies mainly on 3D

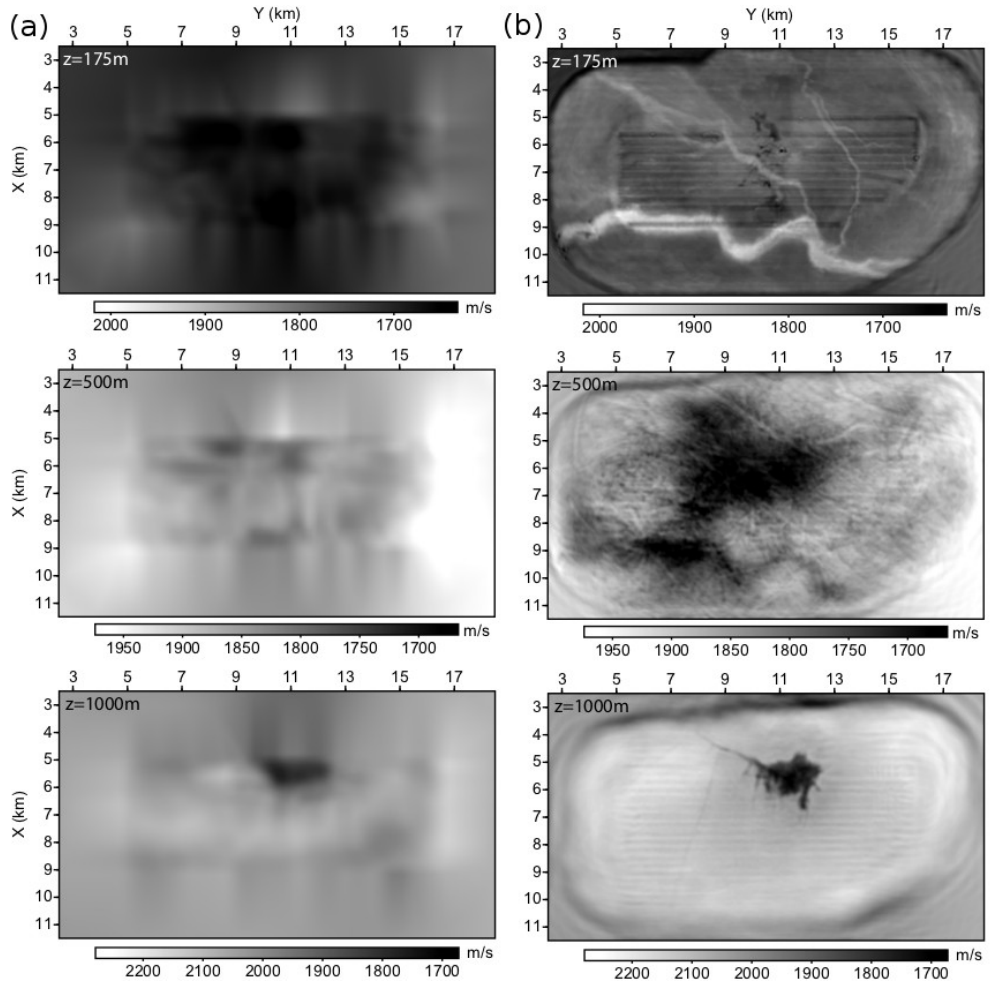

Figure 1.4 – 2D horizontal slices (constant depth z) of the 3D models presented in Figure 1.3. Slices in the initial tomography model (a). Slices in the final FWI model (b). At shallow depth ($z = 175$ m), FWI reveals the imprint of the OBC acquisition system, as well as the presence of paleochannels. At 500 m depth, the elongated structures correspond to scraps left on the sea bottom by deriving icebergs. Deeper ($z = 1$ km), the gas pocket is clearly delineated. The elongated structure on the right side of the gas pocket suggest the presence of a chimney, which could explain the movement of the gas up to the shallowest layers of the model. Figure extracted from (Operto et al., 2015b).

finite element modeling schemes, in particular the spectral element strategy introduced by Patera (1984) and brought later to the seismology community (Komatitsch et al., 2002). The interest in seismology for using FWI compared to more conventional tomography methods is the same than in seismic exploration: this strategy yields estimation of the subsurface mechanical properties with unprecedented resolution. Examples of application at the global scale (Fichtner et al., 2008; Liu & Gu, 2012; Bozdağ et al., 2016) and the regional scales (Fichtner et al., 2009; Tape et al., 2009; Zhu et al., 2012) demonstrate the interest of FWI for seismology.

1.3.2 Near surface imaging with seismic data

FWI starts also being used for near-surface imaging (10 to 100 m depth) using seismic waves. At this scale, the information carried out by the propagation of surface waves is used to infer the properties

of the subsurface. As these waves are highly dispersive, state-of-the-art techniques rely mainly on the interpretation of dispersion curves, based on the assumption of 1D stratified media (Socco et al., 2010). The interest for using FWI at this scale is to overcome this 1D limitation to produce fully 3D high resolution estimates of the subsurface mechanical properties. However, the high frequency content of surface waves and the impact of attenuation make successful applications FWI even harder at this scale. First instance of applications in 2D approximations have been proposed by Groos et al. (2014); Pérez Solano et al. (2014); Schäfer et al. (2014); Masoni et al. (2014a)

1.3.3 Near surface imaging with ground penetrating radar data

Imaging of the subsurface at shorter scales is possible through the analysis of the propagation of electromagnetic waves, which give access to information on the permittivity and the resistivity of the materials. This is the purpose of ground penetrating radar (GPR) acquisition devices. Recovering the electromagnetic properties of the subsurface show to be particularly useful for hydrological and geotechnical applications, these properties being particularly sensitive to the presence of water. Exactly as for seismic imaging, FWI yields high resolution quantitative estimations of these properties. The specificity of the application of FWI to electromagnetic data relies on the higher contrasts of the material properties (1 to 40 for permittivity, while at the exploration scale, velocity contrasts are usually comprised between 1 to 4). The attenuation also plays an important role on the wave propagation, making mandatory to consider simultaneously both permittivity and conductivity in the inverse problem, in a multi-parameter fashion. First instances of applications of FWI to GPR data have been proposed by Busch et al. (2012); Meles et al. (2012); Kalogeropoulos et al. (2013); Klotzsche et al. (2013); Lavoué et al. (2014)

1.3.4 Ultrasound imaging

Finally, FWI starts being considered as a potential strategy to increase the resolution of ultra-sound imaging. This kind of technology is employed in nondestructive testing to assess the quality of industrial materials and to detect potential default/weaknesses. It is also used in medical imaging to control the sanity of the tissues, notably for cancer detection. Again, FWI could be used in this framework to increase the reliability of the inversion and the resolution of the images. These kind of applications are particularly promising for FWI as the data coverage could be much more favorable than for seismic imaging for which the sources and receivers are located only on the surface. The size and configuration of the targets investigated through ultrasound imaging make possible to use acquisition devices with a far better coverage/illumination of the structure. This is for instance the case for breast cancer detection, for which FWI has yield promising results (Pratt et al., 2007; Roy et al., 2016).

1.4 Contributions presented in this manuscript

FWI appears as a very promising technique for general imaging problems, from global and regional seismic scales using earthquake data, to millimeter scales through ultra-sound imaging. However, applying FWI at this diversity of scales raises numerous challenges, asking for a better controlled and robust methodology.

Current issues for FWI consist now in going towards more complex physical modeling of the wave propagation, accounting for elasticity, viscosity and anisotropy. Another topic is related to the design of efficient multi-parameter inversion strategies, going beyond the simple reconstruction of the P-wave velocity model toward the reconstruction of secondary attributes such as density, attenuation factor, and anisotropy parameters. These secondary parameters bring substantial information to better characterize the subsurface. Finally, another general topic relies in the design of more automated full waveform inversion strategies, not requiring a strong human expertise on data processing for successful applications. This mainly relies on the way the misfit function is designed, as the standard least-squares misfit function is non convex, leading the local optimization algorithms to converge towards local minima. My scientific contributions try to address these topics. In the following, I present in details three of these contributions, before giving a more synthetic overview of other topics I have been working on.

The first contribution presented in detail is related to the design of numerical strategies for the modeling of seismic waves in the elastic approximation. More specifically, it is related to the design of suitable absorbing boundary conditions. Except at the global scale, the subsurface is considered as a semi-infinite half-plane, raising the question how to design efficient boundary conditions to prevent from the introduction of parasite reflections in numerical modeling tools. While Perfectly Matched Layers (PML) (Bérenger, 1994) have become the method of choice for this kind of applications, instabilities/amplification have been observed in the framework of elastic modeling, in particular in the presence of (anelliptical) anisotropy. We thus have proposed an alternative absorbing layer strategy, based on the work of Halpern et al. (2011), called SMART layers. This layer is shown to be less accurate than PML ones, however more robust, with a mathematical proof of non-amplification.

The second contribution is related to the use of a second-order optimization scheme for FWI, named truncated Newton strategy. It is based on a more accurate approximation of the inverse Hessian operator (local curvature of the misfit function) than the one obtained through quasi-Newton strategies. We review what is the physical interpretation of this operator in the context of FWI, why it is important to approximate it accurately, especially for multi-parameter problems, and how this is possible, for realistic size problem, through an implementation of Hessian-vector products with a second-order adjoint state method.

The third contribution is related to the design of a misfit function based on an optimal transport distance, to mitigate the non-convexity of the standard least-squares misfit function with respect to the wave velocity model. This non-convexity is one of the main limitation of FWI since it was introduced. Optimal transport distances exhibit very interesting properties that could help to overcome this difficulty. A first implementation on 2D and 3D acoustic time-domain problems is presented.

2 Robust absorbing layers for elastic wave modeling

This section summarizes the work presented in the following publications and expanded abstracts.
Publications:

- Métivier, L., Brossier, R., Labbé, S., Operto, S., & Virieux, J., 2014c. Smart: dissipative absorbing layer technique for general elastodynamics equations. application as s-waves filter in acoustic ti media., *Seismic Technology*, **11**(4), 14
- Tago, J., Métivier, L., & Virieux, J., 2014. SMART layers: a simple and robust alternative to PML approaches for elastodynamics, *Geophysical Journal International*, **199**(2), 700–706
- Métivier, L., Brossier, R., Labbé, S., Operto, S., & Virieux, J., 2014b. A robust absorbing layer for anisotropic seismic wave modeling, *Journal of Computational Physics*, **279**, 218–240

Abstracts:

- Métivier, L., Brossier, R., Operto, S., & Virieux, J., 2014f. Smart: Robust absorbing layer and s-waves filtering for acoustic anisotropic wave simulation, in *Expanded Abstracts, 84^{th} Annual SEG Meeting (Denver)*
- Métivier, L., Brossier, R., Operto, S., & Virieux, J., 2014d. A robust absorbing layer method for seismic wave simulation in anisotropic media, in *Expanded Abstracts, 76^{th} Annual EAGE Meeting (Amsterdam)*

2.1 Context and problematic

First-order absorbing boundary conditions (ABC), also known as radiation boundary conditions, were introduced in the pioneering studies of Clayton & Engquist (1977) and Engquist & Majda (1977). They are easy to implement for simple wave equations models, such as the acoustic wave equation. In a mono-dimensional context, these equations are exact: the outgoing waves are absorbed without introducing spurious reflections. However, the extension to multi-dimensional problems is not trivial. In particular, waves arriving to the boundary with normal incidence are correctly absorbed, while waves propagating to the boundary with grazing incidence angles generate spurious reflections.

An improvement of these ABC can be achieved through the design of higher-order versions, as proposed initially by Collino (1993) or more recently by Givoli (2001). The accuracy of these higher order ABC is better than standard first-order ABC, however their implementation is not trivial, as they imply the use of fractional high-order derivatives, which are nonlocal operators. In addition, the extension of ABC to more complex wave propagation description, such as elasto-dynamic, or acoustic anisotropy, is not straightforward. In this case, the approximation of the differential operator at the boundary yields a complex system of equations to solve. In practice, it is difficult to guarantee the stability of such methods, and a correct absorption of waves at all incidence angles.

An alternative to absorbing boundary conditions consists in the design of absorbing layers: the domain of interest is surrounded with a layer where waves incoming from the domain of interest are artificially damped. This idea has first been promoted by Cerjan et al. (1985) for the second-order in time acoustic equation. Despite this simple formalism, in practice, the design of such absorbing layers is difficult. Except for the 1D problem, the introduction of the layer generates reflections at the interface between the domain of interest and the layer. These reflections can be mitigated by choosing variable damping coefficient that smoothly grow from zero at the interface between the layer and the domain of interest, to the external boundary of the layer. However, this reduces the absorbing capability of the layer and requires to increase its size, which, in turn, increases the overall computational time of the simulation.

The Perfectly Matched Layers (PML) method, introduced by Bérenger (1994), has become rapidly popular as it achieves an excellent trade-off between these two contradictory requirements. The initial method is based on a splitting of the hyperbolic system and the introduction of smooth damping coefficients in the layer. The PML strategy was originally designed for the 2D and 3D Maxwell's equations (Bérenger, 1994, 1996). For these equations, a plane wave analysis demonstrates that the reflection coefficient at the interface between the domain of interest and the layer is null for wave

propagating at all angles. In practice, the reflectivity for the discrete problem is not exactly zero, and when the PML are used in heterogeneous models, a WKB analysis shows that only the leading order of the reflection coefficient is zero (see the review paper by Halpern et al. (2011)). However, in many practical applications the amplitude of the spurious reflected waves remains very small.

Because of this remarkable property, the PML methods is now the standard for the simulation of wave propagation in infinite or semi-infinite medium. The method has been progressively extended to different wave propagation systems, starting from acoustic wave equations, which are in 2D similar to the Maxwell's equations (Qi & Geers, 1998; Diaz & Joly, 2006; Bermúdez et al., 2007). The PML method has then been applied to linearized Euler equations (Hu, 1996; Hesthaven, 1998; Abarbanel et al., 1999), and to linear elasticity (Hastings et al., 1996; Collino & Tsogka, 2001; Basu & Chopra, 2003; Komatitsch & Tromp, 2003; Appelö & Kreiss, 2006)

While the performances of the PML have been particularly convincing for Maxwell's, acoustic and linearized Euler equations, their application to the modeling of elastic seismic waves is less satisfactory. In particular, it has been observed that the waves arriving at grazing angles to the interface between the domain of interest and the layer generate strong spurious reflections. To overcome this difficulty, the design of Convolutional PML (C-PML) (Komatitsch & Martin, 2007) has been proposed. The method is based on a generalization of the complex-valued stretching related to the standard PML formulation in the frequency-domain to achieve a better absorption at grazing angles. This implies a loss of the perfectly matched property of the layer, even at the continuous level. The method can thus be seen as a way of finding a better compromise between a quasi-perfect absorption at near normal incidence and a poorer absorption at grazing angles.

However, when applied to the modeling of wave propagation in anisotropic media, the PML and C-PML methods become amplifying, which causes difficulties for their application to this particular case. This amplification has first been observed in the case of 2D linear elasticity for VTI media by Becache et al. (2003), an later on in the case of 2D acoustic TTI equations (Operto et al., 2009; Duveneck & Bakker, 2011; Zhang et al., 2011). The modification of the PML through the C-PML technique does not prevent from this amplification phenomena (Komatitsch & Martin, 2007).

The amplification analysis proposed by Becache et al. (2003) is based on slowness diagrams. In particular, they show that when the slowness vector and the group velocity are in opposite directions (negative scalar product), the PML becomes amplifying. This amplification phenomenon, numerically experimented in the aforementioned studies, has also been formalized through a WKB analysis by Halpern et al. (2011). Depending on the symbol of the propagation operator associated with the PML strategy, one can see that the amplification phenomenon is indeed common. Acoustic and electromagnetic wave equations are particular cases for which the amplification does not occur, but this is not true in general.

These analysis suggest that the use of PML (or variants such as C-PML) for general elastodynamics equation is not the appropriate choice, raising again the question of a suitable choice of absorbing conditions. A generalization of the standard PML scheme has been proposed under the name of Modified PML (M-PML) by Meza-Fajardo & Papageorgiou (2008). Whereas for standard PML, the damping coefficients introduced in the layer vary only in the direction normal to the interface between the domain of interest and the layer, the M-PML strategy introduces additional absorption terms in the directions tangential to this interface. The introduction of this tangential damping actually can be interpreted as the definition of a trade-off between a standard PML (no tangential absorption) and a standard sponge layer (same level of tangential and normal absorption). The M-PML is thus not perfectly matched (Dmitriev & Lisitsa, 2011). In their study, Meza-Fajardo & Papageorgiou (2008) show that a sufficient level of tangential anisotropy may prevent the amplification phenomenon. However, it is difficult to correctly define this level as this is intrinsically problem dependent. Besides, the required amplitude may be large, making the layer method close from a sponge layer, which may yield high amplitude spurious reflections.

2.2 Proposed approach: SMART layers method

We have investigated how another absorbing layer method, named SMART layers (Halpern et al., 2011), can be applied to the general elastodynamics equations to ensure the robustness of the solution to these equations. The SMART layer is a generalization of a layer method initially introduced by Israeli & Orszag (1981) as an approximation of radiation boundary conditions. A diagonalization of the matrices which compose the leading symbol of the first-order hyperbolic operator yields a natural decomposition of the solution into components propagating inward and outward of the domain of interest in each spatial direction. This decomposition is used to select the components to be damped, depending on the position in the computational domain. In the domain of interest, no damping is applied, while in the left layer, only components of the solution propagating leftward are damped, and similarly for the other sides. This decomposition is applied directly to the continuous equation, and results in the introduction of a zero-order term, in a fairly general formalism.

Compared to the PML method, the reflectivity of the SMART layers at the interface between the computational domain and the layer is larger: the SMART layers are not perfectly matched. However, the amplitude of this reflection can be controlled, as for the PML strategy, through the choice of appropriate absorption coefficients and the definition of the layer width. The main interest of the SMART method compared to the PML is its robustness: we demonstrate that the zero-order term added to the hyperbolic system by the SMART layer method is dissipative as soon as the initial system satisfies a symmetrizability condition. This ensures no artificial amplification of the solution. The symmetrizability condition requires the existence of a symmetrizer common to all the matrices which compose the leading-order symbol of the first-order differential operator.

2.3 Mathematical description

2.3.1 SMART layers model and stability theorem

Consider the general first-order hyperbolic system

$$\begin{cases} \partial_t u + \sum_{j=1}^{d} A_j(x)\partial_j u + A_0(x)u = f(x,t), \quad (x,t) \in \mathbb{R}^d \times [0,T], \quad d \in \mathbb{N}, \\ u(x,0) = u_0(x), \end{cases} \tag{2.1}$$

where d is the dimension of the physical space and

$$\begin{cases} u(x,t) \in L^2\left(\Omega \times [0,T], \mathbb{R}^p\right), \quad f(x,t) \in L^2\left(\Omega \times [0,T], \mathbb{R}^p\right), u_0(x) \in L^2\left(\Omega, \mathbb{R}^p\right), \\ \forall x \in \mathbb{R}^d, \quad A_j(x) \in \mathbb{M}_p\left(\mathbb{R}\right). \end{cases} \tag{2.2}$$

Provided the matrices $A_j(x), j = 1, \dots d$ are diagonalizable with real eigenvalues, the SMART layer is built upon spectral projectors on the eigenspaces of these matrices. For each matrix $A_j(x)$, the solution of the system (2.1) can be decomposed in the basis formed by its eigenvectors. Components of the solution associated with positive eigenvalues (respectively with negative eigenvalues) of the matrix $A_j(x)$ propagate in the direction $x_j > 0$ (respectively in the direction $x_j < 0$). Their propagation velocity is equal to the modulus of the eigenvalue they are associated with. The SMART layer system derived from (2.1) is written as

$$\begin{cases} \partial_t u + \sum_{j=1}^{d} A_j(x)\partial_j u + A_0(x)u + B(x)u = f(x,t), \quad (x,t) \in \Omega \times [0,T], \quad \Omega \subset \mathbb{R}^d, \quad d \in \mathbb{N}, \\ u(y,t) = g(y,t), \quad (y,t) \in \partial\Omega \times [0,T], \\ u(x,0) = u_0(x), \end{cases} \tag{2.3}$$

where Ω is the bounded computational domain, $\partial\Omega$ its boundary, and $g(y,t)$ the boundary condition imposed on $\partial\Omega$.

The additional zero-order term $B(x)$ is a linear combination of the spectral projections on the eigenspaces of the operators $A_j(x)$ in the layer. We introduce the coefficients $d_j^{\pm}(x_j)$ as smooth mono-dimensional functions which are zero in the domain of interest (without the layers) denoted by $\mathring{\Omega}$. These

coefficients have a smooth growth in the layer (cubic or cosine functions are often used in practice). The operator $B(x)$ is

$$B(x) \quad = \quad \sum_{j=1}^{d} \sum_{n=1}^{n_j^+} d_j^+(x_j) B_{j,n}^+(x) + \sum_{j=1}^{d} \sum_{n=1}^{n_j^-} d_j^-(x_j) B_{j,n}^-(x), \tag{2.4}$$

where n_j^+ (respectively n_j^-) denotes the number of positive (respectively negative) eigenvalues of the matrix $A_j(x)$ and $B_{j,n}^+(x)$ (respectively $B_{j,n}^-(x)$) denotes the spectral projector on the corresponding eigenspace. The coefficients $d_j^\pm$ ensure that no attenuation is introduced in the domain of interest $\mathring{\Omega}$, as $B(x)$ is zero in $\mathring{\Omega}$.

For comparison, the standard sponge layer technique amounts to defining $B(x)$ as

$$B(x) = \sum_{j=1}^{d} d_j^\pm(x_j) I = \alpha(x) I, \tag{2.5}$$

where I is the identity operator, and $\alpha(x)$ is defined by

$$\alpha(x) = \sum_{j=1}^{d} d_j^\pm(x_j). \tag{2.6}$$

We now introduce the

Definition 1. *The system* (2.1) *is symmetrizable if and only if there exists a symmetrizer* $S(x) \in \mathbb{M}_p(\mathbb{R})$ *such that*

$$\forall x \in \Omega, \;\; S^T(x) = S(x), \;\; S(x) > 0, \;\; (S(x)A_j(x))^T = S(x)A_j(x), \;\; 1 \le j \le d. \tag{2.7}$$

Definition 2. *The S-norm* $\|.\|_S$ *associated with a symmetric definite matrix* $S(x) \in \mathbb{M}_p(\mathbb{R})$ *is such that*

$$\|u(.,t)\|_S = (Su,u)_{L^2} = \int_\Omega < S(x)u(x,t), u(x,t) > dx, \tag{2.8}$$

where $< .,. >$ *denotes the Euclidean scalar product of* $\mathbb{R}^p$.

With these definitions at hand, we can prove the following theorem.

Theorem 1. *Assume that*

- *the system* (2.1) *is symmetrizable with a symmetrizer* $S(x)$*;*
- *the coefficient of the matrices* $S(x)A_j(x)$ *belong to* $\mathbb{W}^{1,\infty}$ *(regularity condition).*

Then

- *the SMART layer strategy can be applied to the system* (2.1)*: the matrices* $A_j(x)$ *are diagonalizable with real eigenvalues;*
- *the zero-order term* $B(x)$ *added by the SMART layer method is dissipative: this ensures that no spurious amplification will be introduced.*

In particular, we have

$$\frac{d}{dt}\|u(.,t)\|_S^2 \le (\gamma+1)\,\|u(.,t)\|_S^2 + \|f(.,t)\|_S^2 - 2\,(SBu,u)_{L^2}\,, \tag{2.9}$$

where

$$\gamma = \sum_{j=1}^{d} \sup_x |\partial_x(SA_j)| + \sup_x |(SA_0)|, \tag{2.10}$$

and

$$(SBu,u)_{L^2} \ge 0. \tag{2.11}$$

The latter inequality ensures that the SMART zero-order term does not generate any energy growth. No amplification can be caused by its introduction as it is a dissipative term.

2.3.2 Sketch of the proof

The proof is decomposed in two parts. First, we prove that because the matrices A_j are symmetrizable with a symmetric operator S they are diagonalizable with real eigenvalues (trivial). Second, we perform a standard energy estimate for the system (2.3) to obtain the energy inequality (2.9). The difficult part is to show that the quantity $(SBu,u)_{L^2}$ is positive. To do so we use the following lemma

Lemma 1. *Let $A \in \mathbb{M}_p(\mathbb{R})$. Let $S \in \mathbb{M}_p(\mathbb{R})$ symmetric positive definite such that SA is symmetric. Then the eigenspaces of A form an orthogonal direct sum for the scalar product induced by S on $\mathbb{R}^p$.*

The proof of this lemma is straightforward. Let $(u,v) \in \mathbb{R}^p \times \mathbb{R}^p$ be two eigenvectors of A

$$Au = \lambda u, \quad Av = \mu v, \tag{2.12}$$

where λ and μ are the eigenvalues associated with u and v respectively. We have

$$< Su, v >= \frac{1}{\lambda} < SAu, v >= \frac{1}{\lambda} < u, SAv >= \frac{\mu}{\lambda} < u, Sv >= \frac{\mu}{\lambda} < Su, v > . \tag{2.13}$$

Therefore, either

$$< Su, v >= 0, \tag{2.14}$$

and the eigenvectors u, v are orthogonal for the scalar product induced by S or

$$< Su, v >\neq 0, \quad \lambda = \mu, \tag{2.15}$$

which means that u and v belongs to the same eigenspace of A.

Now, as A_j is diagonalizable, the function $u(x,t) \in L^2(\Omega \times [0,T], \mathbb{R}^p)$ can be decomposed in a basis of eigenvectors $a_{j,i}(x)$ of the matrix $A_j(x)$ such that

$$\forall (x,t) \in \Omega \times [0,T], \quad u(x,t) = \sum_i \alpha_{j,i}(x,t) a_{j,i}(x). \tag{2.16}$$

Consider $B^+_{j,1}$, the spectral projector on the eigenspace $\ker(A_j - \lambda_{j,1} I)$, where $I \in \mathbb{M}_p(\mathbb{R})$ is the identity matrix and $\lambda_{j,1}(x)$ the first positive eigenvalue of the matrix $A_j(x)$. We have

$$B^+_{j,1}(x) u(x,t) = \alpha_{j,1}(x,t) a_{j,1}(x). \tag{2.17}$$

Therefore,

$$\begin{aligned} \left(SB^+_{j,1}u, u\right)_{L^2} &= \int_\Omega < S(x) B^+_{j,1}(x) u(x,t), u(x,t) > dx \\ &= \int_\Omega < \alpha_{j,1}(x,t) S(x) a_{j,1}(x), \sum_i \alpha_{j,i}(x,t) a_{j,i}(x) > dx \\ &= \int_\Omega \sum_i \alpha_{j,1}(x,t)\alpha_{j,i}(x,t) < S(x) a_{j,1}(x), a_{j,i}(x) > dx. \end{aligned} \tag{2.18}$$

Using Lemma 1, we find

$$\left(SB^+_{j,1}u, u\right)_{L^2} = \int_\Omega \alpha^2_{j,1}(x,t) < S(x) a_{j,1}(x), a_{j,1}(x) > dx. \tag{2.19}$$

As $S(x)$ is symmetric definite positive, we have

$$\left(SB^+_{j,1}u, u\right)_{L^2} \geq 0. \tag{2.20}$$

The same demonstration is valid for any spectral projectors $B^+_{j,i}$, $B^-_{j,i}$ on the eigenvalues of A_j, therefore, as B is a linear combination of these projectors weighted by positive functions $d^\pm_j(x_j)$ we have

$$(SBu, u)_{L^2} \geq 0. \tag{2.21}$$

The complete proof can be found in Métivier et al. (2014b).

2.4 Application to elastodynamics

Following the theorem 1, the SMART layer technique can be applied to any first-order symmetrizable hyperbolic system. Interestingly, we can prove the following theorem

Theorem 2. *The system of first-order elastodynamics equations is a first-order symmetrizable hyperbolic system.*

The proof is simple and relies on the symmetric definite positive property of the stiffness tensor. I reproduce it here. Under the velocity-stress formulation and the Voigt notations, the elastodynamics equations amount to the first-order hyperbolic system

$$\begin{cases} \partial_t u = \dfrac{1}{\rho}\left(A_x\partial_x\sigma + A_y\partial_y\sigma + A_z\partial_z\sigma\right) \\ \partial_t \sigma = C\left(A_x^T\partial_x\sigma + A_y^T\partial_y\sigma + A_z^T\partial_z\sigma\right), \end{cases} \tag{2.22}$$

where $u(x,t) \in L^2\left(\Omega \times [0,T], \mathbb{R}^3\right)$ and $\sigma(x,t) \in L^2\left(\Omega \times [0,T], \mathbb{R}^6\right)$ are the velocity displacement vector and the stress vector

$$u(x,t) = [u_x, u_y, u_z], \tag{2.23}$$

$$\sigma(x,t) = [\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{xz}, \sigma_{yz}, \sigma_{xy}]. \tag{2.24}$$

In addition, the quantity ρ is the density, C is the stiffness tensor reduced as the 6×6 matrix C_{ij}, and the matrices A_x, A_y, A_z are

$$A_x = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad A_y = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad A_z = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}. \tag{2.25}$$

The stiffness tensor C is by definition symmetric and positive definite. Its inverse $M = C^{-1}$, known as the compliance matrix, is thus also symmetric positive definite. A simple manipulation of the system (2.22) yields

$$\begin{cases} \rho\partial_t u = A_x\partial_x\sigma + A_y\partial_y\sigma + A_z\partial_z\sigma \\ M\partial_t\sigma = A_x^T\partial_x\sigma + A_y^T\partial_y\sigma + A_z^T\partial_z\sigma. \end{cases} \tag{2.26}$$

We define $w \in L^2\left(\Omega \times [0,T], \mathbb{R}^9\right)$ as

$$w = [u, \sigma]^T. \tag{2.27}$$

We rewrite (2.26) as

$$S\partial_t w = \widetilde{A}_x\partial_x w + \widetilde{A}_y\partial_y w + \widetilde{A}_z\partial_z w, \tag{2.28}$$

where

$$S = \begin{pmatrix} \rho I_3 & 0 \\ 0 & M \end{pmatrix}, \quad \widetilde{A}_j = \begin{pmatrix} 0 & A_j \\ A_j^T & 0 \end{pmatrix}, \quad j \in \{x, y, z\}. \tag{2.29}$$

In (2.29), I_3 denotes the identity matrix of $M_3(\mathbb{R})$. It is important to note that the matrix S is therefore symmetric positive definite, as it is block diagonal with positive definite blocks. In addition, the matrices $\widetilde{A}_j$ are symmetric. The elastodynamics equations (2.22) are thus equivalent to the first-order hyperbolic system

$$\partial_t w = S^{-1}\widetilde{A}_x\partial_x w + S^{-1}\widetilde{A}_y\partial_y w + S^{-1}\widetilde{A}_z\partial_z w. \tag{2.30}$$

An obvious symmetrizer for this system is the operator S.

This simple demonstration, inspired from the work of Burridge (1996), enlightens the symmetrizability of the elastodynamics equations, through the definition of the compliance matrix M.

2.5 Numerical studies

As an illustration, the method has first been tested on so-called "acoustic anisotropic" equations. These equations are used within the seismic exploration community to account for anisotropic effects (tilted transverse isotropy) while still focusing on the propagation of pressure waves only as it is the case for the acoustic approximation. The motivation is to improve the accuracy of the wave model while maintaining the computational cost of the acoustic approximation (one to two orders of magnitude lower to the one associated with elastic equations). These equations are derived from the elastodynamics equations including anisotropy and setting the shear-wave velocity to zero (Alkhalifah, 1998, 2000; Duveneck & Bakker, 2011). In 2D, these equations can be written as

$$\left\{\begin{array}{rcl}
\partial_t u_x &=& \frac{1}{\rho}\left[\partial_x\left(\cos^2\theta\sigma^\theta_{xx}+\sin^2\theta\sigma^\theta_{zz}\right)+\partial_z\left(\sin\theta\cos\theta\left(\sigma^\theta_{zz}-\sigma^\theta_{xx}\right)\right)\right]\\
\partial_t u_z &=& \frac{1}{\rho}\left[\partial_z\left(\sin^2\theta\sigma^\theta_{xx}+\cos^2\theta\sigma^\theta_{zz}\right)+\partial_x\left(\sin\theta\cos\theta\left(\sigma^\theta_{zz}-\sigma^\theta_{xx}\right)\right)\right]\\
\partial_t \sigma^\theta_{xx} &=& \rho v_P^2\left(1+2\epsilon\right)\left[\cos^2\theta\partial_x u_x-\sin\theta\cos\theta(\partial_x u_z+\partial_z u_x)+\sin^2\theta\partial_z u_z\right]\\
&& +\rho v_P^2\sqrt{1+2\delta}\left[\sin^2\theta\partial_x u_x+\sin\theta\cos\theta(\partial_x u_z+\partial_z u_x)+\cos^2\theta\partial_z u_z\right]\\
\partial_t \sigma^\theta_{zz} &=& \rho v_P^2\sqrt{1+2\delta}\left[\cos^2\theta\partial_x u_x-\sin\theta\cos\theta(\partial_x u_z+\partial_z u_x)+\sin^2\theta\partial_z u_z\right]\\
&& +\rho v_P^2\left[\sin^2\theta\partial_x u_x+\sin\theta\cos\theta(\partial_x u_z+\partial_z u_x)+\cos^2\theta\partial_z u_z\right].
\end{array}\right. \tag{2.31}$$

where $v_P(x)$ is the P-wave velocity, $\rho(x)$ is the density, and the anisotropy is described by the Thomsen parameter $\epsilon(x), \delta(x)$, and the tilt angle $\theta(x)$. When $\theta = 0$, the anisotropy is vertical: the local vertical velocity (along the z axis) is different from the horizontal velocity (along the x axis). The tilt angle represents the local rotation of the anisotropy orientation. For general "anelliptical" anisotropy ($\epsilon > \delta$), the PML are instable, while for the particular case of "elliptical anisotropy" ($\epsilon = \delta$) the PML are stable. The results presented here are extracted from (Métivier et al., 2014c,d,f) and focus on a finite-difference discretization of these 2D acoustic anisotropic equations.

2.5.1 Comparison with PML: stability

This experiment is based on the model presented in Figure 2.1, which is a part of a benchmark model issued by BP in 2007. This model contains a vertical P-wave velocity model as well as Thomsen anisotropy parameters ϵ and δ (Thomsen, 1986) and a tilt angle map θ.

The source is located close from the top of the model such that $x_1 = 7$ km, $x_2 = 0.05$ km. A reference solution is computed in a domain large enough to ensure no reflection at the boundaries. We then compute solutions using PML and SMART layers. We compare snapshots of the PML and SMART wavefields with the reference wavefields in Figure 2.2. As can be seen, the PML solution suffers from an exponentially growing mode which contaminates the whole domain of interest at the end of the computation. In comparison, the SMART layer solution remains stable (no growing modes appear). The SMART layer wavefield shows a satisfactory agreement with the reference solution. This experiment emphasizes the robustness of the SMART layer in a strongly heterogeneous and anisotropic model, compared to the PML approach.

2.5.2 Accuracy comparison between PML, SPONGE, and SMART layers

We compare the accuracy of the PML, SMART and SPONGE layer strategies in terms of reflectivity at the interface between the domain of interest and the layer. Here we consider a homogeneous medium with elliptical anisotropy so that the PML are stable. We compute a reference seismogram for a source located at $x_1 = 1$ km, $z_2 = 0.05$ km and an array of receivers located all along the surface at the same depth as the source. This reference solution is computed in a domain large enough to ensure no reflection at the boundaries during the recording time. It is presented in Figure 2.3, along with residual seismograms obtained from the difference between the reference seismogram and the seismograms obtained using the

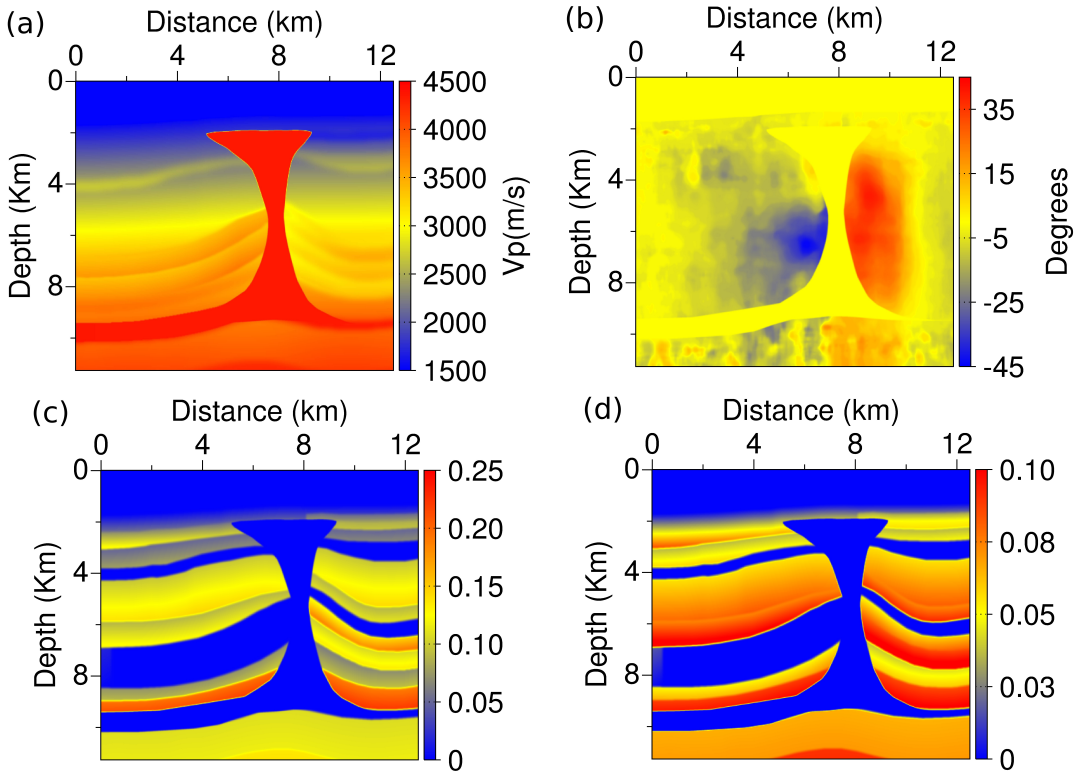

Figure 2.1 – (a) vertical P-wave velocity, (b) tilt angle, (c) Thomsen parameter ϵ, (d) Thomsen parameter δ.

different layer methods, with different thickness. We observe from Figure 2.3 the excellent accuracy of the PML solution in this case. The color-scale for the residual seismogram is two orders of magnitude lower than the color-scale used for the reference seismogram, and the residual energy in the PML seismogram is weak. Using the same layer thickness (15 discretization points), the SPONGE layer method yields significantly more energetic residuals. In comparison, the SMART layer appears to be less reflective than the SPONGE layer: the amplitude of the spurious reflections generated by the SMART layer is weaker. When the layer thickness is increased to 25 discretization points, the SPONGE layer still produces non negligible spurious reflections, while the SMART layer almost reaches the accuracy of the PML layer with 15 discretization points.

From this comparison, we see that the SMART layer is not as accurate as the PML method. This is expected, as the SMART method is not perfectly matched: the reflection coefficient at the interface between the domain of interest and the layer is zero only for waves arriving at normal incidence. On the other hand, the SMART layer method represents a significant improvement with respect to the SPONGE layer in terms of accuracy, as the amplitude of spurious reflection seems to be significantly reduced. We thus conclude that the SMART layer strategy is a better choice than PML or SPONGE layers when PML amplification problem arises.

2.6 Short conclusion

SMART layers represent an interesting alternative to PML strategies for robust simulation in elastic media when anisotropy has to be taken into account. Tests similar to the one presented here for acoustic transverse isotropy have been conducted on elastodynamics equations discretized with a finite-element discontinuous Galerkin strategy, leading to the same conclusions (Tago et al., 2014). One remaining open question for this strategy is related to the fact that the method is designed for first-order hyperbolic systems. Current FWI applications in the elastic approximation are usually formulated using second-

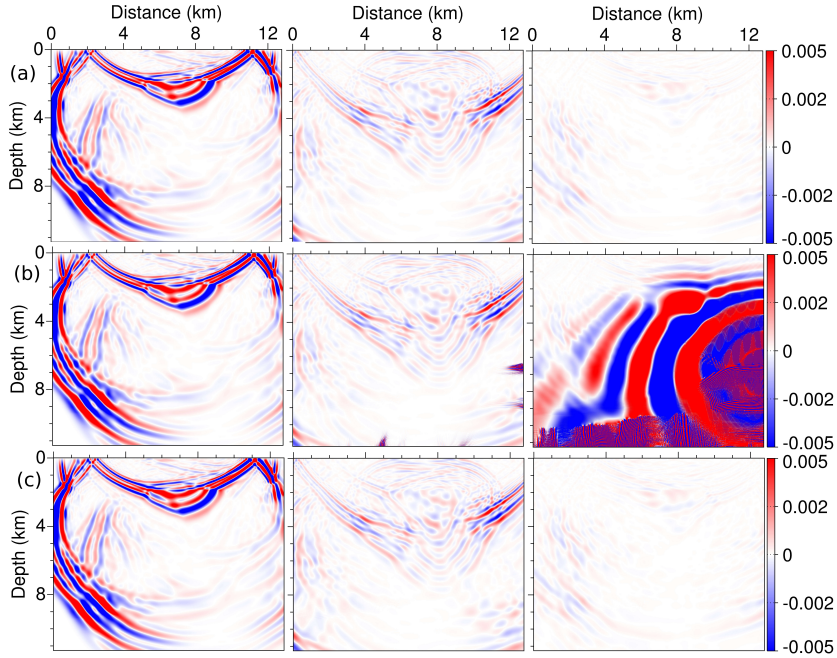

Figure 2.2 – Pressure wavefield snapshots in the domain of interest at time $t = 4.2$ s, 7.8 s and 12 s (from left to right). Reference wavefield (row a). PML wavefield (row b), SMART wavefield (row c). The PML amplification starts at time $t = 7.8$ s on the right and bottom edges of the domain and contaminates almost all the domain at final time $t = 12$ s

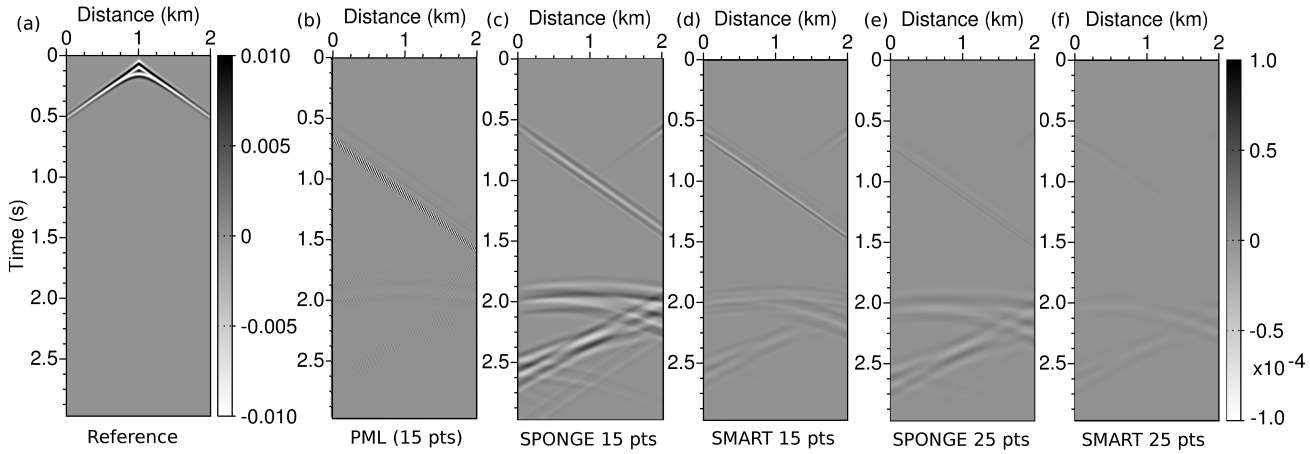

Figure 2.3 – Reference seismogram (a). Residuals seismogram with PML 15 grid points width (b), SPONGE with 15 grid points, SMART with 15 grid points (c), SPONGE with 25 grid points (d), SMART with 25 grid points (e). The scale of the residual is 2 orders of magnitude lower than for the reference wavefields.

order systems based on velocity displacement only. This allows a simple reduction from a system of 9 equations to a system of 3 equations, therefore decreasing the computation cost, either in the time or in the frequency-domain. Up to now, there is no straightforward extension of the SMART layers approach to second-order systems, as the projection operators are based on the first-order velocity-stress system.

Deriving a proper extension to second-order equations thus requires further work to be performed.

3 Truncated Newton optimization and multiparameter imaging

This section summarizes the work presented in the following publications and expanded abstracts.
Publications:

- Métivier, L., Brossier, R., Operto, S., & J., V., 2017. Full waveform inversion and the truncated Newton method, *SIAM Review*, **59**(1), 153–195
- Métivier, L. & Brossier, R., 2016a. The seiscope optimization toolbox: A large-scale nonlinear optimization library based on reverse communication, *Geophysics*, **81**(2), F11–F25
- Métivier, L., Brossier, R., & Virieux, J., 2015c. Combining asymptotic linearized inversion and full waveform inversion, *Geophysical Journal International*, **201**(3), 1682–1703
- Métivier, L., Bretaudeau, F., Brossier, R., Operto, S., & Virieux, J., 2014a. Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures, *Geophysical Prospecting*, **62**, 1353–1375
- Operto, S., Brossier, R., Gholami, Y., Métivier, L., Prieux, V., Ribodetti, A., & Virieux, J., 2013. A guided tour of multiparameter full waveform inversion for multicomponent data: from theory to practice, *The Leading Edge*, **Special section Full Waveform Inversion**(September), 1040–1054
- Métivier, L., Brossier, R., Virieux, J., & Operto, S., 2013. Full Waveform Inversion and the truncated Newton method, *SIAM Journal On Scientific Computing*, **35(2)**, B401–B437

Abstracts:

- Métivier, L., Brossier, R., Operto, S., & Virieux, J., 2015a. Acoustic multi-parameter FWI for the reconstruction of P-wave velocity, density and attenuation: preconditioned truncated newton approach, in *Expanded Abstracts, 85th Annual Meeting*, pp. 1198–1203, SEG
- Métivier, L., Brossier, R., Operto, S., & Virieux, J., 2014e. Multi-parameter FWI - an illustration of the Hessian operator role for mitigating trade-offs between parameter classes, in *Expanded Abstracts, 6^{th} EAGE St-Petersbourg International Conference & Exhibition*
- Métivier, L., Brossier, R., Virieux, J., & Operto, S., 2012a. Optimization schemes in FWI: the truncated Newton method, in *2012 SEG Abstracts*
- Métivier, L., Brossier, R., Virieux, J., & Operto, S., 2012b. Toward Gauss-Newton and exact Newton optimization for full waveform inversion, in *EAGE, 74th Conference and Exhibition*

3.1 Context and problematic

3.1.1 Notations

For the sake of clarity, we introduce the following notations:

- the domain of interest in which we want to reconstruct the subsurface mechanical properties is denoted by $\Omega \subset \mathbb{R}^d$;
- the sources and receivers positions are denoted by x_s and x_r, they are defined on the boundary of Ω
$$x_s \in \Gamma_s \subset \partial\Omega, \quad x_r \in \Gamma_r \subset \partial\Omega; \tag{3.1}$$
- the collection of the subsurface mechanical properties we want to reconstruct are called the model parameters, they are gathered in a vector $m(x) = [m_1(x), \dots m_N(x)]$ for $x \in \Omega$ where N is the number of different parameter classes;
- the observed seismic data associated with a source located at position x_s is a function $d_{obs,s}(x_r, t)$ with t in the time interval $[0, T]$, T being the recording time, and $x_r \in \Gamma_r$;
- the corresponding calculated seismic data associated with a source located at position x_s and a model parameter $m(x)$ is denoted by $d_{cal,s}[m](x_r, t)$.

3.1.2 Position of the problem

In its conventional formulation, FWI corresponds to the following nonlinear least-squares minimization problem

$$\min_m f(m) = \frac{1}{2}\sum_s \sum_r \int_t |d_{cal,s}[m](x_r,t) - d_{obs,s}(x_r,t)|^2 dt. \tag{3.2}$$

The computation of the synthetic data $d_{cal,s}[m](x_r,t)$ is performed through the computation of the wavefield

$$u_s[m](x,t), \quad (x,t) \in \Omega \times [0,T], \tag{3.3}$$

as the solution of a set of partial differential equations representing the wave propagation in the medium of interest Ω,

$$A(m)u = s(x,t), \quad (x,t) \in \Omega \times [0,T], \quad u(x,0) = 0, \tag{3.4}$$

where $s(x,t)$ represents the source term. We introduce $s(x,t)$ as

$$s(x,t) = \delta(x - x_s)\varphi(t), \tag{3.5}$$

where δ is the Dirac delta function and $\varphi(t)$ is the temporal signature of the source. The medium is considered to be at rest at the initial time, therefore the zero initial condition in equation (3.4).

In (3.4), the operator $A(m)$ is a compact notation for any wave propagation operator, from the simple acoustic case to the general visco-elastodynamics case. In the acoustic approximation, we would have for instance

$$m = [c \;\; \rho]^T, \quad A(m) = \partial_{tt} + \rho c^2 \mathrm{div}\frac{1}{\rho}\nabla. \tag{3.6}$$

Given a wavefield $u_s[m](x,t)$, the synthetic data $d_{cal}[m](x_r,t)$ is computed through a restriction operator R, extracting the value of the wavefield at the receiver locations x_r

$$d_{cal,s}[m](x_r,t) = Ru[m] = \int_\Omega u[m](x,t)\delta(x - x_r)\, dx. \tag{3.7}$$

Note that the nonlinearity of the misfit function with respect to $m(x)$ is related to the nonlinear relation between the wavefield $u_s[m](x,t)$ and the model parameter $m(x)$, even though the wave propagation PDE is linear in u.

In the general case, without strong prior information on the subsurface structure (*i.e.* layered medium) a fine discretization of $m(x)$ (down to a fraction of the shortest propagated wavelength) is used to represent the subsurface mechanical properties. This leads to large scale nonlinear minimization problems, for which local optimization strategies have to be used. These schemes are based on the iteration

$$m^{k+1} = m^k + \alpha^k \Delta m^k, \tag{3.8}$$

starting from an initial guess m^0, where $\alpha^k \in \mathbb{R}_*^+$ is a steplength (scalar parameter) computed through linesearch or trust-region strategies (Nocedal & Wright, 2006; Bonnans et al., 2006), and Δm^k is the model update at iteration k also called descent direction.

Local optimization schemes depend on the definition of the descent direction Δm^k. First order strategies are based on the opposite of the gradient (steepest descent, nonlinear conjugate gradient). Second-order schemes rely on the solution of the Newton equation

$$H(m^k)\Delta m^k = -\nabla f(m^k), \tag{3.9}$$

where $H(m^k)$ is the Hessian operator, the matrix of second-order derivatives of the misfit function $f(m)$. The most popular one is the l-BFGS strategy (Nocedal, 1980) which approximates the inverse Hessian operator $H(m^k)^{-1}$ through finite differences of the l previous gradient and model updates $\nabla f(m_k), \ldots, \nabla f(m^{k-l+1})$, $m^k, \ldots, m^{k-l+1}$.

3.1.3 Physical interpretation of the gradient and Hessian operators

Geophysicists have inferred insightful physical interpretations of the gradient and Hessian operators (Pratt et al., 1998), which appear from the previous discussion as the key components of the FWI process. It is worth recalling the main results here.

We consider the general multi-parameter case. We assume the subsurface model m is discretized in space. Each discrete model parameter can thus be indexed by ij where i refers to the *spatial location* and j to the *parameter class* (velocity, density, attenuation for instance).

In this framework, the ijth component of the gradient of the misfit function (3.2) is given by

$$\nabla f(m)_{ij} = \sum_s \sum_r \int_0^T \frac{\partial d_{cal,s}[m](x_r,t)}{\partial m_{ij}} \left(d_{cal,s}[m](x_r,t) - d_{obs,s}(x_r,t)\right) dt. \tag{3.10}$$

The partial derivatives of the synthetic data $\frac{\partial d_{cal,s}[m](x_r,t)}{\partial m_{ij}}$ represent the wavefield scattered by a small perturbation of the parameter m_{ij} recorded at the position x_r, all the other parameters being kept constant (Fig. 3.1). This is obtained mathematically by reminding that from the restriction (3.7) we have

$$\frac{\partial d_{cal,s}[m](x_r,t)}{\partial m_{ij}} = R\frac{\partial u_s[m]}{\partial m_{ij}} \tag{3.11}$$

and by deriving the forward problem (3.4) with respect to the model parameter m_{ij}

$$A(m)\frac{\partial u_s[m]}{\partial m_{ij}} = -\frac{\partial A}{\partial m_{ij}} u_s[m]. \tag{3.12}$$

From (3.12) we see that the wavefield $\frac{\partial u_s[m]}{\partial m_{ij}}$ is the solution of the same wave equation as the incident wavefield $u_s[m](x_r,t)$ only with a different source term $\frac{\partial A}{\partial m_{ij}} u_s[m]$ which represents the perturbation of the parameter m_{ij}.

The component ij of the gradient is therefore given by the cross-correlation between the data residuals $(d_{cal,s}[m](x_r,t) - d_{obs,s}(x_r,t))$ and this scattered response recorded at the receiver level. As such, building the gradient can be interpreted as systematically probing the scattering response of the medium to perturbation of each model parameters m_{ij} and measuring the correlation of this response with the unexplained part of the seismic data. The gradient introduces missing heterogeneities in the medium to reduce the data residuals relying on a single scattering approximation of the wave propagation.

The way the wavefield is scattered by the local heterogeneities of the medium actually depends on the type of parameter heterogeneity and the choice of the parameterization to represent the medium. This is what is called the radiation (or scattering) pattern (Fig. 3.2), which is associated with the operator $\frac{\partial A}{\partial m_{ij}}$. The amplitude of the scattered signal depends on the parameterization and the illumination angle θ formed by the ray connecting the source to the diffraction point, and the ray connecting the receiver to the diffraction point (Fig. 3.1 d). The differences between the gradients associated with different parameter classes result only from their different radiation patterns (equation (3.10)). Therefore, if two different parameter classes have radiation patterns which significantly overlap over a significant range of θ, deciphering between these parameter classes using only the gradient is difficult: cross-talks (or trade-offs) impact their reconstruction. In other words, perturbations of different parameter classes might explain similarly the seismic data. This already highlights the complexity of multi-parameter FWI.

If now we analyze the Hessian operator, we have, from (3.2), the general formulation

$$H(m) = B(m) + C(m), \tag{3.13}$$

where

$$B(m)_{ij,kl} = \sum_s \sum_r \int_0^T \frac{\partial d_{cal,s}[m](x_r,t)}{\partial m_{ij}} \frac{\partial d_{cal,s}[m](x_r,t)}{\partial m_{kl}} dt, \tag{3.14}$$

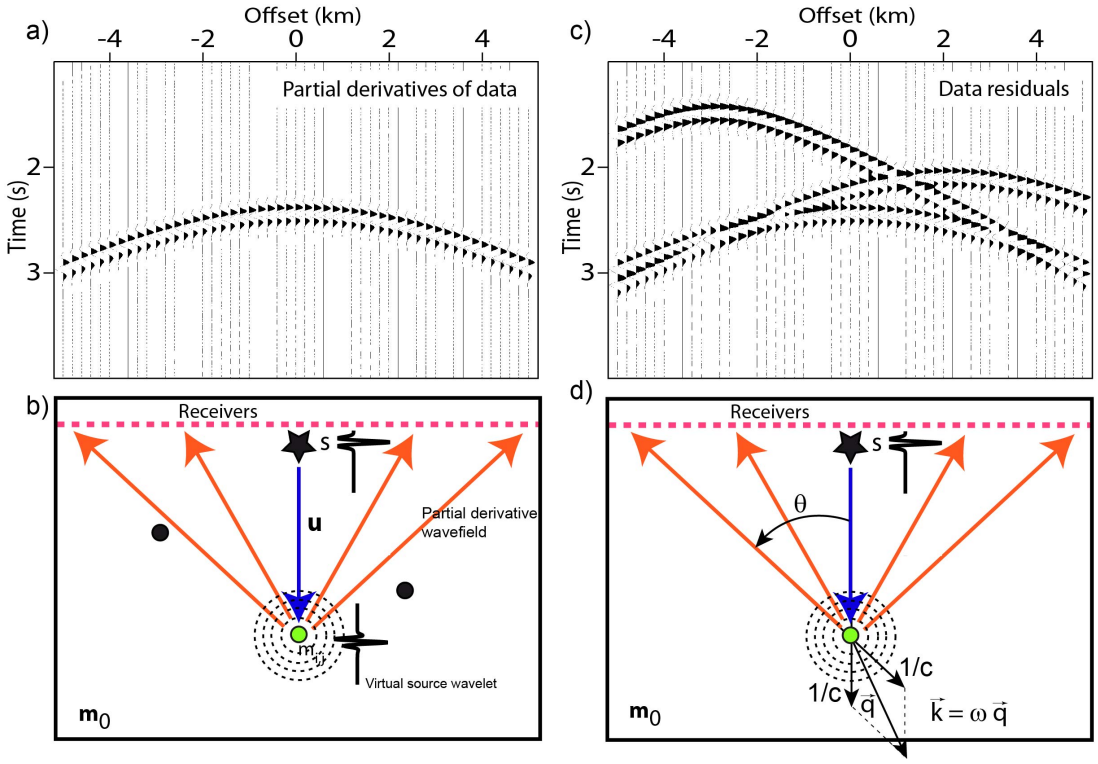

Figure 3.1 – Physical interpretation of the gradient. The true medium contains three diffracting points (filled circles in (b)) and the background model is homogeneous. The data residuals (difference between observed and synthetic data) recorded at receivers (b, thick dashed line,) correspond to the wavefield scattered by the three diffracting points (c). The partial derivative of the wavefield with respect to the model parameter located at the position of the middle diffracting point (b, green circle) corresponds to the wavefield scattered by this diffracting point normalized by the value of the wave velocity perturbation. The radiation pattern of the virtual source located at the diffracting point is illustrated by the thin dashed lines in (b) (here, an isotropic radiation pattern). The zero-lag correlation between the data residuals (c) and the partial derivative wavefield taken at the receiver positions (a) provides the unscaled contribution of the source s at the position m_{ij}. Figure taken from (Operto et al., 2013).

is the Gauss-Newton approximation of the Hessian operator, and $C(m)$ is defined by

$$C(m)_{ij,kl} = \sum_s \sum_r \int_0^T \frac{\partial^2 d_{cal,s}[m](x_r,t)}{\partial m_{ij} \partial m_{kl}} \left(d_{cal,s}[m](x_r,t) - d_{obs,s}(x_r,t) \right) dt. \tag{3.15}$$

Assuming the model parameters are organized by parameter classes, the matrices $B(m)$ and $C(m)$ have a block structure, as illustrated in Figure 3.3 for a three parameter classes problem in the acoustic approximation. The parameters which are considered in this example are the P-wave velocity, the density, and an attenuation factor. The upper left diagonal block is associated with the P-wave velocity, the central diagonal block is associated with the density, the bottom right diagonal block is associated with the attenuation parameter. The off-diagonal blocks are associated with cross-talks between these three parameters.

From (3.14), we see that $B_{ij,kl}$ is the correlation of the wavefield scattered by perturbations of the couple of parameters m_{ij}, m_{kl}. The matrix $B(m)$ thus measures the similarity of this scattering response. As such, it can be interpreted as a measure of the coupling between discrete parameters when relying only on the gradient interpretation.

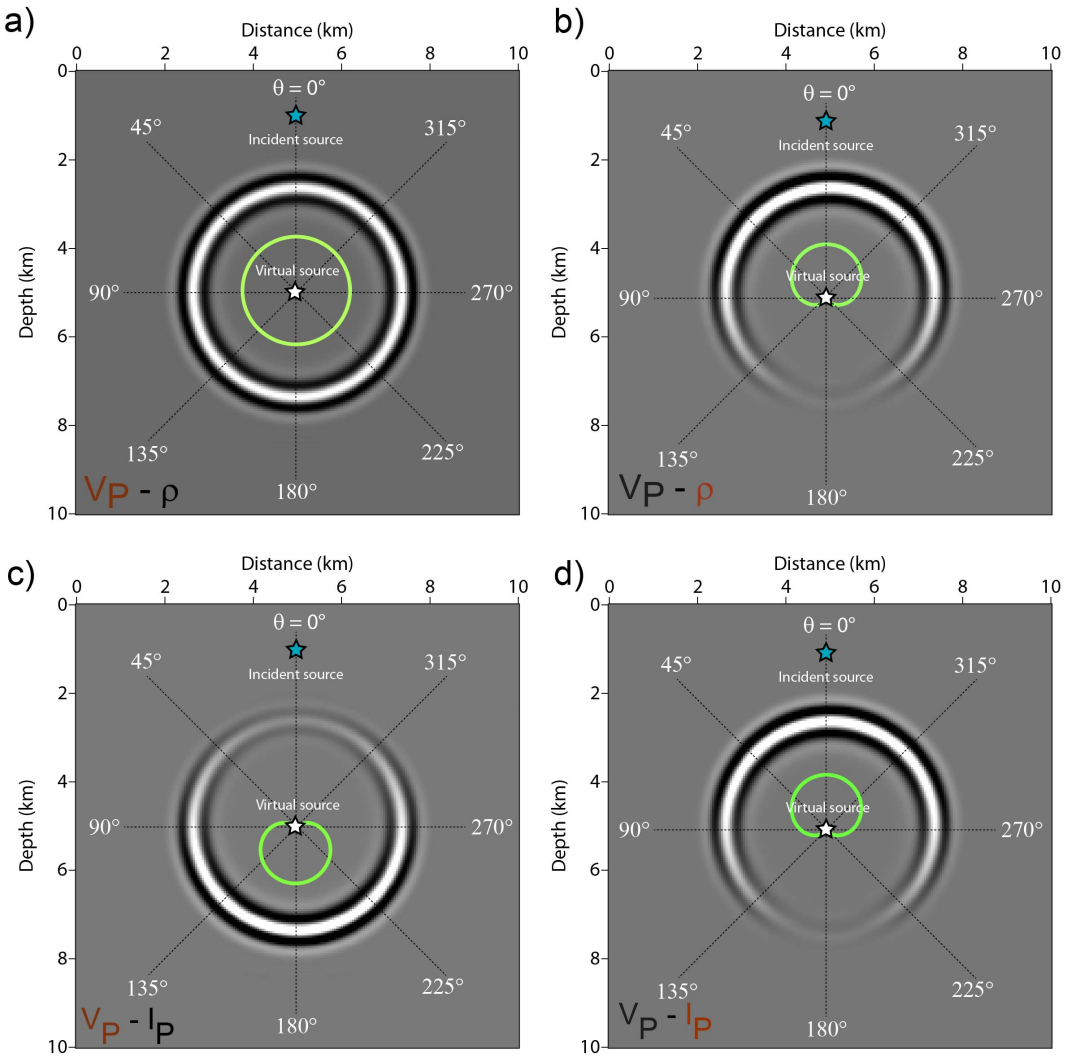

Figure 3.2 – Radiation patterns of the virtual sources of the partial derivative wavefields for an acoustic medium parameterized by (a-b) (velocity v_P, density ρ) and (c-d) (velocity v_P , acoustic impedance I_P). The figure shows a snapshot of a wavefield that propagates in a homogeneous background model and which is scattered by a point perturbation located in the middle of the grid. Only one model parameter of the subsurface parameterization is perturbed, the other one being kept fixed. (a) v_P perturbation, with ρ fixed. (b) ρ perturbation, with v_P fixed. (c) v_P perturbation, with I_P fixed. (d) I_P perturbation, with v_P fixed. The scattering angle is labeled by θ. The virtual source at the diffracting point is denoted by the white-filled star. The incident source, at the vertex of the model perturbation, is denoted by the blue-filled star. The amplitude variations of the wavefront around the diffracting point result from the radiation pattern of the virtual sources. The analytical radiation patterns are superimposed (green curves). Figure taken from (Operto et al., 2013).

For diagonal blocks such that $j = l$, this coupling is measured between parameters of the same class j, at different spatial positions i and k. The entry $B_{ij,kj}$ decreases with the spatial distance which separates m_{ij} and m_{kj} and reaches its maximum for the autocorrelation of the two derivatives of the wavefields ($i = k$). In the high-frequency approximation, the zero-lag correlation of the derivatives of the wavefield with respect to m_{ij} and m_{kj} would be zero and the blocks of $B(m)$ would be diagonal.

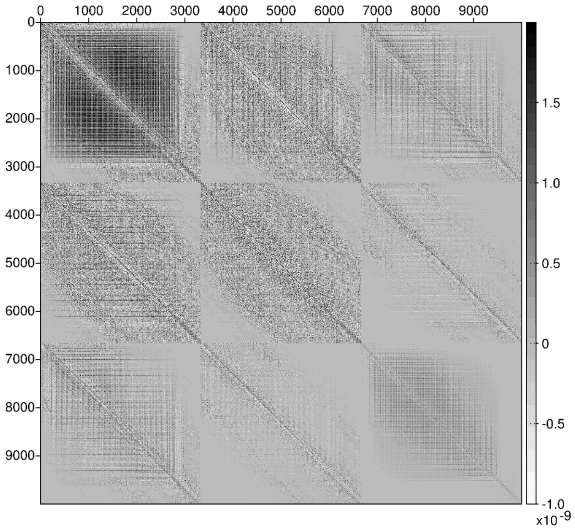

Figure 3.3 – Hessian operator $H(m_0)$ corresponding to a 2D frequency-domain multi-parameter FWI problem for the reconstruction of the P-wave velocity, the density, and an attenuation parameter.

The limited frequency content of the seismic data translates in the banded aspect of these blocks.

For non-diagonal blocks such that $j \neq l$, the coupling is measured between parameters of two different classes j and l. The diagonal entries of the non-diagonal blocks $B_{ij,il}$ measure the coupling between parameters of class j and l at the same spatial position i. In the most general settings ($i \neq k$ and $j \neq l$), $B_{ij,kl}$ measures the coupling between parameters of two different classes j and l, at two different spatial positions i and k.

Similarly, the expression (3.15) shows that $C_{ij,kl}$ is the correlation of the residuals with the second-order derivatives of the wavefield recorded at the receivers, with respect to the parameters m_{ij} and m_{kl}. This second-order wavefield is also called a double-scattered wavefield: it represents a recorded signal that has been scattered twice, by the parameter m_{ij} and m_{kl}.

3.2 Proposed approach: truncated Newton strategy

From this analysis, we see that relying only on the gradient to perform FWI present two difficulties.

- When considering multi-parameter reconstructions, trade-offs/coupling effects between parameters might contaminate the reconstruction, due to an overlap of their radiation patterns over a significant range of diffraction angles θ.
- When strong multiple reflections are recorded, the gradient might interpret secondary and higher order reflections as single scattered events and consequently misposition heterogeneities in the medium (reflectors). This issue might be mitigated by iterating the descent algorithm, however the convergence might be extremely slow which would make the method improper for practical applications.

Filtering the gradient $\nabla f(m)$ with the matrix $H(m)^{-1}$ to compute the model update following the Newton equation (3.9) acts simultaneously as

- a spatial refocusing filter for each parameter class;
- a decoupling filter to mitigate trade-offs between parameter classes.

In addition, the second-order part of the Hessian operator $C(m)$ allows one to compensate for the artifacts generated by double-scattered waves.

This analysis underlines the crucial importance of the Hessian operator in the solution of the FWI problem (3.2). This has been the main motivation to study a specific local optimization scheme named truncated Newton method (also known as inexact Newton, Newton-CG, or Newton-Krylov method) (Nash, 2000). Contrary to quasi-Newton algorithms, this method is not based upon an approximation of the inverse Hessian operator. Instead, the descent direction Δm^k is computed as an approximate solution of (3.9) using the conjugate gradient algorithm (Saad, 2003). In this framework, the approximation of the inverse Hessian operator at a given iteration k is local: it only involves first-and second-order derivatives of the misfit function computed at the model m^k. It is an advantage over the l-BFGS algorithm, which builds an approximation of the inverse Hessian operator which is non-local, as it is based on the gradient values computed at the l previous iterations $k, k-1, \ldots, k-l+1$. This non locality may be prejudicial to an accurate approximation of the inverse Hessian operator for strongly nonlinear misfit functions, and only a rough approximation is possible in the first iterations, as it is only based on the first gradients.

Nonetheless, the implementation of truncated Newton strategies for large scale FWI problems is not trivial, mainly due to the potentially high computational cost of the method as it requires direct approximation of the second-order derivatives of the misfit function. We thus have proposed a series of contributions to design a feasible implementation of this strategy for realistic scale FWI problems.

The guideline is to work in a matrix-free format: the storage or the computation of the Hessian operator would be prohibitively expensive. The Newton equation (3.9) is solved with a conjugate gradient iterative solver which is naturally matrix-free. The difficulty thus relies on an efficient computation of Hessian-vector products $H(m)v$. We have proposed a second-order adjoint technique to this purpose (Métivier et al., 2012b). An adapted stopping criterion is also mandatory, to adapt the level of accuracy to which is solved the Newton equation (3.9) to guarantee the convergence of the nonlinear optimization. We have proposed to use the criterion defined by Eisenstat & Walker (1994) (Métivier et al., 2013; Métivier et al., 2017). Finally, we have developed specific preconditioning techniques to accelerate the convergence of the conjugate gradient solver (Métivier et al., 2014a, 2015a), adapted to the multi-parameter FWI case.

We present these different contributions in more details in the following section, before providing some illustrations of the interest of this method.

3.3 Mathematical description

3.3.1 Hessian-vector products computation through second-order adjoint techniques

Adjoint state strategy for computing the gradient. The standard approach to compute the misfit gradient $\nabla f(m)$ relies on the adjoint-state technique, which has been introduced in optimal control theory by (Lions, 1968), and later on adapted to parameter identification problems by Chavent (1971) and to weather forecasting by Le Dimet & Talagrand (1986). A survey of its use in seismic imaging has been proposed by Plessix (2006). The purpose of this strategy is to avoid the computation of the partial derivative wavefield $\frac{\partial d_{cal,s}[m](x_r,t)}{\partial m_{ij}}$ for each model parameter m_{ij}, which would not be affordable for realistic size FWI problems.

Here we present the general principle of the strategy. For the sake of clarity, we assume that the number of seismic sources is equal to 1 and we drop the index s in the sequel (the generalization is straightforward by summation).

The Lagrangian function associated with the problem (3.2) is

$$L(m,u,\lambda) = \frac{1}{2}\sum_r \int_0^T |(Ru)(x_r,t) - d_{obs}(x_r,t)|^2 \, dt + (A(m)u - s, \lambda)_{\mathcal{W}}\,, \tag{3.16}$$

where $(.,.)_{\mathcal{W}}$ is the L^2 scalar product in the wavefield space

$$(u,v)_{\mathcal{W}} = \int_0^T \int_\Omega u(x,t)v(x,t)\, dx\, dt, \tag{3.17}$$

and λ is the adjoint wavefield. Denoting $u[m]$ the solution of (3.4), we have

$$L(m,u[m],\lambda) = f(m). \tag{3.18}$$

Deriving (3.18) with respect to m yields

$$\frac{\partial L(m,u[m],\lambda)}{\partial m} = \nabla f(m), \tag{3.19}$$

which is equivalent to

$$\frac{\partial}{\partial m}\left(A(m)u[m],\lambda\right)_{\mathcal{W}} + \frac{\partial L(m,u[m],\lambda)}{\partial u}\frac{\partial u[m]}{\partial m} = \nabla f(m). \tag{3.20}$$

The adjoint state $\lambda[m]$ is defined as the solution of

$$\frac{\partial L(m,u[m],\lambda)}{\partial u} = 0, \tag{3.21}$$

which is equivalent to

$$A(m)^T\lambda = R^T\left(d_{obs}(x_r,t) - Ru[m](x_r,t)\right) = \sum_r \left(d_{obs}(x_r,t) - Ru[m](x_r,t)\right)\delta(x-x_r), \tag{3.22}$$

where $A(m)^T$ stands for the adjoint wave propagation operator. In the time-domain, this operator is equivalent to a backpropagation operator. Indeed, through time-integration by parts, the initial homogeneous condition of the incident wavefield turns to an homogeneous final condition for the adjoint wavefield. In the frequency domain, this backpropagation turns into a complex conjugation. The formula (3.22) is important as it reveals that the adjoint is computed as the backpropagation of the residuals from the receivers location.

We obtain the gradient formula

$$\nabla f(m) = \int_0^T \left(\frac{\partial A(m)}{\partial m}\overline{u}[m](x,t)\right)\overline{\lambda}[m](x,t)\,dt. \tag{3.23}$$

Using the adjoint formalism, the gradient is thus computed as the correlation between the incident wavefield $u[m]$ and the adjoint wavefield $\lambda[m]$. From a computational point of view, the computation of the first-order scattered wavefields $\frac{\partial d_{cal,s}[m](x_r,t)}{\partial m_{ij}}$ is replaced by the solution of one incident and one adjoint propagation problems, which is much more efficient.

The correlation of these two wavefields to assemble the gradient through the formula (3.23) requires specific attention in the time-domain as the incident and adjoint wavefields are not in principle accessed at the same time step following standard time marching schemes. Different options exist to overcome this difficulty, either relying on storing the incident wavefield during its computation, or recomputing it from the final step after a first propagation stage. The choice depends of course of the problem size and the computational architecture available. Combination of these two strategies might be employed: see for instance checkpointing strategies (Griewank & Walther, 2000), or more recently the CARFS method we have designed within SEISCOPE (Yang et al., 2016b,a)).

Extension to the computation of Hessian-vector products. To compute Hessian vector products $H(m)v$ for a given vector v in the model space, the previous strategy is extended to second-order. This extension simply relies on the definition of a function $g_v(m)$ such that

$$g_v(m) = \left(\nabla f(m), v\right)_{\mathcal{M}}, \tag{3.24}$$

where this time $(.,.)_{\mathcal{M}}$ denotes the L^2 scalar product in the model space $\mathcal{M}$

$$(m,p)_{\mathcal{M}} = \sum_{j=1}^{N}\int_\Omega m_j(x)p_j(x). \tag{3.25}$$

The gradient of the function $g_v(m)$ is, by definition

$$\nabla g_v(m) = H(m)v. \tag{3.26}$$

The same adjoint-state formalism can thus be used to compute the gradient of the function $g_v(m)$. We do not give the details of this derivation here: they can be found for instance in (Métivier et al., 2017). As a result, the Hessian-vector products can be expressed as the sum of three correlation terms (similar to expression (3.23))

$$\begin{aligned}(H(m)v)(x) = \quad & \int_0^T \sum_j \int_\Omega \left(\frac{\partial^2 A}{\partial m_j \partial m} u[m](x,t) v_j(x) \lambda[m](x,t) dt \right) + \\ & \int_0^T \frac{\partial A}{\partial m} \lambda[m](x,t) \mu_1[m](x,t)\ dt + \\ & \int_0^T \frac{\partial A}{\partial m} u[m](x,t) \mu_2[m](x,t)\ dt,\end{aligned} \tag{3.27}$$

involving the computation of secondary incident and adjoint wavefields $\mu_1[m]$ and $\mu_2[m]$ solutions of

$$\left\{ \begin{array}{l} A(m)\mu_1(x,t) = -\sum_j \left(\frac{\partial A}{\partial m_j} v_j(x) u[m] \right)(x,t), \\ A(m)^T \mu_2 = -R^T R \mu_1(x,t) - \sum_j \left(\frac{\partial A^T}{\partial m_j} v_j(x) \lambda[m] \right)(x,t). \end{array} \right. \tag{3.28}$$

Following this strategy, the Hessian-vector product can be expressed only in terms of four wavefields: the primary incident and adjoint wavefields $u[m]$ and $\lambda[m]$, together with the secondary incident and adjoint wavefields $\mu_1[m]$ and $\mu_2[m]$. This thus opens the way to efficient implementation, the formalism being matrix free. Additional savings can be obtained in the Gauss-Newton approximation for which only three wavefields need to be computed (the term involving $\lambda[m]$ vanishes in this case). Depending on the problem size and available resources, the primary fields might also not be recomputed and kept in memory, yielding substantial savings.

This mathematical development has been the starting point for the implementation of the truncated Newton strategy for FWI.

3.3.2 Dynamic stopping criterion for the Newton equation

The second key point for an efficient implementation of the truncated Newton strategy for large scale nonlinear optimization relies on the definition of the stopping criterion to solve the inner linear system (3.9). In the Newton optimization framework, the computation of the descent direction Δm_k amounts to minimizing the local quadratic approximation of the misfit function $f(m)$ defined by

$$q_k(\Delta m_k) = f(m_k) + (\nabla f(m_k), \Delta m_k) + \frac{1}{2}(H(m_k)\Delta m_k, \Delta m_k). \tag{3.29}$$

If this approximation is accurate, there is interest in computing a precise solution to the problem (3.9). Conversely, when this quadratic approximation is inexact, computing an exact solution of the system (3.9) might generate ascent directions. The accuracy with which the linear system (3.9) should be solved thus depends on the accuracy of this quadratic approximation.

This issue has been clearly pointed out by Eisenstat & Walker (1994) who provide three stopping criteria related to the accuracy of the local quadratic approximation. For each of these criteria, a convergence proof of the truncated Newton method is provided. These criteria take the following form: the linear iterations are stopped whenever

$$\|H(m_k)\Delta m_k + \nabla f(m_k)\| \leq \eta_k \|\nabla f(m_k)\|. \tag{3.30}$$

where the quantity η_k is defined as the forcing term. The value of η_k controls the required accuracy on the solution of the system (3.9).

Among the three possible definitions of η_k, we have illustrated in a numerical study (Métivier et al., 2013) that the following choice appears as the most appropriate for FWI applications

$$\eta_k = \frac{\|\nabla f(m_k) - \nabla f(m_{k-1}) - \alpha_{k-1} H(m_{k-1}) \Delta m_{k-1}\|}{\|\nabla f(m_{k-1})\|}, \tag{3.31}$$

The measure of the accuracy of the quadratic approximation of the misfit function is obtained through a comparison between the gradient and its first-order Taylor expansion

$$\nabla f(m_k) \simeq \nabla f(m_{k-1}) + \alpha_{k-1} H(m_{k-1}) \Delta m_{k-1} + o(\|\Delta m_{k-1})\|. \tag{3.32}$$

The choice of the forcing term and the definition of the associated stopping criterion for the linear iteration is complemented with an appropriate strategy to deal with the detection of negative curvatures. Indeed, while the Gauss–Newton approximation of the Hessian $B(m)$ is symmetric positive, the second-order part $C(m)$ may render the full Hessian operator $H(m)$ symmetric indefinite. Therefore, during the solution of the linear system (3.9) with the conjugate gradient algorithm, the probability of encountering a curvature associated with a negative eigenvalue of the operator $H(m)$ cannot be ignored. In this case, the linear iterations are stopped and the last value of the descent direction Δm_k which is computed is returned. If this negative curvature is met at the very first linear iteration, the steepest-descent direction is returned. This safeguard provides a very robust stopping criterion: the model update computed in the inner conjugate gradient loop is always guaranteed to be a descent direction for the misfit function $f(m)$.

3.3.3 Preconditioning techniques

Improving the convergence rate when solving the Newton equation (3.9) with CG iterations is crucial for an efficient implementation of the truncated Newton strategy as for realistic applications, controlling the computational time might require to perform only few CG iterations (up to 10 for instance, for linear systems involving millions of unknowns or more). This requires the design of a suitable preconditioner for the linear system (3.9).

The first possibility we have explored consists in coupling the truncated Newton and l-BFGS strategies, where the l-BFGS approximation of the inverse Hessian operator is used as a preconditioner (Métivier et al., 2013). However, this approach suffers from the same limitation than the l-BFGS strategy itself: the inverse Hessian approximation is not accurate at the first iterations of the method, therefore the quality of the preconditioner might not be sufficient in the first iteration.

The second strategy we have explored is related to what is referred to as the pseudo-Hessian operator in the FWI community (Shin et al., 2001; Choi & Shin, 2008). This is an approximation $\widetilde{B}$ of the Gauss-Newton part of the Hessian such that

$$\widetilde{B}(m)_{ij,kl} = \sum_s \int_\Omega \int_0^T u_s[m](x,t) \frac{\partial A}{\partial m_{ij}} u_s[m](x,t) u_s[m](x,t) \frac{\partial A}{\partial m_{kl}} u_s[m](x,t) dx dt. \tag{3.33}$$

This rather crude approximation is based on an approximation of the Green function at a receiver position by the incident wavefield $u_s[m]$ itself, which is a quantity already computed when the gradient is assembled.

In the mono-parameter case, a crude approximation of H^{-1} is deduced from (3.33) by selecting only the diagonal components of this operator, such that the preconditioner P is

$$P = \left(\mathrm{diag} \widetilde{B}_{ii} \right)^{-1}. \tag{3.34}$$

We have generalized this approximation to the multi-parameter settings in Métivier et al. (2015a), following a block diagonal approximation initially proposed in Innanen (2014) and (Korta et al., 2013). The principle is that each block of the Hessian operator is approximated by its diagonal, which leads to the preconditioner

$$P = \begin{bmatrix} \mathrm{diag}\widetilde{B}_{ii,11} & \mathrm{diag}\widetilde{B}_{ii,12} & \dots & \mathrm{diag}\widetilde{B}_{ii,1N} \\ \mathrm{diag}\widetilde{B}_{ii,21} & \mathrm{diag}\widetilde{B}_{ii,22} & \dots & \mathrm{diag}\widetilde{B}_{ii,2N} \\ \vdots & \ddots & & \\ \mathrm{diag}\widetilde{B}_{ii,NN} & \mathrm{diag}\widetilde{B}_{ii,NN} & \dots & \mathrm{diag}\widetilde{B}_{ii,NN} \end{bmatrix}^{-1}. \tag{3.35}$$

These off-diagonal blocks contributions approximate the trade-off between parameter classes, neglecting the spatial interactions between parameters. The structure of P implies that its application only requires

the solution of a $N \times N$ linear system at each grid points which is negligible from a computational point of view, as in practice the number of parameter classes simultaneously inverted will not exceed the order of 10 (the maximum number of independent parameter in the stiffness matrix is equal to 21 to which could be added parameters associated with attenuation mechanisms).

3.4 Numerical studies

We now provide several illustrations of the interest of the truncated Newton strategy in the framework of FWI. These examples are obtained on synthetic cases, in the 2D acoustic approximation, using a frequency-domain FWI algorithm. We first focus on the mono-parameter case for the reconstruction of the P-wave velocity only, to emphasize the interest for accounting for second-order scattering when strong amplitude multiple reflections are recorded. Two illustrations are provided. We then present a three parameters inversion for the simultaneous reconstruction of the P-wave velocity, the density, and an adimensional factor accounting for the subsurface attenuation. We emphasize the interest of the preconditioned truncated Newton strategy in this case to better decouple these parameters.

3.4.1 Mono-parameter velocity reconstruction with large amplitude multiple reflections

Civil engineering application. We first consider a schematic synthetic case study, inspired from a civil engineering application. The model to reconstruct is composed of a homogeneous background at 300 m.s^{-1} and two superimposed concrete structures at 4000 m.s^{-1} (Fig. 3.4). The initial model misses the two concrete structures. Detecting and correctly imaging these structures is challenging due to the strong multiple reflections generated by the interaction between the two structures. A full acquisition system with four line of sources and receivers (one line on each side) is used. The inversion is performed in the frequency-domain: we invert 9 data-sets from 100 to 300 Hz at 25 Hz interval. No preconditioning is used for this case study (no improvements were observed in this case).

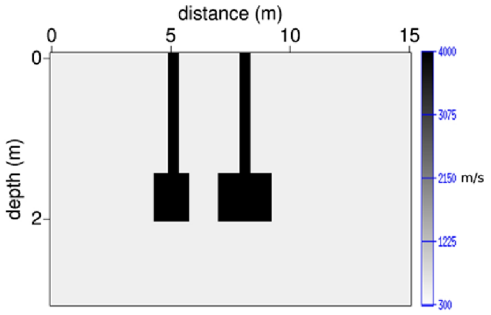

Figure 3.4 – Exact pressure wave velocity model.

Two time-domain data-sets computed with the exact wave-velocity model and with the initial homogeneous-velocity model are presented in Figure 3.5. These two data-sets are obtained using a source located at the surface, between the two concrete structures, at $x = 6.75$ m. While the direct waves are well predicted by the initial model, all the multiples coming from the reverberation between the two structures are missing in the data modeled in the initial model.

We compare the convergence of the steepest descent algorithm, l-BFGS algorithm, and the truncated Newton algorithm either in the Gauss-Newton or full Newton settings (Fig 3.6). The convergence is observed in terms of the number of solutions of wave propagation problems: while the truncated Newton method might converge more rapidly than gradient based algorithms, each nonlinear iteration is more expensive as the computation of Hessian-vector products implies the solution of additional wave propagation problems. It it thus more "fair" to compare the convergence of the different algorithms in functions of these wave propagation solves. In this case, only the truncated Newton method using the full Hessian operator is able to substantially decrease the misfit. The three other methods, steepest descent, l-BFGS, truncated Newton in the Gauss-Newton approximation, fail to converge.

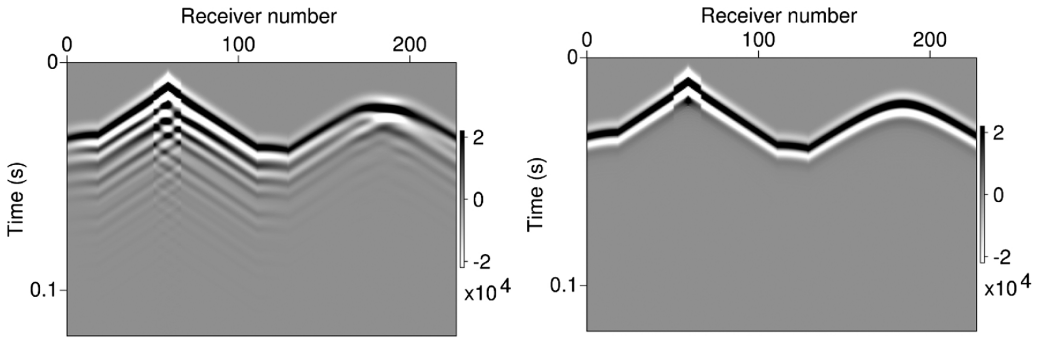

Figure 3.5 – Dataset computed in the exact model (left) and in the initial homogeneous model (right).

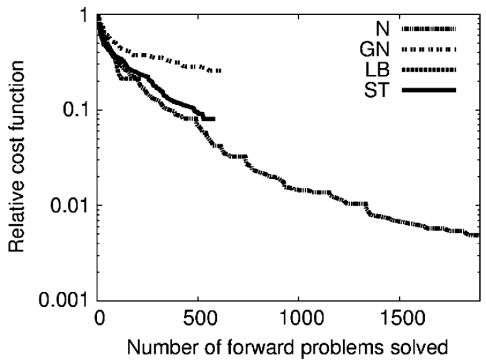

Figure 3.6 – Convergence curves: N: truncated Newton, GN: truncated Gauss–Newton, LB: l-BFGS, ST: steepest-descent.

The corresponding inversion results are presented in Figure 3.7. The four results are plotted with the same color scale. As indicated by the misfit function decrease, the two concrete structures are better reconstructed using the truncated Newton method based on the full Hessian operator, even if the wave-velocity amplitude is underestimated (this value reaches 810 $\mathrm{m{\cdot}s^{-1}}$, which is still far from the real value which is 4000 $\mathrm{m{\cdot}s^{-1}}$). The structures and their edges appear more clearly.

We compare the synthetic data computed in the time domain associated with each of the estimated wave-velocity models. The corresponding residuals (difference with the data computed in the exact model) are presented in Figure 3.9.

These figures illustrate that the multi-scattered waves appearing between the two concrete structures are only correctly interpreted by the truncated Newton strategy using the full Hessian operator.

From these results, we see that this strategy is preferable in this case than l-BFGS and first-order strategies. The imprint of high order reflections in the residuals is high, therefore an accurate estimation of the Hessian operator is required at the first iteration of the inversion. In addition, we see that the complete Hessian is required to reach satisfactory results: this is expected as the Gauss-Newton estimation discards the Hessian term correcting for second-order reflections.

The 2004 BP model. The experiment on the 2004 BP model can be seen as an exploration scale extension of the previous case study (Fig. 3.10a.). Instead of considering buried concrete structure, we consider salt bodies (red structures with velocity reaching 4500 $\mathrm{m.s^{-1}}$ in Fig. 3.10) below a deep water layer (blue top layer with velocity equal to 1500 $\mathrm{m.s^{-1}}$ in Fig. 3.10). The contrast is lower than in the previous case, however it is sufficient to create energetic multiple reflections. This effect is amplified by the introduction of a free surface condition at the air/water layer interface at the top of the model. Such geological configuration can be for instance found in the Mexico gulf environment. The 2004 BP model is representative of this environment (Billette & Brandsberg-Dahl, 2004).

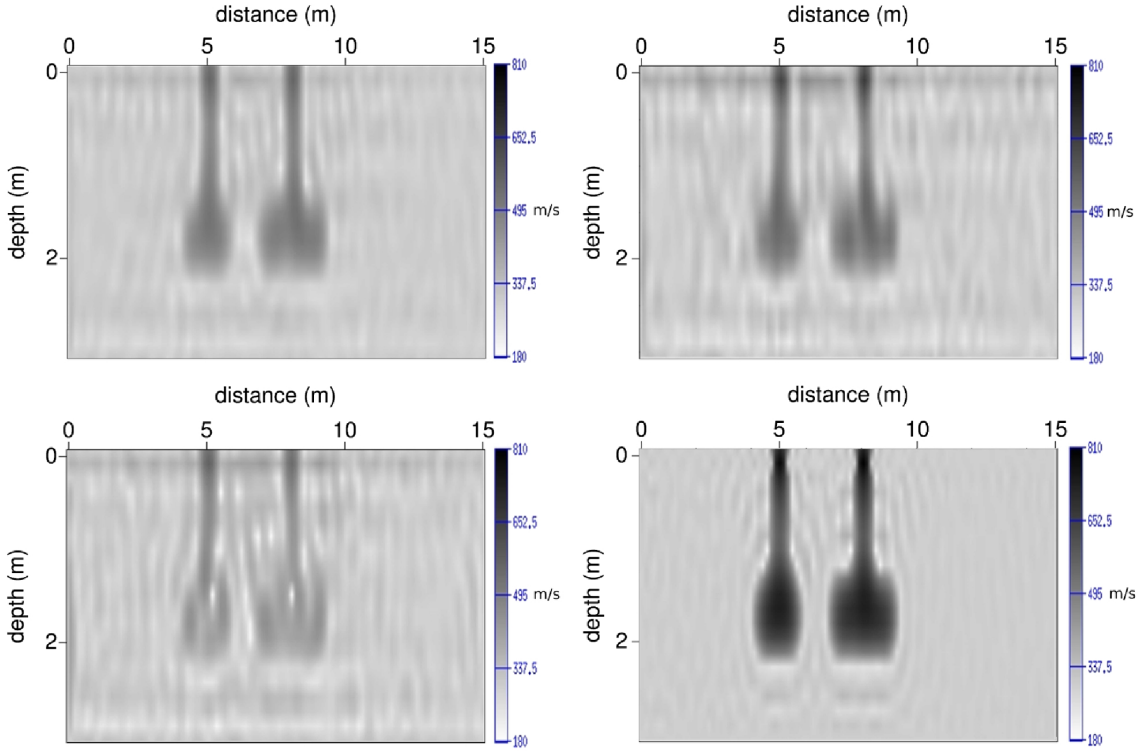

Figure 3.7 – Pressure wave velocity estimation, steepest-descent (top left) l-BFGS (top right), truncated Gauss-Newton (bottom left), truncated Newton (bottom right).

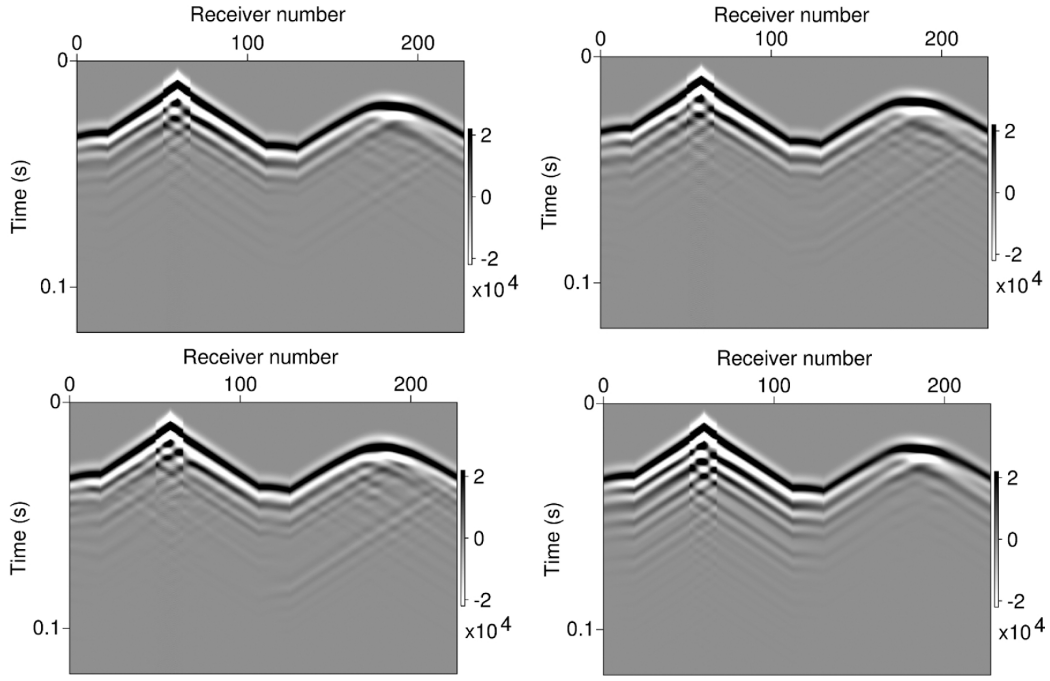

Figure 3.8 – Dataset computed in the estimated models. Steepest-descent (top left), l-BFGS (top right), Gauss–Newton (bottom left), Newton (bottom right).

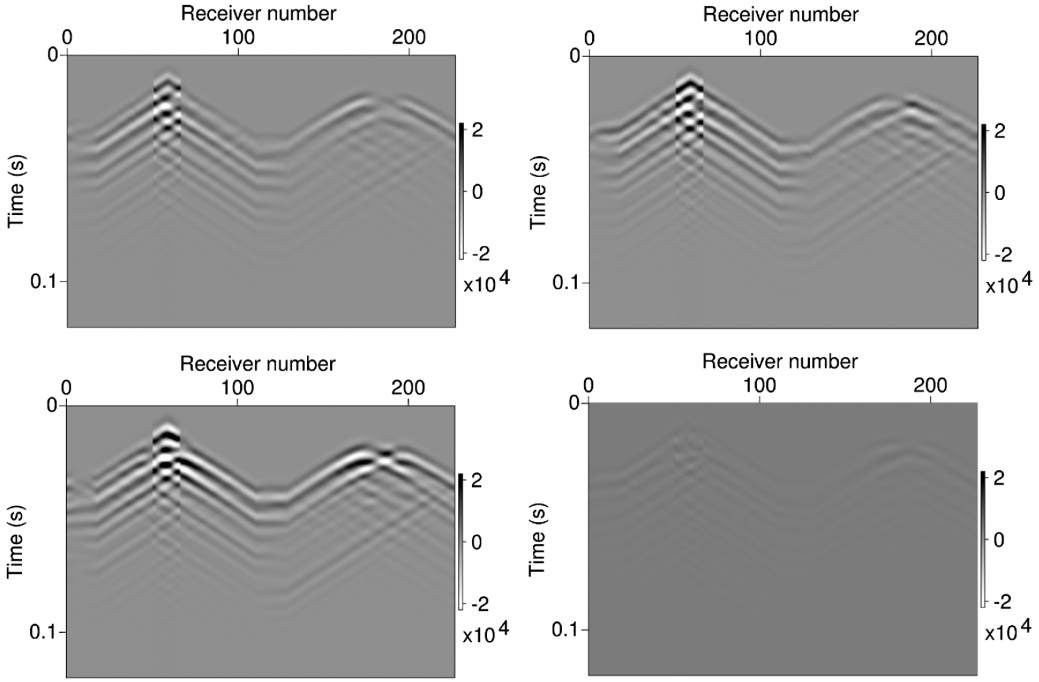

Figure 3.9 – Residuals associated with the estimated models. Steepest-descent (top left), l-BFGS (top right), Gauss–Newton (bottom left), Newton (bottom right).

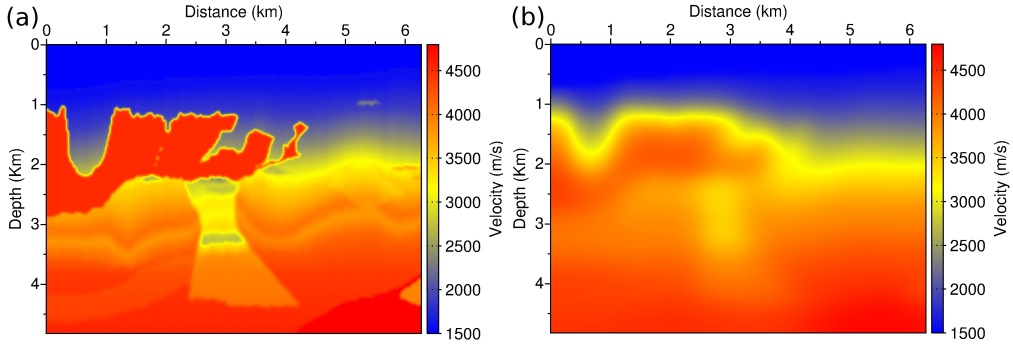

Figure 3.10 – BP 2004 synthetic exact model (a), initial model (b).

We use a surface acquisition configuration with 62 sources and 248 receivers, from $x = 50$ m to $x = 6225$ m at 25 m below the sea-level. The spatial sampling of the receivers and the sources is set up to 25 m and 100 m, respectively. A diagonal preconditioner based on the pseudo-Hessian approach is used in this experiment (equation (3.34)).

The cycle of FWI starts with an initial model which is a smooth version of the exact model. A Gaussian smoothing with 500 m correlation length (Fig. 3.10b) is used to generate it. A hierarchical frequency strategy is set up. We generate 27 data sets, from 2 Hz frequency to 19.5 Hz gathered into 8 overlapping subgroups, as presented in Table 3.1.

The four models obtained using nonlinear conjugate gradient, l-BFGS, and the truncated Newton strategy (Gauss-Newton and full Newton settings) are presented in Figure 3.11. From $x = 1$ km to 6 km, the top salt-structure is correctly delineated in the four estimations. The reconstruction of the basin between $x = 0$ km and $x = 1$ km is more difficult. This basin is responsible for high amplitude multi-scattered waves difficult to interpret, and is located at one extremity of the acquisition. The nonlinear

Group 1	2 Hz	2.25 Hz	2.5 Hz	2.75 Hz	
Group 2	2.5 Hz	3 Hz	3.5 Hz	4 Hz	
Group 3	4 Hz	4.5 Hz	5 Hz	5.5 Hz	
Group 4	5.5 Hz	6 Hz	6.5 Hz	7 Hz	
Group 5	7 Hz	7.5 Hz	8 Hz	8.5 Hz	
Group 6	8.5 Hz	9.5 Hz	10.5 Hz	11.5 Hz	
Group 7	11.5 Hz	12.5 Hz	13.5 Hz	14.5 Hz	15.5 Hz
Group 8	15.5 Hz	16.5 Hz	17.5 Hz	18.5 Hz	19.5 Hz

Table 3.1 – Frequency group strategy for the BP 2004 case study

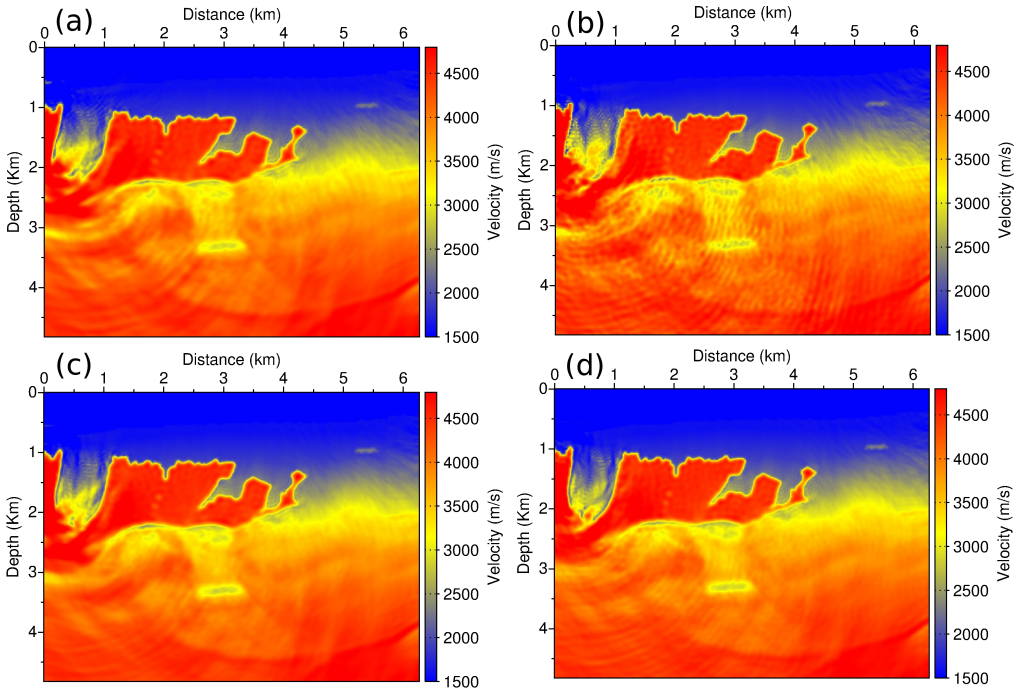

Figure 3.11 – Estimated models for the BP case study. Nonlinear conjugate gradient (a), l-BFGS method (b), truncated Gauss-Newton method (c), truncated Newton method (d).

conjugate gradient method and the l-BFGS method seem to be the most affected by this particular configuration. The geometry of the basin is not recovered and is filled with high amplitude velocities. These perturbations are also responsible for obscuring the sub-salt targets, and creating erroneous slow velocity anomalies.

Conversely, the results provided by the truncated Newton are more reliable, both using the Gauss-Newton approximation or the full Hessian operator. The best estimation is provided when using the full Hessian operator. The geometry of the basin is better recovered, and the sub-salt slow velocity anomalies better reconstructed. The possible enhancement of the inverse Hessian approximation yielded by the truncated Newton method may explain this improvement in the resolution and the stability of the inversion. The regularization effect of the truncation strategy (Kaltenbacher et al., 2008) may also contribute to the better quality of the subsurface BP 2004 model estimations. We see particularly that the l-BFGS estimation (and the nonlinear conjugate gradient estimation to a lesser extent) are affected by high frequency artifacts. Conversely, the truncated Newton and Gauss-Newton estimations are significantly smoother.

3.4.2 Multi-parameter reconstruction

The multi-parameter experiment we present is based on a 2D synthetic model extracted from the Valhall case study (Fig. 3.12). In this experiment we aim at reconstructing simultaneously the P-wave velocity v_P as well as the density ρ and an attenuation factor Q_P. This attenuation factor is introduced in the 2D frequency-domain modeling as a complex part of the velocity, following the Kolsky (1956) attenuation model

$$v_P = \overline{v}_P \left(1 - \frac{i}{2Q_P}\right), \tag{3.36}$$

where i is the imaginary unit.

We are interested in the simultaneous reconstruction of these parameters, intending to mitigate as much as possible potential trade-offs. We compare the efficiency of a standard l-BFGS approach with the truncated Newton method. The results obtained with and without the multi-parameter preconditioner (equation (3.35)) are presented to emphasize its role. We refer to the methods as (P)LB and (P)TRN in the sequel for the sake of concision.

A fixed-spread surface acquisition with sources and receivers located along the surface each 25 m is used. Six synthetic data-sets are generated from 3 Hz to 8 Hz with 1 Hz sampling. The smooth initial P-wave model is obtained using a Gaussian smoothing of the exact P-wave velocity model with a correlation length of 500 m. The Gardner's law is then used to generate an initial density model. The initial Q_P value is equal to 150 except for the water layer where it is set to 1000.

The convergence of the different strategies over 40 iterations is presented in Figure 3a. The TRN benefits from a faster convergence rate in the early iterations. The effect of the preconditioner on the convergence profile is visible only after 20 iterations for the TRN method. The LB and PLB methods follow approximately the same path excepted that PLB stops at iteration 18 due to a linesearch failure, which indicates a poor behavior of the preconditioner for the LB method.

To further compare the methods, the model error

$$e_{model}(m) = \frac{100}{M} \sum_{j=1}^{M} \frac{|m_j - \overline{m}_j|}{\overline{m}_j} \tag{3.37}$$

is introduced, where $\overline{m}$ denotes the exact model. In Figures 3(b-d) the evolution of these model errors with respect to the decrease of the data misfit is presented. The starting point (first iteration) is located at the top right side of the figures (the data misfit is normalized and is equal to 1).

The decrease of e_{v_P} (Fig. 3b) is monotonic for the four methods. The LB method provides the smallest e_{v_P} model. Conversely, e_ρ increases after approximately 10 iterations for LB and PLB (Fig 3c). This illustrates the trade-off between ρ and v_P: while the v_P estimation is improved, the ρ estimation is degraded. In particular, high frequency artifacts are introduced, and density values in the gas layers are overestimated (see also Fig. 3.12d). For PLB, the preconditioner accelerates the decrease of e_ρ at the first iterations but does not prevent its increase after 10 iterations. This could be interpreted as follows: as v_P and ρ exact models share some similarities, the cross-talk between the two parameters does not hamper their reconstruction in the early stage of the inversion. However, after few iterations, the cross-talks cause the ρ estimation to diverge from the true one.

Comparatively, TRN and PTRN provide a monotonic decrease of e_ρ (Fig 3c). The estimation of the Hessian thus seems to prevent from cross-talk effects. In addition, PTRN provides a faster reduction of the ρ error, which indicates that the preconditioner enhances the decoupling effect of the Hessian operator. The evolution of e_{Q_P} presents the same pattern: TRN and PTRN benefit from an (almost) monotonic decrease, while LB and PLB suffer from cross-talks after approximately 10 iterations (Fig 3d). The preconditioner accelerates significantly the reconstruction of Q_P both for PLB and PTRN.

Parameter estimations obtained after 40 iterations using the methods LB, TRN and PTRN are presented in Figure 3.12(c-f). Results obtained with LB after 10 iterations, before the density starts degrading, are added (denoted by LB10 in the sequel) (Fig 2c). Results obtained with PLB are omitted as they are similar to those obtained with LB.

The TRN and PTRN v_P estimations are slightly more resolved than the one provided by LB10. The gas layers are well reconstructed as well as the cap-rock above the reservoir. Due to attenuation,

almost no updates are brought below the cap-rock. The LB v_P estimation after 40 iterations has better amplitude values in the gas layers but also seems to be corrupted by high frequency artifacts (Fig 2d). The ρ estimations provided by LB10, TRN and PTRN are similar. The gas layers and the cap-rock reflector are well delineated while very small updates are brought below the cap rock (similar to v_P). The LB10 Q_P estimation shows few modifications from the starting model, although the highly attenuating shallow sediment layer starts being reconstructed. The imprint of the gas layers and the cap-rock in the attenuation model can also be seen but is weak. After 40 iterations, the amplitude of these features is better estimated however some high frequency artifacts appear. The TRN Q_P model is close from the one obtained with LB10 however the amplitude is slightly better estimated. The Q_P result provided by PTRN is the closest from the true model, even if the imprint of the gas layers is overestimated (Fig 2f).

Vertical logs extracted at $x = 7.5$ km, traversing the gas layers, are presented in Figure 4. The final LB estimation is omitted (only LB10 is presented). For v_P and ρ the estimations are similar. The trend of the rapid variations corresponding to the layers are correctly recovered, although the amplitudes are underestimated. The strong reflector corresponding to the cap-rock is not reconstructed. For Q_P, the LB10 and TRN estimations present few modifications of the initial model. Conversely, the PTRN estimation recovers correctly the attenuation value in the shallow layer. The variation of the shallowest gas layers seems to be shifted in depth. Deeper gas layers are correctly positioned, although the deepest one is clearly overestimated. The imprint of the cap rock reflector is not recovered.

3.5 Short conclusion

The truncated Newton approach allows to better account for the inverse Hessian operator within the FWI scheme. This operator has a crucial important in the subsurface parameter reconstruction process, particularly when several parameter classes are reconstructed simultaneously. We have been able to show on simple 2D acoustic frequency-domain synthetic case studies the benefit one could expect from a truncated Newton approach for the reconstruction of P-wave velocity when high amplitude multiple reflections are recorded. In a multi-parameter approach, for the simultaneous reconstruction of the P-wave velocity, density, and an attenuation parameter, the truncated Newton approach combined with a suitable preconditioner also appears to provide better results.

In the particular 2D acoustic frequency-domain context, substantial computational savings can be done by storing incident and adjoint wavefields. The feasibility of the method for time-domain approach thus still needs to be assessed. This is a very important step for moving towards more realistic large scale 3D and/or elastic FWI case studies, for which only time-domain approaches are affordable at this time. This is the topic of ongoing studies led within the SEISCOPE group.

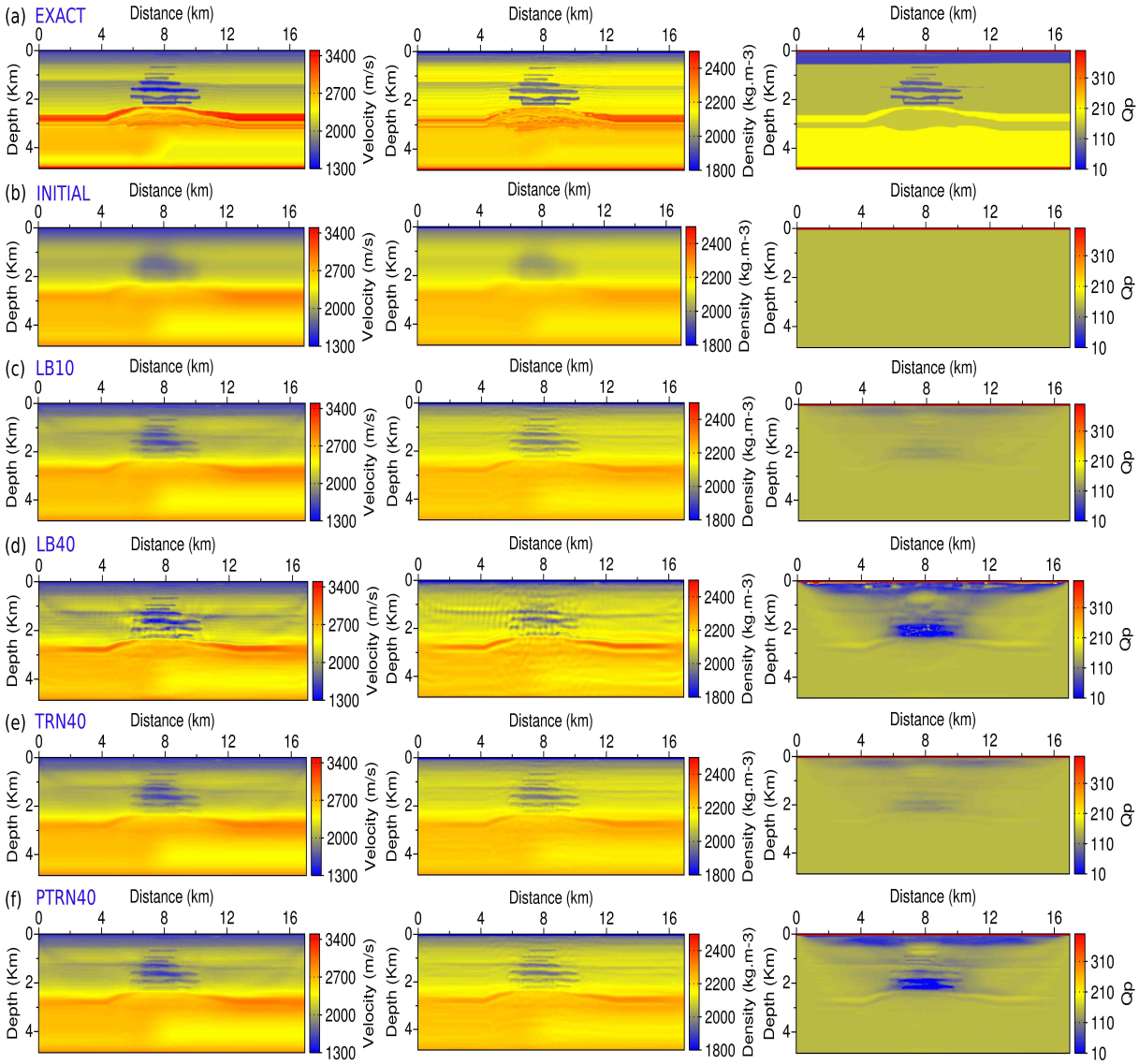

Figure 3.12 – P-wave velocity (left column), density (middle column) and attenuation (right column). Exact models (a), initial models (b), estimated models with 10 l-BFGS iterations (c), 40 l-BFGS iteration (d), 40 truncated Newton iterations (e), 40 preconditioned truncated Newton iterations (f).

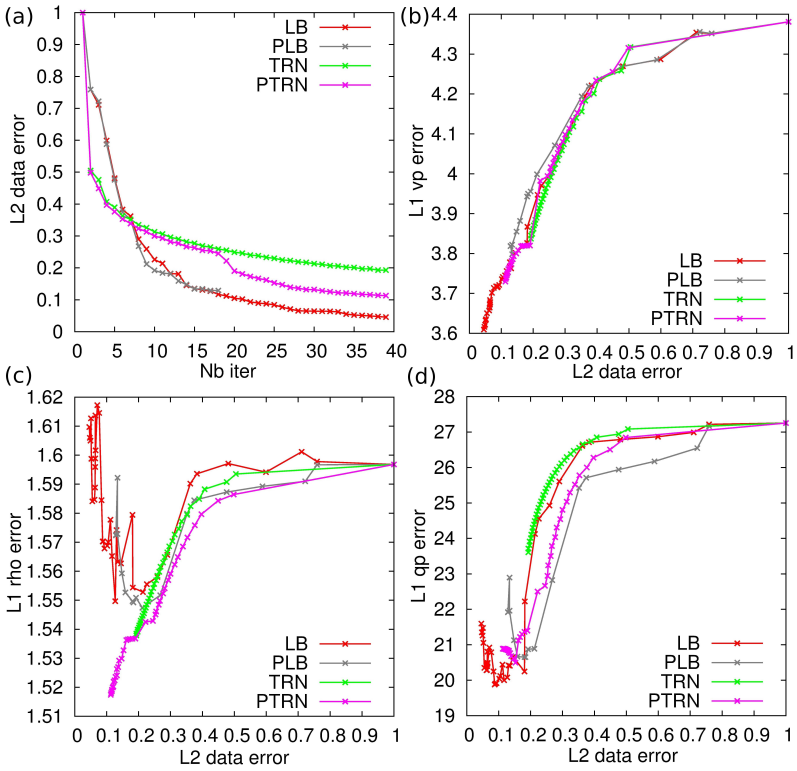

Figure 3.13 – Convergence profiles of LB, PLB, TRN, PTRN. Misfit reduction depending on the number of iterations (a). Evolution of e_{v_P} (b), e_ρ (c), e_{Q_P} (d) depending on the misfit reduction.

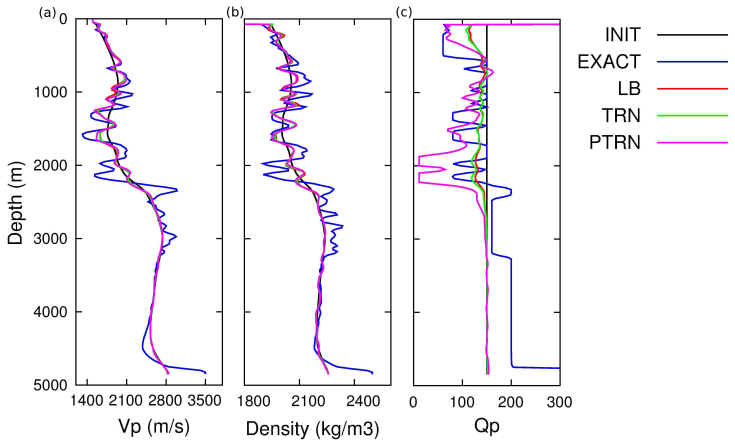

Figure 3.14 – Vertical logs taken at $x = 7.5$ km. Results after 10 l-BFGS iterations, 40 TRN iterations, 40 PTRN iterations. P-wave velocity (a), density model (b), attenuation model (c).

4 Optimal transport distance for full waveform inversion

This section summarizes the work presented in the following publications and expanded abstracts. Publications:

- Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J., 2016b. An optimal transport approach for seismic tomography: Application to 3D full waveform inversion, *Inverse Problems*, **32**(11), 115008

- Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J., 2016c. Increasing the robustness and applicability of full waveform inversion: an optimal transport distance strategy, *The Leading Edge*, **35**(12), 1060–1067

- Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J., 2016d. Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion, *Geophysical Journal International*, **205**, 345–377

Abstracts:

- Métivier, L., Brossier, R., Oudet, E., Mérigot, Q., & Virieux, J., 2016e. An optimal transport distance for full-waveform inversion: Application to the 2014 chevron benchmark data set, in *SEG Technical Program Expanded Abstracts 2016*, pp. 1278–1283

- Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J., 2016a. Overcoming cycle skipping in FWI: An optimal transport approach, in *Expanded Abstracts, 78th Annual EAGE Meeting (Vienna)*

4.1 Context and problematic

The work presented in this Section is related to one of the most stringent limitation for the applicability of FWI at the exploration scale, which is usually referred to as cycle skipping, or phase ambiguity. This difficulty appears for the reconstruction of wave velocity models, the primary parameter usually inverted through FWI. The cycle skipping is a difficulty coming from the least-squares formalism of FWI. Each sample of the observed data is compared to the sample of the synthetic data at the same position in time and/or in space. This choice is problematic: if the initial velocity model predicts the signal with a time shift larger than half a period, minimizing the least-squares distance between observed and calculated data amounts to match the observed data up to one or several phase shifts. This yields an incorrect estimation of the subsurface model which cannot be overcome through iterations: the optimization is locked into a local minimum. An illustration of this phenomenon, where the seismic data is considered schematically as a sinusoidal temporal signal, is presented in Figure 4.1.

Overcoming this difficulty has been a recurrent objective since the introduction of FWI by Lailly (1983b) and Tarantola (1984). We start this section by summarizing the different approaches which have been proposed to mitigate this difficulty in the last decades.

The probably first strategy to overcome this limitation has been proposed by Pratt (1999). It consists in matching the data following a multi-scale (hierarchical) approach. The lowest frequency components of the data are matched first. This increases the attraction valley of the misfit function as, in this case, the initial velocity model should only explain the data up to half the period corresponding to the lowest frequency components that have been extracted. The result of the first inversion serves as an initial model for an inversion of data containing higher frequencies. This procedure can be iterated until the whole seismic data has been interpreted. This strategy is used for instance in Bunks et al. (1995); Sirgue & Pratt (2004) and Operto et al. (2004).

This multi-scale approach can be complemented with offset and time-windowing strategies. Time-windowing is used to select the diving waves and remove the reflected energy from the observed seismograms. The offset is increased progressively, as large offsets correspond to diving waves traveling across a long distance between the subsurface, therefore containing a large number of oscillations. These waves are more subject to cycle skipping. This time-windowing and offset selection is also known as layer stripping technique: the shallow part of the subsurface is first reconstructed, the depth of investigation being progressively increased by this data selection strategy. See Shipp & Singh (2002); Wang & Rao

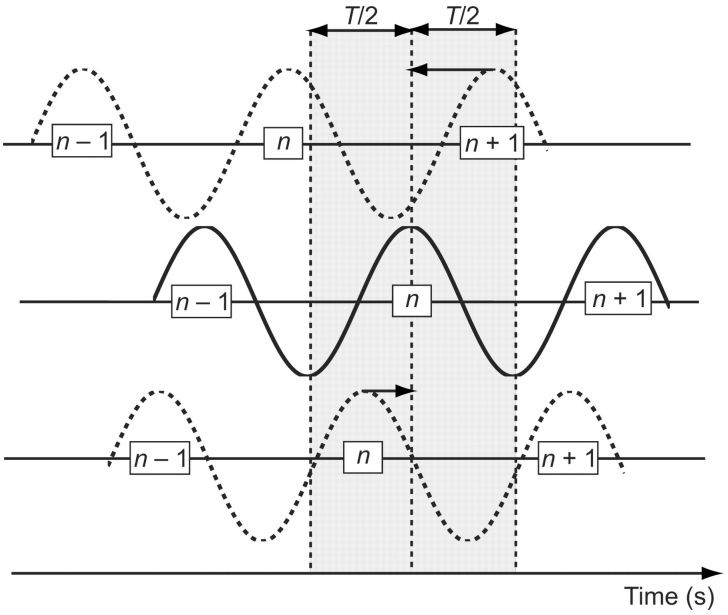

Figure 4.1 – Schematic example of the cycle skipping/phase ambiguity issue on sinusoidal signals. As soon as the initial shift is larger than half a period of the signal, the fit of the signal using a least-squares distance is performed up to one or several phase shifts. On may try to fit the $n+1$ dashed wriggle of the top signal with the n continuous wriggle of the middle signal moving to the wrong direction. The bottom dashed signal predicts the n wriggle in less than half-period leading to a correct updating direction (figure from Virieux & Operto, 2009).

(2009) and Brossier et al. (2009) for examples of applications in the 2D acoustic (respectively elastic) approximation.

Despite these successful applications, the multi-scale approach does not really overcome the cycle skipping limitation. Instead, the data interpretation is re-organized in such a way that this limitation does not preclude the estimation of the velocity through FWI. Commonly encountered difficulties for real data application preventing this strategy to produce reliable velocity estimations encompass: the impossibility of building an accurate enough and kinematically compatible initial velocity model, the presence of strong noise corrupting the low frequency part of the data, or offset limitations in the acquisition design.

Several attempts have thus been made to modify the FWI misfit function itself, to avoid comparing the seismic signal using the L^2 distance, and to yield a more robust, convex misfit function, less prone to cycle skipping. Two classes of strategies designed to achieve this objective can be identified, referred to as data-domain and image-domain techniques in the following.

The underlying concept of data-domain technique relies so far on a hybridization between tomography methods and FWI. These hybrid methods try to emphasize the matching of travel-times instead of the full signal, to recover the properties of tomography methods, while still benefiting from the expected high resolution power of FWI. One of the first attempt is the design of the wave-equation tomography (WETT) proposed by Luo & Schuster (1991). This is a tomography method, aiming at matching travel-times. However, while classical tomography methods rely on travel-time picking in the observed data (a possibly heavy pre-processing step) and the computation of travel-times through asymptotic approaches, the travel-times misfit is directly estimated from the cross-correlation of the observed and synthetic traces. This method is interesting as it bridges the gap between tomography and FWI from a formal point of view: a full wave modeling engine is used to compute the synthetic data, and the method can be interpreted as a modification of the FWI misfit function, making possible to use the adjoint formalism to compute the associated gradient, as is commonly done in FWI. Originating from

exploration geophysics, this strategy has been adopted by the seismology community as the finite-frequency tomography method (Dahlen et al., 2000; Montelli et al., 2004; Tromp et al., 2005; Nolet, 2008).

An attempt to enhance the robustness of the automatic travel-time misfit computation through warping has been proposed by Ma & Hale (2013). Dynamic image warping is a technology originally designed for pattern recognition in signal processing. In a recent study, Hale (2013) has demonstrated that this method could be applied to determine time shifts between seismograms. However, in the presence of non-predicted events (i.e. reflections), the estimation of the time-shifts through cross-correlation or warping technique collapses.

More recently, the design of a misfit function based on deconvolution has been proposed by Luo & Sava (2011). The travel-time misfit between the synthetic and observed data is computed through a matching filter obtained by the deconvolution of the synthetic data by the observed data. The FWI problem is reformulated as the focusing of the energy at zero-lag for these matching filters. This method has started to be applied to realistic scale case-studies in seismic exploration and seems to provide a more robust misfit function, less prone to cycle skipping (Warner & Guasch, 2014). The main difficulty for the application of this strategy consists in the proper design of the penalization function which is used to refocus the energy at zero-lag, which might be application dependent.

In seismology, other data-domain modifications of the misfit function have been proposed. Fichtner et al. (2008) propose to use a time-frequency analysis of the data through a Gabor transform in order to extract both the travel-times and the amplitude envelope information from the seismic signal. This allows to define a misfit function as a sum of two terms measuring the misfit between travel-times and amplitude envelope separately. A similar strategy has been proposed by Bozdağ et al. (2011) where the amplitude and travel-time information are computed following the Hilbert transform instead of the Gabor transform. Both strategies can be used in combination with different time-windowing strategies (Maggi et al., 2009). Envelope inversion has also been investigated in the context of exploration seismology (Luo & Wu, 2015).

Parallel to the development of these data-domain techniques, the development of image-domain techniques started with the design of Differential Semblance Optimization (DSO) (Symes & Kern, 1994) and later on wave equation migration velocity analysis (WEMVA) (Sava & Biondi, 2004a,b; Symes, 2008). These methods rely on the separability of scales assumption: the velocity model is decomposed as the sum of a smooth background model and a high wavenumber reflectivity model. The reflectivity is related to the smooth background model through an imaging condition: it is the sum for each source of the cross-correlation between the incident wavefield and the back-propagated residuals computed in the smooth background velocity model. This imaging condition can be extended using either an offset selection (Symes & Kern, 1994) or an illumination angle selection (Biondi & Symes, 2004) in the residuals (the angles are easily accessible when the reflectivity is computed through asymptotic techniques), or a time lag in the cross-correlation (Faye & Jeannot, 1986; Sava & Fomel, 2006; Biondi & Almomin, 2013).

Within this framework, an extended image consists in a collection of reflectivity models depending on one of these additional parameters (offset, angle, time lag). This extended image is used to probe the consistency of the smooth background velocity model: the uniqueness of the subsurface implies that for the correct background, the energy should be focused in the image domain, either along the offset/angle dimension, or at zero lag. A new optimization problem is thus defined, either as the penalization of the defocusing of the energy, or as the maximization of the coherency of the energy in the image domain. The corresponding misfit function is minimized iteratively, following standard numerical optimization schemes.

As for the deconvolution approach, one drawback of these strategies rely on the potential sensitivity to the choice of the penalization function (also called annihilator function). Another difficulty is related to their computational cost, because of the repeated construction of reflectivity images, which has to be performed at each iteration of the reconstruction of the smooth background velocity model. This high computational cost seems to have precluded the use of these techniques for 3D waveform inversion up to now. It should also be noted that these methods are based on the assumption that only primary reflections will be used to generate the extended image through migration, which requires non negligible data pre-processing. Locally coherent events in the image-domain associated with, for instance, multiple

reflections, would yield inconsistent smooth background velocity models (Lambaré, 2002). Accounting for higher order reflection within this framework is nontrivial (Cocher & Chauris, 2014; Cocher et al., 2015)

4.2 Proposed approach: a misfit function based on an optimal transport distance

Recently, a new data-domain modification of the misfit function based on optimal transport distances has been proposed by Engquist & Froese (2014). We first introduce briefly the main concepts of optimal transport. For the sake of simplicity, this introduction relies on a discrete formulation of the problem, even if the formalism of optimal transport is much more general (the Wasserstein distance introduced hereafter is defined on measure spaces).

In the last fifteen years, the field of optimal transport has been put on the front scene through the work of many mathematicians, as testified by the books of Villani (2003, 2008); Ambrosio et al. (2008); Santambrogio (2015). In particular, the metric underlain by the optimal transport distance has been used to establish new existence results of solution to nonlinear partial differential equations such as the Boltzmann equations. Applications of optimal transport to image processing have also been investigated and have yield very convincing result, for instance for contrast and color mappings, pattern recognition, shape classification, texture synthesis and texture mixing (see Lellmann et al. (2014) and references therein for a detailed state-of-the-art in image processing)

Optimal transport finds its roots back in 1780, in the work of the French mathematician Gaspard Monge. Monge supervised a bridge building site. Piles of sand needed to be displaced to fill in holes. Monge expressed mathematically how this displacement could be achieved optimally, to minimize the effort of the workers. More than 150 years later, a modern and suitable mathematical formulation for this problem was introduced by Kantorovich (1942), as a relaxation of the Monge problem. In his description, the initial piles can be associated with N quantities μ_i, located at the points x_i, for $i = 1, \dots N$. The configuration of the holes is associated with M quantities ν_j, located at point y_j, for $j = 1, \dots, M$. The total quantity of sand requested to fill in the holes ν_j is assumed to be exactly equal to the total quantity of available sand μ_i (mass conservation assumption).

$$\sum_{i=1}^{N} \mu_i = \sum_{j=1}^{M} \nu_j. \tag{4.1}$$

Kantorovich considers the ensemble of displacements making possible to fill the holes with the quantities ν_j from the sand piles μ_i. These displacements can be represented as matrices γ with N rows and M columns. An entry γ_{ij} tells how much from the pile μ_i should be moved to fill in the hole ν_j. Mapping the ensemble of piles μ onto the holes ν requires that the sum of the elements of the ith row of γ_{ij} is equal to μ_i, while the sum of the elements of the jth column of γ_{ij} is equal to ν_j. A matrix satisfying this assumption is called a "transport plan". An example of such transport plan is presented in Figure 4.2.

An infinity of transport plan allowing to move the sand piles μ to the holes ν exists. The optimal transport plan $\overline{\gamma}$ minimizes a function measuring the total displacement cost. This cost is the sum of the elementary costs associated with the elementary displacements. The cost of an elementary displacement between x_i and y_j is measured as the product between the mass which is actually transferred γ_{ij}, multiplied by the distance between x_i and y_j. This measure implies that a balance has to be found between the amount of mass which is transported, and the distance on which it is transported. Mathematically, this is formulated as the linear programming problem

$$\min_{\gamma_{ij} \geq 0} \sum_{ij} \gamma_{ij} \|x_i - y_j\|, \text{ subject to } \sum_j \gamma_{ij} = p_i, \ \sum_i \gamma_{ij} = q_j. \tag{4.4}$$

where $\|x_i - y_j\|$ denotes a distance between x_i and y_j, called the ground distance (often the Euclidean distance).

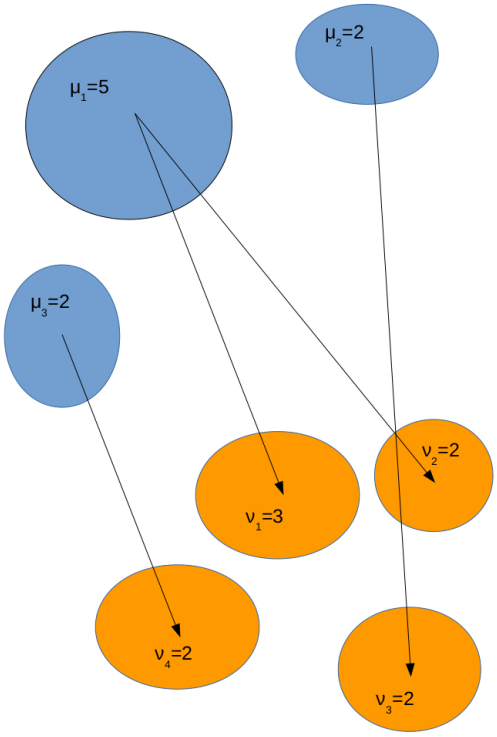

Figure 4.2 – An example of transport plan for the solution of the transport problem between a distribution μ_1, μ_2, μ_3 towards a distribution $\nu_1, \nu_2, \nu_3, \nu_4$. The matrix γ representing this transport plan is

$$\gamma = \begin{pmatrix} 3 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}. \tag{4.3}$$

The solution of the problem (4.4) defines a distance between μ and ν, named Wasserstein distance, defined by

$$W_p(\mu,\nu) = \left(\min_{\gamma_{ij} \geq 0} \sum_{ij} \gamma_{ij} \|x_i - y_j\|^p \text{ subject to } \sum_j \gamma_{ij} = p_i, \ \sum_i \gamma_{ij} = q_j \right)^{1/p}. \tag{4.5}$$

These distances are of particular interest for FWI because of their convexity with respect to shifts between the distribution μ and ν. This property is demonstrated by Engquist & Froese (2014) for the hat function, and illustrated numerically for a Ricker function decomposed in its positive in negative part. In short, considering two real positive functions $f_1(t)$ and $f_2(t)$ such that

$$f_2(t) = f_1(t - t_0), \tag{4.6}$$

one can expect that

$$W_p^p(f_1, f_2) \simeq |t_0|^p. \tag{4.7}$$

This property is extremely interesting in the context of seismic imaging and FWI. Velocity perturbations are mainly responsible for shifting in time the seismic data. Therefore a FWI misfit function which is convex with respect to time shifts should be convex with respect to the velocity. This has been the main motivation for investigating the use of these distances in the framework of FWI.

Measuring the misfit between seismic data using an optimal transport distance is nonetheless not straightforward. Two requirements must be fulfilled by the data for the optimal transport problem (4.4) to have a solution.

- The data must be positive: $f_1(t) \geq 0, \quad f_2(t) \geq 0$;
- The total mass has to be conserved: $\int_t f_1(t)dt = \int_t f_2(t)dt$.

The positivity assumption is clearly not satisfied by seismic data, as it is oscillatory in essence. On the other hand the second assumption should be satisfied, as the integral in time of a seismic signal amounts to its zero frequency, which is zero in practice. In addition to this positivity issue, the size of the discrete problem which has to be solved is a severe limitation for the use of optimal transport for FWI. Standard seismic data involve from $O(10^6)$ discrete data for 2D acquisition to $O(10^9)$ discrete data in 3D. Standard linear programming algorithms based on the simplex strategy can not be used to treat these problems as they have a cubic complexity in $O(N^3)$.

4.3 Mathematical description

4.3.1 The Kantorovich-Rubinstein norm

We have proposed a strategy to overcome these two limitations which is based on the W_1 distance. Our strategy relies on the dual formulation of the Kantorovich problem, which is expressed in discrete form as

$$\max_{\varphi_i,\psi_j} \sum_i \varphi_i \mu_i + \sum_j \psi_j \nu_j, \quad \varphi_i + \psi_j \leq \|x_i - y_j\|^p. \tag{4.8}$$

(see for instance Villani (2003) or Santambrogio (2015) for this duality result).

In the particular case where μ and ν are discretized in the same space, and $p = 1$, the dual problem (4.8) can be simplified into the linear programming problem

$$\max_{\varphi_i} \sum_i \varphi_i(\mu_i - \nu_i), \quad \varphi \in \mathrm{Lip}_1, \tag{4.9}$$

where Lip_1 denotes the space of 1 Lipschitz function *i.e.*

$$\mathrm{Lip}_1 = \{\varphi : \forall i, j, \quad |\varphi_i - \varphi_j| \leq \|x_i - x_j\|\}. \tag{4.10}$$

The duality result in this particular case is known as the Kantorovich-Rubinstein theorem (Santambrogio, 2015, chap. 3.2.1). While, under its primal form (4.4), the optimal transport problem has a solution only for data satisfying the positivity and mass conservation assumption, the dual problem (4.9) has a solution even if the positivity assumption breaks down. This interesting property makes this problem an interesting tool for comparing seismic data.

In addition, in case the mass conservation is not satisfied, a generalization of (4.9) can be easily found: this generalization consists in complementing the 1-Lipschitz constraint with a bound constraint on the infinity norm of the dual variable φ. The problem (4.9) thus becomes

$$\max_{\varphi_i} \sum_i \varphi_i(\mu_i - \nu_i), \quad \varphi \in \mathrm{Lip}_1, \quad \forall i, \quad |\varphi_i| \leq \lambda. \tag{4.11}$$

The generalization (4.11) corresponds to the definition of the Kantorovich-Rubinstein (KR) norm (Bogachev, 2007). Besides the link with optimal transport, the KR norm can also be interpreted as a generalization of the L^1 norm (in a similar sense that the generalization from Total Variation to Total Generalized Variation norms), and shares some properties with the Meyer's G-norm. These similarities are studied in detail by Lellmann et al. (2014), where the use of the KR norm is proposed as an alternative to the L^1 norm in a Total Variation denoising problem.

The distance based on (4.11) is the one we have decided to implement in the framework of FWI. The first reason is its ability to deal with non-positive data. The second reason is, in case the mass conservation is not strictly satisfied because of numerical approximation, there is a straightforward fix to this difficulty.

4.3.2 Numerical strategy

At this stage, the difficulty relies in the design of an efficient enough numerical scheme to estimate this distance. We give a description of this scheme in the following. We assume in this description that the dimension d is set to 3. This means that we are going to compute the distance between data cubes μ and ν, where the vertical dimension corresponds to the time t, and the horizontal x and transverse dimensions y correspond to the location of receivers at the surface. Using a triplet of index ijk to denote one position in the data cube, the problem (4.11) is discretized as

$$\begin{array}{l} \max\limits_{\varphi_{ijk}} \sum\limits_{ijk} \varphi_{ijk}\left(\mu_{ijk}-\nu_{ijk}\right), \;\; s.c. \\ \left\{ \begin{array}{ll} \forall\,(i,j,k),(l,m,n) & |\varphi_{ijk}-\varphi_{lmn}| < \|(t_i,x_j,y_k)-(t_l,x_m,y_n)\|, \\ \forall\,(i,j,k) & |\varphi_{ijk}| \leq \lambda. \end{array} \right. \end{array} \tag{4.12}$$

Computing a numerical approximation of the solution of (4.12) requires the solution of a convex optimization problem involving $O(N^2)$ linear constraints which would lead to an unacceptable computational time for the large scale problems induced by FWI applications. For this reason, we consider an alternative problem where the Lipschitz constraint is expressed in terms of the ℓ_1 norm instead of the standard Euclidean norm $\|.\|$. This leads to

$$\begin{array}{l} \max\limits_{\varphi_{ijk}} \sum\limits_{ijk} \varphi_{ijk}\left(\mu_{ijk}-\nu_{ijk}\right), \;\; s.c. \\ \left\{ \begin{array}{ll} \forall\,(i,j,k),(l,m,n) & |\varphi_{ijk}-\varphi_{lmn}| < |t_i-t_l|+|x_j-x_m|+|y_k-y_n|, \\ \forall\,(i,j,k) & |\varphi_{ijk}| \leq \lambda. \end{array} \right. \end{array} \tag{4.13}$$

This change amount to a change of the ground distance from the Euclidean to the ℓ_1 distance in the primal optimal transport problem (4.4).

We then use a property of the ℓ_1 norm on $\mathbb{R}^d$ to reduce the number of constraints from $O(N^2)$ to $O(N)$. We can prove indeed that the two following assertions are equivalent

$$(A1) \quad \forall\,(i,j,k),(l,m,n)\; |\varphi_{ijk}-\varphi_{lmn}| < |t_i-t_l|+|x_j-x_m|+|y_k-y_n|,$$

$$(A2) \quad \left\{ \begin{array}{ll} \forall\,(i,j,k) & |\varphi_{i+1,j,k}-\varphi_{ijk}| < |t_{i+1}-t_i|, \\ \forall\,(i,j,k) & |\varphi_{i,j+1,k}-\varphi_{ijk}| < |x_{j+1}-x_j|, \\ \forall\,(i,j,k) & |\varphi_{i,j,k+1}-\varphi_{ijk}| < |y_{k+1}-y_k|. \end{array} \right. \tag{4.14}$$

$(A1)$ obviously implies $(A2)$. To prove the reciprocal implication, consider a pair of points on the mesh denoted by u and v, such that

$$u=(t_i,x_j,y_k), \;\; v=(t_l,x_m,y_n). \tag{4.15}$$

A sequence of points $w_q=(t_{i_q},x_{j_q},y_{k_q}),\; q=1,\ldots,M$ can be selected to form a path on the mesh from u to v, such that $w_1=u$, $w_M=v$, and w_q are all adjacent on the grid, with monotonically varying coordinates. The key is to see that, for such a sequence of points, the ℓ_1 distance on $\mathbb{R}^d$ ensures that

$$||v-u||_1 = \sum_{q=1}^{M} ||w_{q+1}-w_q||_1. \tag{4.16}$$

(see Fig. 4.3 for an illustration). Now, consider a function φ satisfying $(A2)$. The triangle inequality yields

$$||\varphi(v)-\varphi(u)||_1 \leq \sum_{q=1}^{M} ||\varphi(w_{q+1})-\varphi(w_q)||_1. \tag{4.17}$$

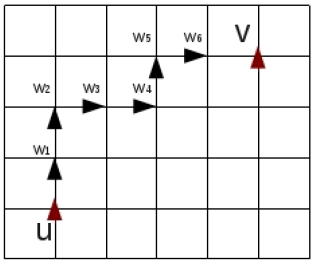

Figure 4.3 – Illustration of the property (4.16) for a 2D mesh. Considering two points u and v, a sequence of adjacent points $w_1, \ldots, w_6$ with monotonically varying coordinates is found to connect them. Such a sequence always exists and is non-unique.

As the points w_q are adjacent, the local inequalities described by $(A2)$, satisfied by φ, yield

$$\sum_{q=1}^{M} ||\varphi(w_{q+1}) - \varphi(w_q)||_1 \leq \sum_{q=1}^{M} ||w_{q+1} - w_q||_1. \tag{4.18}$$

Putting together equations (4.17), (4.18) and (4.16) yields

$$||\varphi(v) - \varphi(u)||_1 \leq ||v - u||_1, \tag{4.19}$$

or

$$|\varphi_{ijk} - \varphi_{lmn}| < |t_i - t_l| + |x_j - x_m| + |y_k - y_n|, \tag{4.20}$$

which proves the proposition.

Using the equivalence (4.14), the problem (4.12) can be rewritten in its equivalent form

$$\begin{array}{l} \max\limits_{\varphi_{ijk}} \sum\limits_{ijk} \varphi_{ijk} \left(\mu_{ijk} - \nu_{ijk}\right), \quad s.c. \\ \left\{ \begin{array}{ll} \forall\ (i,j,k), & |\varphi_{i+1,j,k} - \varphi_{ijk}| < |t_{i+1} - t_i| = h_t, \\ \forall\ (i,j,k), & |\varphi_{i,j+1,k} - \varphi_{ijk}| < |x_{j+1} - x_j| = h_x, \\ \forall\ (i,j,k), & |\varphi_{i,j,k+1} - \varphi_{ijk}| < |y_{k+1} - y_k| = h_y, \\ \forall\ (i,j,k), & |\varphi_{ijk}| < \lambda. \end{array} \right. \end{array} \tag{4.21}$$

The problem (4.21) is equivalent to (4.12) and only involves $O(N)$ constraints. This reduction of the order of the number of constraints gives the possibility to design an efficient numerical strategy to compute the KR norm.

After this first step, we reformulate the problem (4.21) as the convex non-smooth problem

$$\max_{\varphi} f_1(\varphi) + f_2(A\varphi), \tag{4.22}$$

where

$$f_1(\varphi) = \sum_{i,j,k \in \mathcal{A}} \varphi_{ijk} \left(\mu_{ijk} - \nu_{ijk}\right), \quad f_2(\varphi) = i_K\left(\varphi\right), \tag{4.23}$$

with K the unit hypercube

$$K = \left\{x \in \mathbb{R}^P,\ |x_i| \leq 1,\ i = 1, \ldots P\right\}, \tag{4.24}$$

where P is the total number of linear constraints, i_K the indicator function of K

$$i_K(x) = \left| \begin{array}{lll} 0 & \text{if} & x \in K \\ +\infty & \text{if} & x \notin K, \end{array} \right. \tag{4.25}$$

and $A \in \mathbb{M}_{P,N}(\mathbb{R})$ is a rectangular real matrix with P rows and N columns (N being the total number of discrete unknowns) such that

$$A = \left[D_t \;\; D_x \;\; D_y \;\; \frac{1}{\lambda} I_N \right]^T, \tag{4.26}$$

where I_N is the real identity matrix of size N and D_t, D_x, D_y are the forward finite differences operators

$$\begin{cases} (D_t\varphi)_{ijk} = \dfrac{\varphi_{i+1,j,k} - \varphi_{ijk}}{h_t}, \\ (D_x\varphi)_{ijk} = \dfrac{\varphi_{i,j+1,k} - \varphi_{ijk}}{h_x}, \\ (D_y\varphi)_{ijk} = \dfrac{\varphi_{i,j,k+1} - \varphi_{ijk}}{h_y}. \end{cases} \tag{4.27}$$

Convex optimization problems of type (4.22) involving at least one non differentiable functions (here $f_2(\varphi)$) can be efficiently solved through proximal splitting techniques. From our numerical experiments, among this class of methods, the simultaneous-direction method of multipliers (SDMM) was found to achieve the fastest convergence (Combettes & Pesquet, 2011).

At each iteration of the SDMM algorithm, the computation of the proximity operators of the two functions f_1 and f_2 (scaled by a positive factor γ) is required. Closed-form formulations for these operators can be found such that

$$\text{prox}_{\gamma f_1}(\varphi) = \varphi - \gamma(\mu + \nu), \tag{4.28}$$

$$\forall i = 1, \ldots, P, \;\; \left(\text{prox}_{\gamma f_2}(x)\right)_i = \left(\text{prox}_{i_K}(x)\right)_i = \left| \begin{array}{lll} x_i & \text{if} & -1 \leq x_i \leq 1 \\ 1 & \text{if} & x_i > 1 \\ -1 & \text{if} & x_i < -1. \end{array} \right. \tag{4.29}$$

Note that the scaling γ only acts on the proximity operator of γf_1 as $\gamma i_K = i_K$. The closed-form formulations (4.28) and (4.29) are inexpensive to compute with an overall complexity in $O(N)$ operations.

In addition, the SDMM algorithm requires at each iteration the solution of a linear system involving the matrix $I + A^T A$, which is the most time-consuming part of the algorithm. **The important result we have obtained** in (Métivier et al., 2016b) is to show that the operator $A^T A$ corresponds to the second-order finite difference discretization of the Laplacian operator with Neumann homogeneous boundary conditions. Therefore, the inversion of the matrix $I + A^T A$ can be performed in quasi-linear complexity ($(O(N \log N))$ through FFT based solvers Swarztrauber (1974) or in linear complexity ($O(N)$) through multigrid algorithms Brandt (1977).

The combination of the reduction of the number of constraints using the property of the ℓ_1 distance and the observation that the matrix appearing in the SDMM strategy actually corresponds to the discretization of the Poisson's equation offers the possibility to design an efficient numerical method to compute the KR norm for large scale problems.

4.3.3 Implementation within the FWI scheme

The implementation of this new misfit function within the FWI scheme requires the ability to compute its gradient, such that it can be minimized following a local optimization scheme (nonlinear conjugate gradient, l-BFGS). The adjoint state formalism for the computation of the misfit function presented in the previous Section provides a very systematic way to investigate this question. Following this formalism, it can be shown that a modification of the misfit function corresponds, for the gradient computation, to a modification of the source term for the adjoint wavefield equation.

In the particular case of the KR norm considered here, one can show that this source term is actually the dual variable $\overline{\varphi}$ which realizes the maximum of the maximization problem (4.22).

$$\overline{\varphi} = \arg\max_{\varphi} f_1(\varphi) + f_2(A\varphi), \tag{4.30}$$

The demonstration of this property is given in (Métivier et al., 2016d,b). In other words, at each FWI nonlinear iteration, only one optimal transport problem has to be solved to compute the value of the

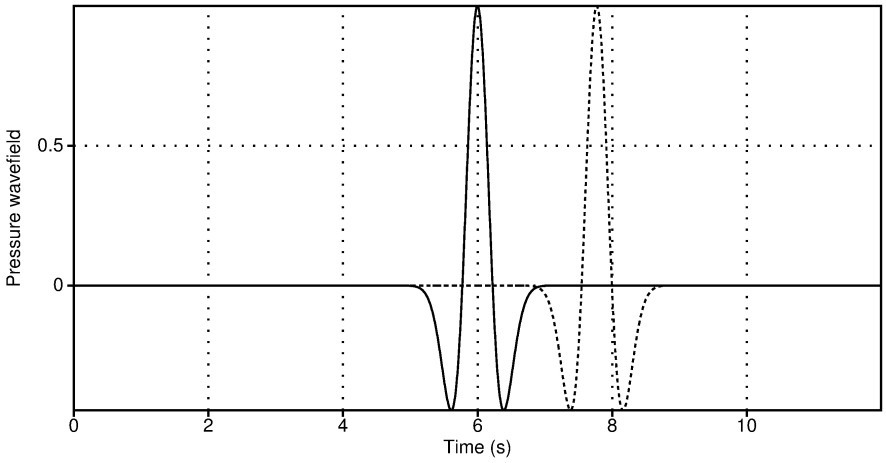

Figure 4.4 – Ricker signal playing the role of observed data (thick line). Example of shifted in time Ricker (dash line).

misfit function and its gradient. The value of the misfit function is the direct output of the optimal transport problem. The gradient is obtained by correlating the incident and adjoint wavefields, the adjoint wavefields being computed as the backpropagation of a source term which is the argmax of the same optimal transport problem.

4.4 Numerical studies

We now provide a summary of the different numerical illustrations we have been able to conduct to assess the interest of the KR distance for FWI, compared to the standard least-squares approach. We first compare the misfit function itself with respect to the least-squares distance in simplified configuration. We then present 2D FWI results on three case studies: Marmousi, BP2004, Chevron 2014 data. We finally provide an example of 3D FWI results on the Overthrust SEG/EAGE model.

4.4.1 Misfit functions comparison

1D Ricker time functions. We start with a schematic 1D experiment similar to the one proposed in Engquist & Froese (2014). A Ricker time signal serves as observed data, and the predicted data corresponds to this same Ricker signal, shifted in time (Fig. 4.4). The L^2 and KR distances, as functions of the time shift, are compared. The corresponding misfit profiles are presented in Figure 4.5. The misfit based on the L^2 distance presents two local minima, typical of cycle skipping, apart the global minimum. The misfit based on the KR distance presents a single minimum, which indicates a better robustness to the time shift. The misfit function appears as not differentiable at its minimum and concave. The non-differentiability at the minimum is reminiscent of the L_1 norm, which might not be surprising according to the strong relation between the KR norm and this norm (Lellmann et al., 2014).

The increase of the attraction valley is therefore not as important as one would expect from a strategy supposed to improve the convexity of the misfit function, however recovering a single minimum is already an interesting feature. In addition, the non-differentiability at the minimum should not be a strong problem for applications as noise (coming from data, imperfect modeling, for instance) always prevent to converge towards the exact minimum.

2D misfit map with velocity gradient and surface acquisition. Going further in the analysis of misfit maps, a 2D configuration is considered. A fixed-spread surface acquisition is used, constituted

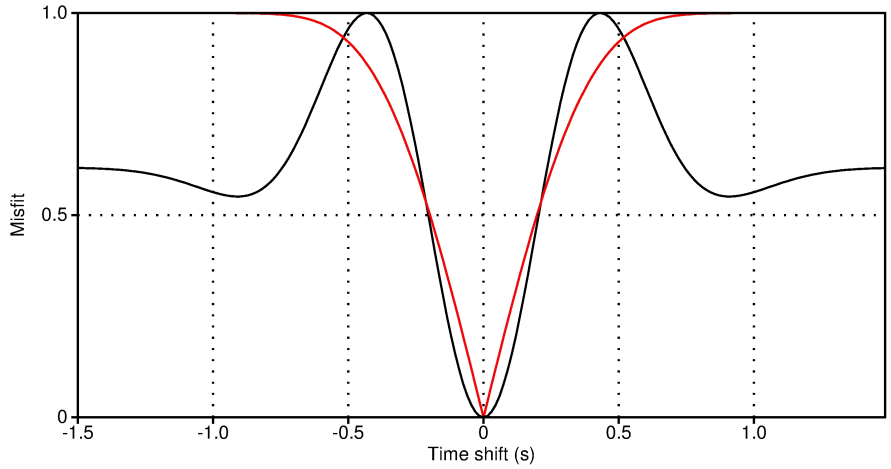

Figure 4.5 – Misfit function depending on the time shift of the Ricker signal, using the L^2 distance (black) and the KR distance (red).

of 168 receivers equally spaced each 100 m, at 50 m depth, from $x = 0$ km to $x = 16.85$ km. A single source is located at $x = 8.45$ km. Similar to the experiment presented in Mulder & Plessix (2008) to emphasize the local minima of the L^2 misfit function, the velocity model is assumed to vary linearly in depth such that

$$v_P(x, z) = v_{P,0} + \alpha z. \tag{4.31}$$

The P-wave velocity is thus parameterized by the velocity at the origin $v_{P,0}$ and the velocity vertical gradient α. The reference velocity model is chosen so that $v_{P,0} = 2000$ m s^{-1} and $\alpha = 0.7$ s^{-1}. A reference data set is computed in this model. The L^2 and KR misfit functions are then evaluated on a grid of 41×41 points such that

$$v_{P,0} \in [1750, 2250], \quad \alpha \in [0.4, 0.9], \tag{4.32}$$

with discretization steps $\Delta v_{P,0} = 12.5$ m s^{-1} and $\Delta\alpha = 0.015$ s^{-1}. The results are presented in Figure 4.6. Interestingly, the L^2 misfit function presents two narrow valleys of attraction, on both sides of the valley where the global minimum is located. Finding this global minimum with local descent methods thus requires to start in the correct valley. Conversely, even if the KR misfit function still possesses local minima, the two narrow valleys of attraction on both sides of the central valley containing the global minima have been lifted up. The valley on the left is not anymore an obstacle to converge to the central valley. The valley on the right still plays the role of a barrier, however the height of the barrier has been significantly reduced compared to the L^2 case. This is an indication of a better behavior of the KR misfit function compared to the L^2 misfit function.

Contrary to the previous 1D experiment, the misfit function appears as more regular. A better insight on the shape of the global minimum is provided in Figure 4.7 where misfit function profiles along the velocity gradient α and the background velocity $v_{P,0}$ respectively are presented. These profiles well illustrate how the secondary valley of attraction are lifted up for the KR distance. They also show that even if the KR misfit function is smoother, in the vicinity of the global minimum, the misfit function exhibits similarities with the L^1 norm and appears as not differentiable at the global minimum.

In terms of resolution power, it should be noted that the width of the global valley of attraction is almost the same for the L^2 and KR misfit functions. This is different from what is observed when cross-correlation or deconvolution approaches are used to reduce the sensitivity to the initial model and cycle skipping: in this case the valley of attraction is strongly widen. This reflects a resolution loss of the imaging method as, near the solution, models possibly quite different yield approximately the same misfit. The KR misfit function is thus expected to keep the same resolution as the L^2 misfit function while relaxing the constraint on the choice of the initial model.

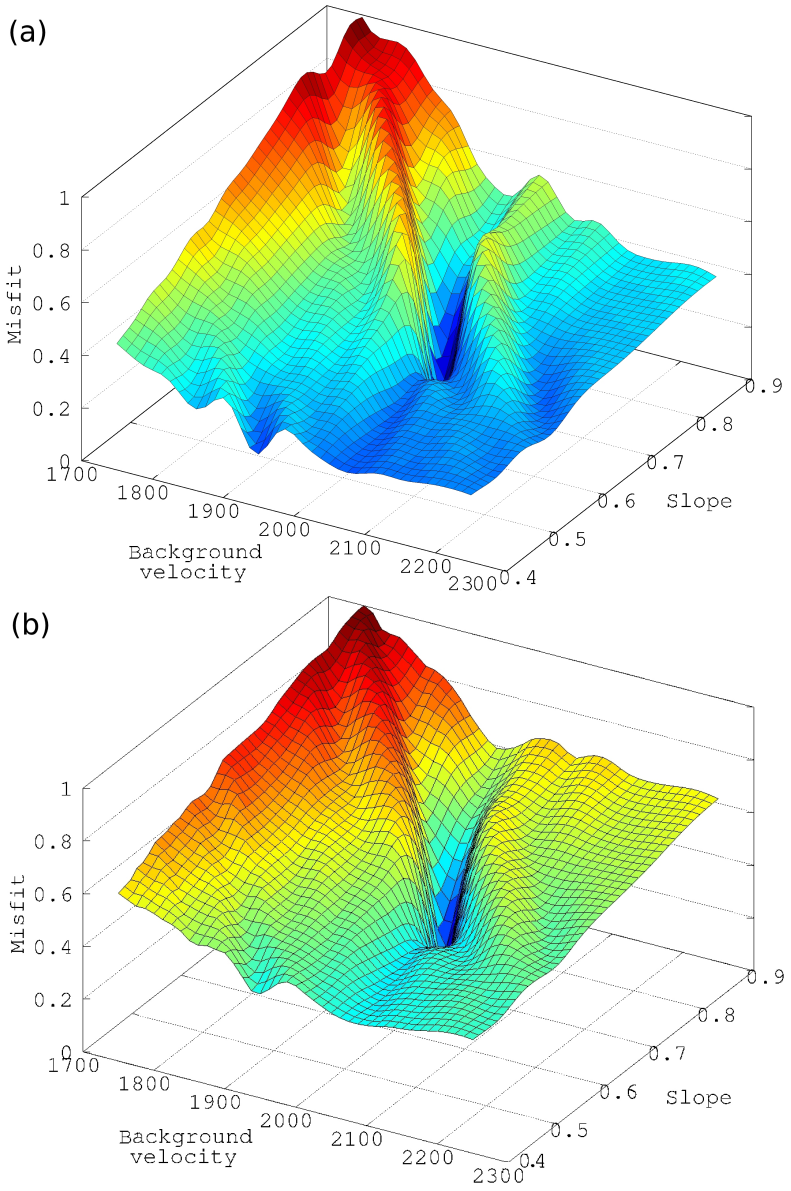

Figure 4.6 – L^2 misfit function (a) and KR misfit function (b) depending on the background velocity $v_{P,0}$ and the slope α.

4.4.2 2D FWI results

Marmousi 2 model. The P-wave velocity of the Marmousi 2 benchmark model is presented in Figure 4.8a. A fixed-spread surface acquisition with 128 sources each 125 m and 168 receivers each 100 m, at 50 m depth, is considered. The synthetic data is generated using a Ricker source function centered on 5 Hz. The frequency content of the source is low-pass filtered below 3 Hz to mimic realistic seismic data for which this frequency band is contaminated by noise and therefore unavailable for inversion. Two initial models are considered: the first contains the main features of the exact model, only with smoother interfaces. The second is a strongly smoothed version of the exact model with very weak lateral variation and underestimated growth of the velocity in depth.

The results obtained using the L^2 and KR distances are presented in Figure 4.8(d-g). The conver-

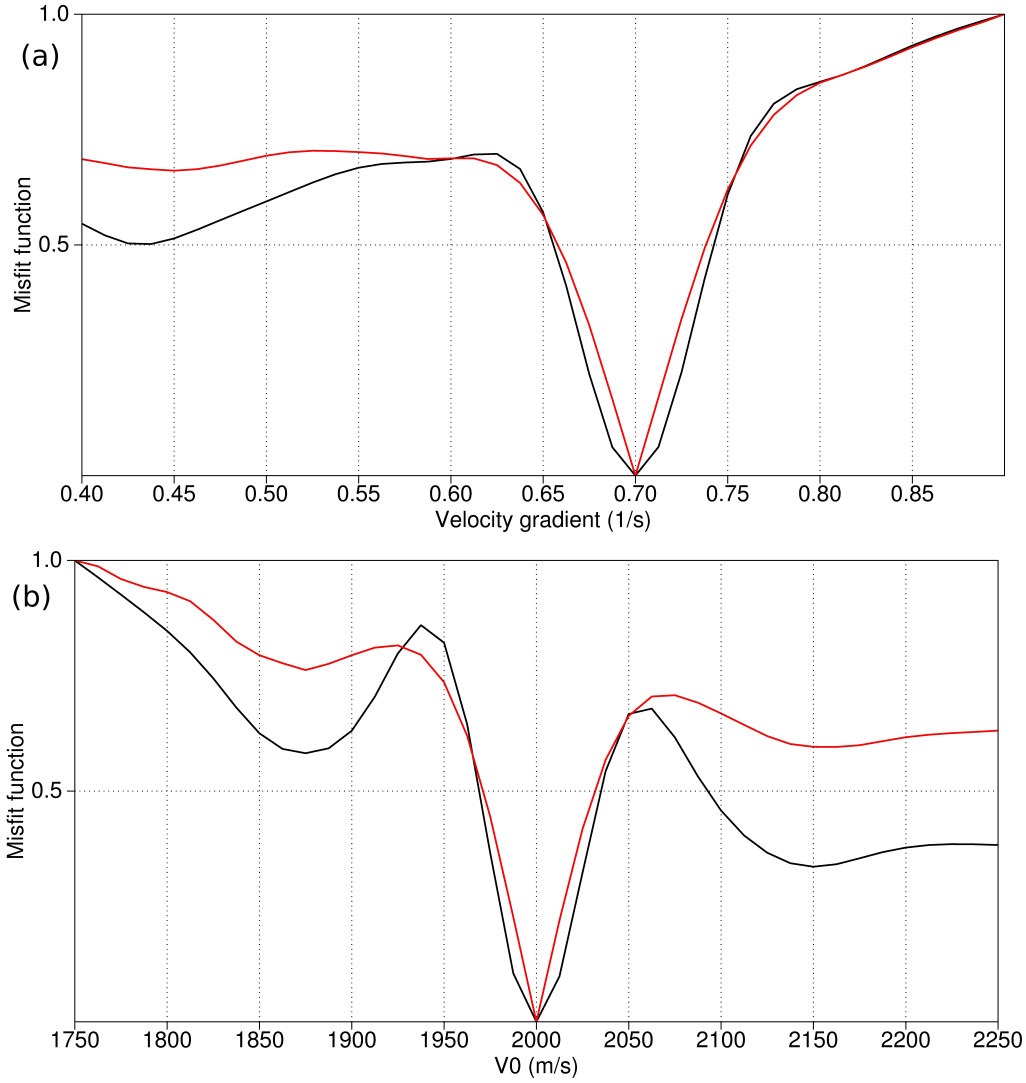

Figure 4.7 – L^2 (black) and KR (red) misfit function depending on the velocity gradient α for a constant $v_{P,0} = 2000$ m.s^{-1} (a). L^2 (black) and KR (red) misfit function depending on $v_{P,0}$ for a constant velocity gradient $\alpha = 0.7$ (b).

gence to a correct estimation of the P-wave velocity model is reached using both the L^2 and KR distances starting from the first initial model. Starting from the second initial model, only the results obtained using the KR distance are meaningful (Fig.4.8g). The poor initial approximation of the P-wave velocity is responsible for cycle skipping and the L^2 estimation converges towards a local minimum (Fig.4.8f). The estimation obtained with the KR norm is significantly closer from the true model, despite low velocity artifacts in the shallow part and in depth. This result highlights the strong potential of the KR approach for FWI. When the inaccuracy of the initial model would yield to cycle skipping issues using the L^2 distance, the KR distance is able to produce a meaningful estimation of the true model.

BP 2004 model. To further assess the interest of the KR approach, we investigate the BP 2004 model. This model is representative of the geology of the Gulf of Mexico (Billette & Brandsberg-Dahl, 2004). It is presented in Figure 4.10a. This area is characterized by a deep water environment and the

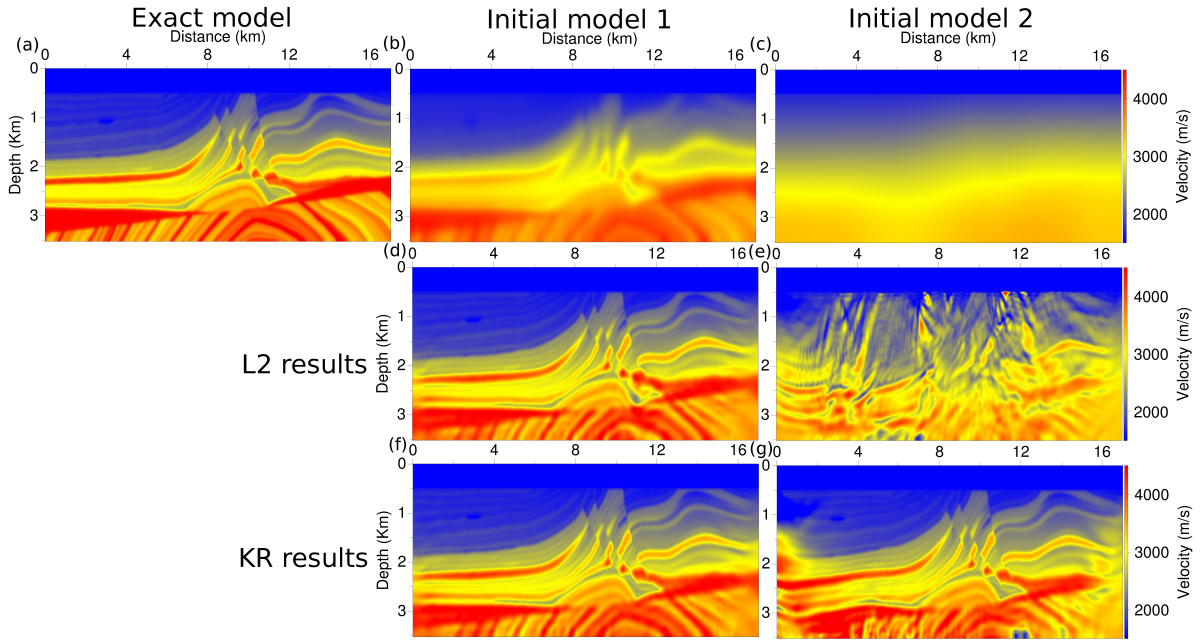

Figure 4.8 – Marmousi model case study. Exact model (a), initial model 1 (b), initial model 2 (c), results obtained with the L^2 distance starting from model 1 (d), from model 2 (e), results obtained with the KR distance starting from model 1 (f), from model 2 (g).

presence of complex salt structures. The large P-wave velocity value of the salt structures is responsible for most of the energy of the seismic signal to be reflected back to the receivers from the interface between the water layer and the salt. Only a few percentage of energy of the seismic signal travels within the structure and below before being recorded. This particular configuration makes seismic imaging in the presence of salt structures challenging. In this experiment, we want to assess if the KR distance can help in dealing with this issue and with the reconstruction of salt bodies. To this end, the initial model has been designed such that the imprint of the salt structure has been totally removed (Fig 4.10b).

A fixed-spread surface acquisition is used, with 128 sources and 161 receivers distant from 125 m and 100 m respectively. The depth of the sources and receivers is set to $z = 50$ m. The density model is assumed to be homogeneous such that $\rho_0 = 1000$ kg.m^{-3}. The wavelet used to generate the data is based on a Ricker wavelet centered on 5 Hz. A whitening of the frequency content is performed before a minimum phase Butterworth low-pass and high-pass filters are applied. The spectrum of the resulting wavelet is within an interval from 3 Hz to 9 Hz (Fig. 4.9).

From the starting model, FWI using a standard L^2 distance fails to produce meaningful results (Fig. 4.10c). The time-window is reduced to 4.6 s to focus the inversion on the shallowest part of the model and to reduce cycle skipping issues, however this does not prevent the minimization from converging towards a local minimum. The incorrect P-wave velocity estimation of the starting model prevents the FWI algorithm from locating the reflectors associated with the top of the salt. Instead, diffracting points are created to match the most energetic events without lateral coherency.

In comparison, the same experiment is performed using the optimal transport distance (Fig. 4.10d). As can be seen, the top of the salt structure is correctly delineated. Synthetic shot-gathers corresponding to the source located at $x = 8$ km, computed in the exact model, initial model, L^2 estimation, and optimal transport estimation, are presented in Figure 4.11. The strong reflection coming from the top of salt is inaccurately predicted by the L^2 estimation; in particular, the reflected energy which is introduced is discontinuous (Fig.4.11c). In comparison, the optimal transport estimation yields a correct prediction of this reflection (Fig.4.11d).

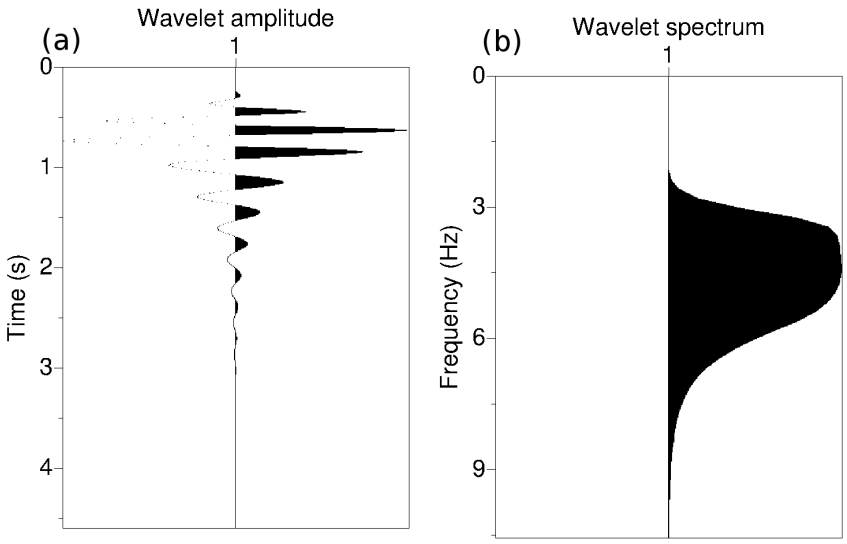

Figure 4.9 – Source wavelet used to generate the synthetic dataset on the BP 2004 model (a). This source is obtained from a Ricker wavelet centered on 5 Hz. A whitening of its frequency content is performed before a low-pass and high-pass filter are applied, so that the corresponding spectrum spans an interval from 3 Hz to 9 Hz (b).

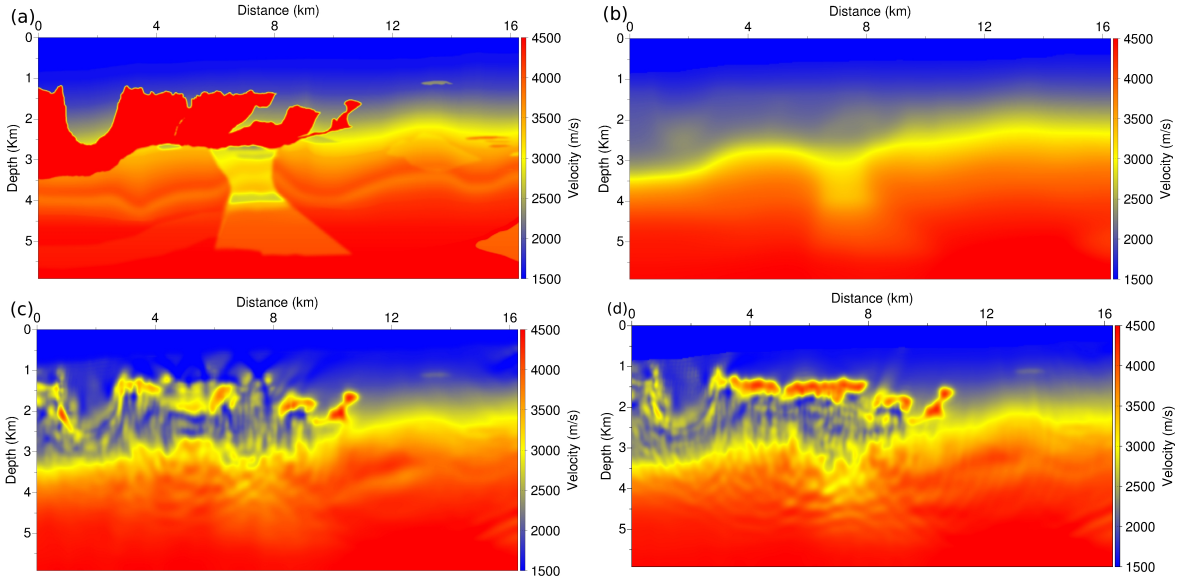

Figure 4.10 – BP 2004 exact model (a) and initial model (b). P-wave velocity estimation with a standard L^2 norm on short-time window data (4.6 s) (c). The same with the optimal transport distance (d).

Building on this result, a layer stripping workflow is suggested. Five increasing time-windows are defined, with recording time equal to 4.6 s, 5.75 s, 6.9 s, 9.2 s, and finally 10.3 s. For each time window, several inversions are performed, with decreasing levels of smoothing, for a total of 15 inversions across all the time windows.

Intermediate to final results are presented in Figure 4.12. As can be seen, the salt structure is practically entirely recovered at the end of the cycle of inversions (Fig 4.12f). A continuous progression

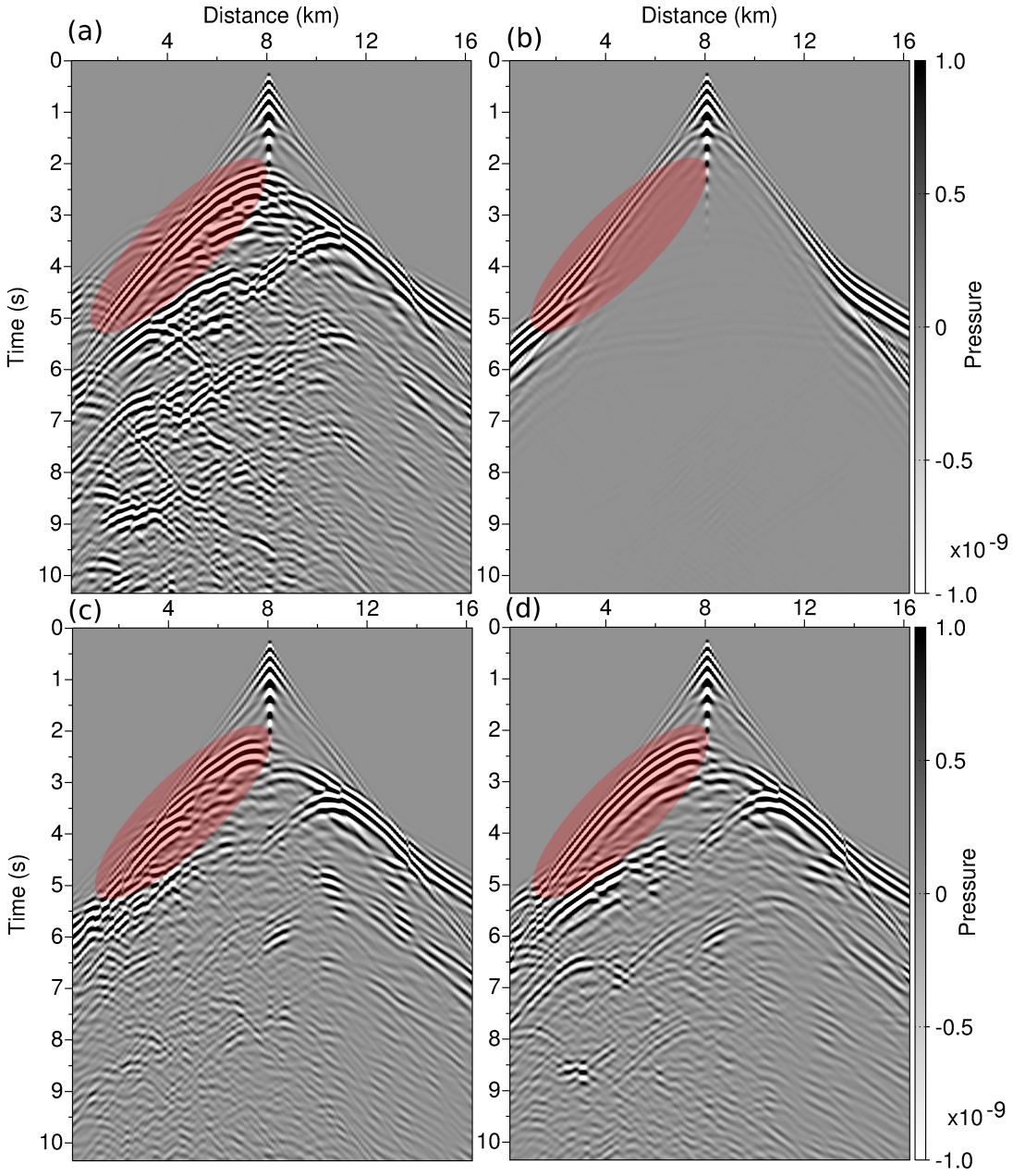

Figure 4.11 – BP 2004 exact data (a) and initial data (b). Predicted data in the final model using a standard L^2 norm (c). Predicted data in the final model using the optimal transport distance using together with a layer stripping workflow (d). The red ellipses highlight the reflection on the salt roof. This reflection is not present in the initial data (b). Its reconstruction using the L^2 distance is discontinuous (c). The use of the optimal transport distance yields a better reconstruction of this event (d).

is achieved from the initial delineation of the top of the salt structure to the full reconstruction of its deeper parts. The subsalt zone, however, whose reconstruction is critical, is not satisfactorily recovered. To this purpose, a possibility would consist in building an initial model from this reconstruction by freezing the salt, which is correctly delineated, and smoothing below the salt. From such an initial model, our previous study show that FWI based on the the L^2 distance with a truncated Newton optimization

strategy should be able to reconstruct accurately the subsalt region (Métivier et al., 2014a).

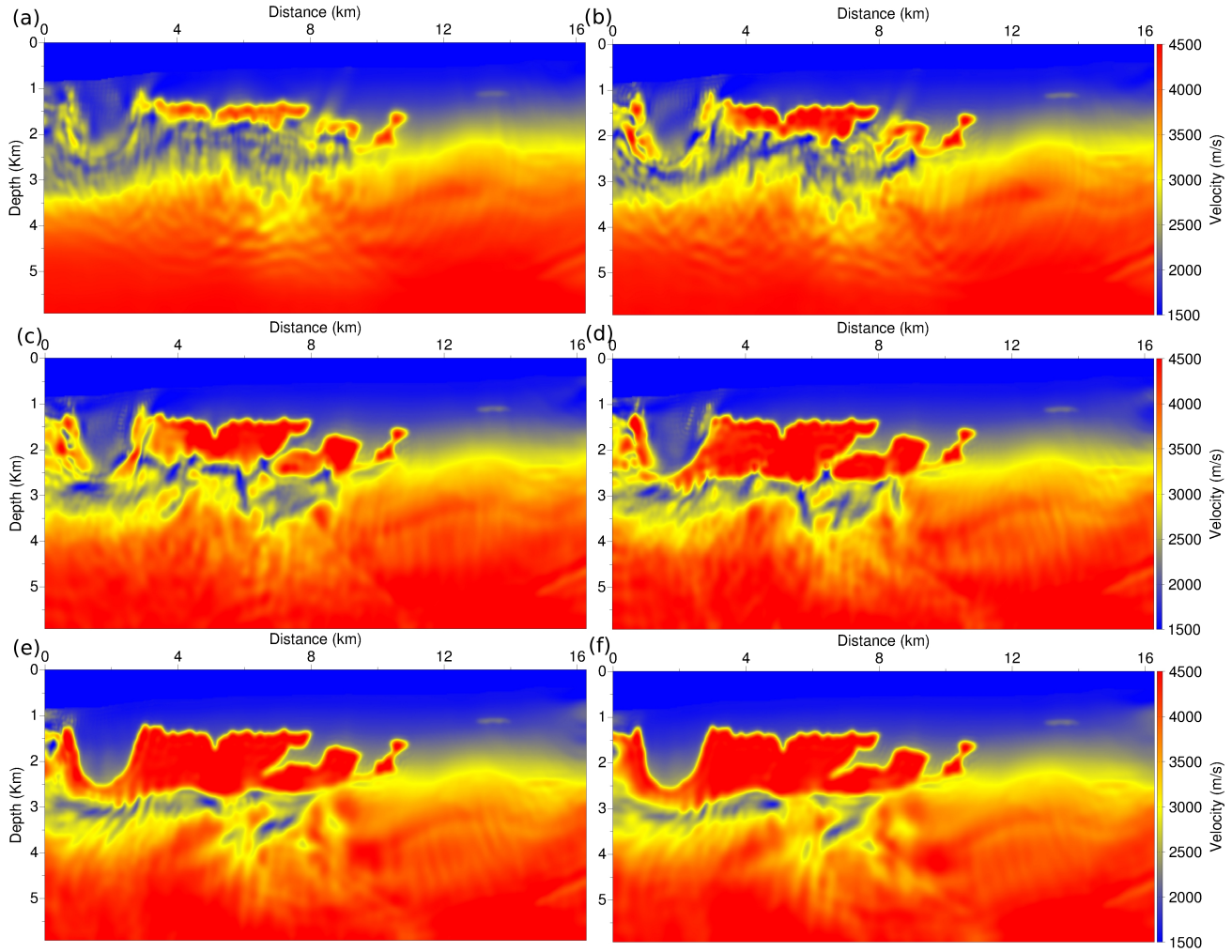

Figure 4.12 – BP 2004 P-wave velocity estimation computed after the 1st(a), 3rd (b), 6th (c), 9th (d), 12th (e), and 15th (f) inversion using the optimal transport distance.

A better insight of the reconstruction process is given by the synthetic data computed in intermediate models throughout the different steps of the workflow presented in Figure 4.13. The shot-gathers are computed for a source located at $x = 8$ km. A particular attention should be accorded to the left part of the seismogram (red rectangles), as this part corresponds to the main salt structure in the exact model. After interpreting correctly the reflections coming from the salt roof (Fig.4.13a), the transmitted wave traveling within and below the salt is progressively adjusted while deeper reflections are also progressively integrated (Fig.4.13b to Fig.4.13f). This behavior is in contrast with standard multi-scale approaches for which the transmitted energy is fitted prior to the reflected energy. However, this might not be input to the use of the optimal transport distance. Due to the high velocity contrast, the reflected energy dominates the transmitted energy in the data. This, in conjunction with the layer stripping strategy which focuses the prior steps of the inversion toward short offset data, favors the fit of the reflections prior to the diving waves.

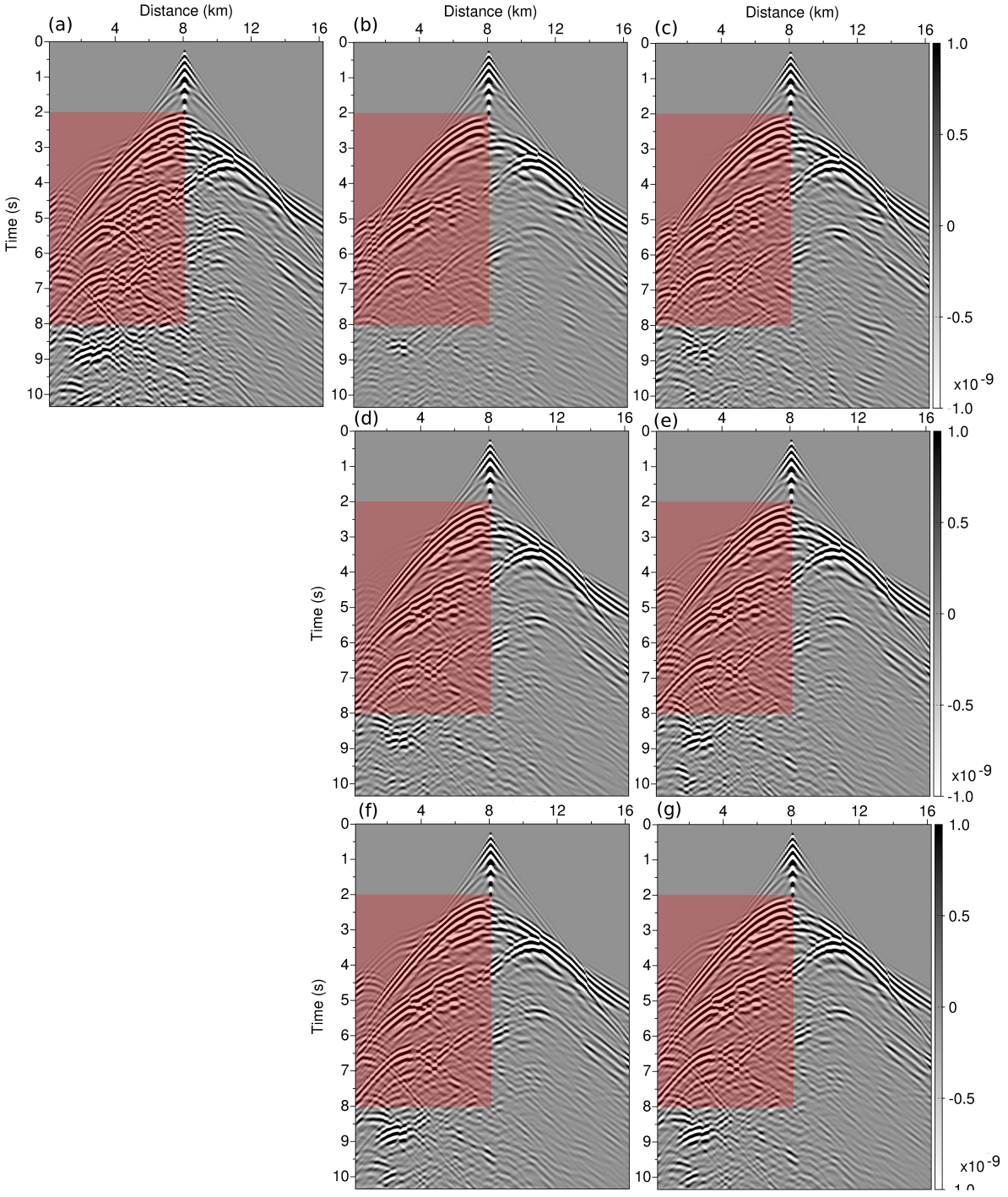

Figure 4.13 – Synthetic data in the exact model (a) and in the intermediate models obtained with FWI using an optimal transport distance after the 1st(b), 3rd (c), 6th (d), 9th (e), 12th (f), and 15th (g) inversion. The red rectangles highlight the shot-gather zone associated with the diving waves traveling within the salt dome and the reflections generated by deeper interfaces.

Chevron 2014 data. In 2014, the Chevron oil company has issued a blind benchmark synthetic dataset for FWI. The aim of such blind benchmark is to provide realistic exploration seismic data to practitioners with which they can experiment various FWI workflow and test methodological developments. As the exact model which has served to build the data is not known, such a case study is closer from an application to field data than synthetic experiments for which the exact model is known.

The Chevron 2014 benchmark dataset is built from a 2D isotropic elastic modeling engine. A frequency-dependent noise has been added to the data to mimic a realistic dataset. Especially, the Signal Over Noise Ratio (SNR) for low frequencies (below 3 Hz) is much less than for higher frequencies. Free surface multiples are incorporated in the data. A streamer acquisition is used, with a maximum of 8 km offset, with 321 receivers by sources equally spaced each 25 m. The depth of the sources and receivers is $z = 15$ m. Among the 1600 available shots gathers, 256 have been used in this study, with a distance of 150 m between each sources. A frequency continuation strategy similar to the one proposed by Bunks et al. (1995) is implemented: Butterworth low-pass and high-pass filters are applied to the selected shot-gathers to generate an ensemble of 15 datasets with an increasing bandwidth from $2 - 4$ Hz to $2 - 25$ Hz.

The shot-gathers corresponding to the source located at $x = 150$ m are presented for the 1st, 5th, 10th and 15th frequency bands in Figure 4.14. As mentioned previously, the noise imprint is clearly stronger for the first frequency bands.

The initial model provided by Chevron is presented in Figure 4.15a. This is a 1D layered model with no horizontal variations except for the water layer on top for which the correct bathymetry has been incorporated. The P-wave velocity in the water layer is set to 1500 m.s^{-1}. The initial model incorporates an important feature: a low velocity layer is located between the depth $z = 2.3$ km and $z = 3$ km. This velocity inversion and the relatively short available offsets (only 8 km) prevent diving waves from sampling the deepest part of the model. This makes the benchmark data challenging as only reflection information is available for constraining the deep part of the model.

The results obtained after inverting the data up to 4 Hz (frequency band 1), 10 Hz (frequency band 8), 16 Hz (frequency band 12) and 25 Hz (frequency band 15) are presented in Figure 4.15b,c,d,e. Three shallow low velocity anomalies are recovered at approximately 500 m depth and at the lateral positions $x = 12$ km, $x = 18$ km and $x = 30$ km. An additional small scale low velocity anomaly appears at $x = 14.75$ km and $z = 1$ km in the highest resolution estimation. The original layered structure of the initial model is tilted in the final estimation. The upper (faster) layers bend downward (from left to right), while the low velocity layer at depth $z = 2.5$ km bends upward. Three high velocity anomalies are also recovered on top of the layer above the low velocity layer, at depth 1.8 km and lateral positions $x = 8$ km, $x = 19$ km, $x = 22$ km. The deeper part of the model, below 3 km depth, seems less well reconstructed, as it could be expected from the lack of illumination of this zone. However, a curved interface seems to be properly recovered at a depth between 4.5 and 5 km. A flat reflector is also clearly visible at the bottom of the model, at depth $z = 5.8$ km.

As the exact model is not known, it is important to perform quality controls of the computed P-wave velocity estimation. A first indication of the result quality is its similarity with results presented by other teams from the industry and the academy who have worked on the same dataset.

A standard quality control relies on the comparison between synthetic and observed data. To this end, a synthetic shot-gather in the model estimated at 25 Hz is computed and compared to the corresponding benchmark shot-gather in Figure 4.16. The similarity between the two gathers is important. The kinematic of the diving waves is correctly predicted. Most of the reflected events are in phase. Destructive interference due to free surface effects are also correctly recovered. A slight time-shift can however be observed for the long-offsets diving waves. This time-shift is not in the cycle skipping regime. A similar phenomenon is observed in Operto et al. (2015b) where FWI is applied to invert the 3D Valhall data. As mentioned in this study, this time-shift may be due to the accumulation of error with propagating time or an increasing kinematic inconsistency with large scattering angles.

Another quality control consists in comparing the reconstructed P-wave velocity model with *in situ* measurements in some wells. To this end, Chevron has provided a vertical well log of the exact P-wave velocity model at $x = 39375$ m, at a depth between 1000 m and 2450 m. The corresponding log is extracted from the final estimation obtained at 25 Hz maximum frequency and compared to this log

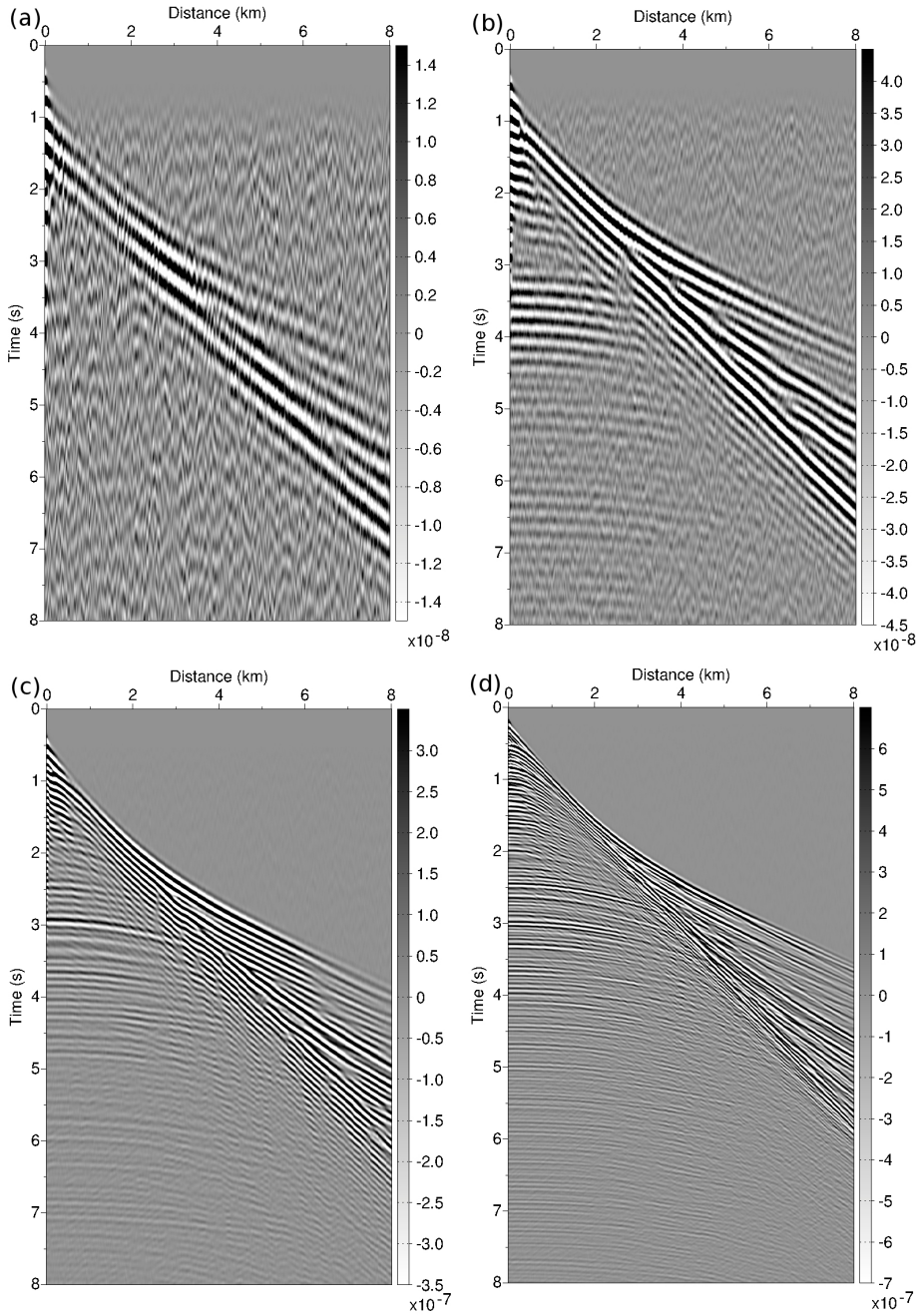

Figure 4.14 – Chevron 2014 dataset. Common shot-gather for the source situated at $x = 0$ km for the frequency bands 1 (a), 5 (b), 10 (c), and 15 (d).

in Figure 4.17. This provides another criterion to assess the quality of the estimation. As can be seen in Figure 4.17, the agreement between the exact and estimated logs is excellent. However, only the shallowest part of the model is constrained here. A deeper exact log would be interesting to have quality control on the deeper part of the model, which is more challenging to recover in this configuration.

To emphasize the benefits provided by using the optimal transport distance, the same frequency continuation workflow is applied to the Chevron 2014 benchmark dataset, with a FWI algorithm based

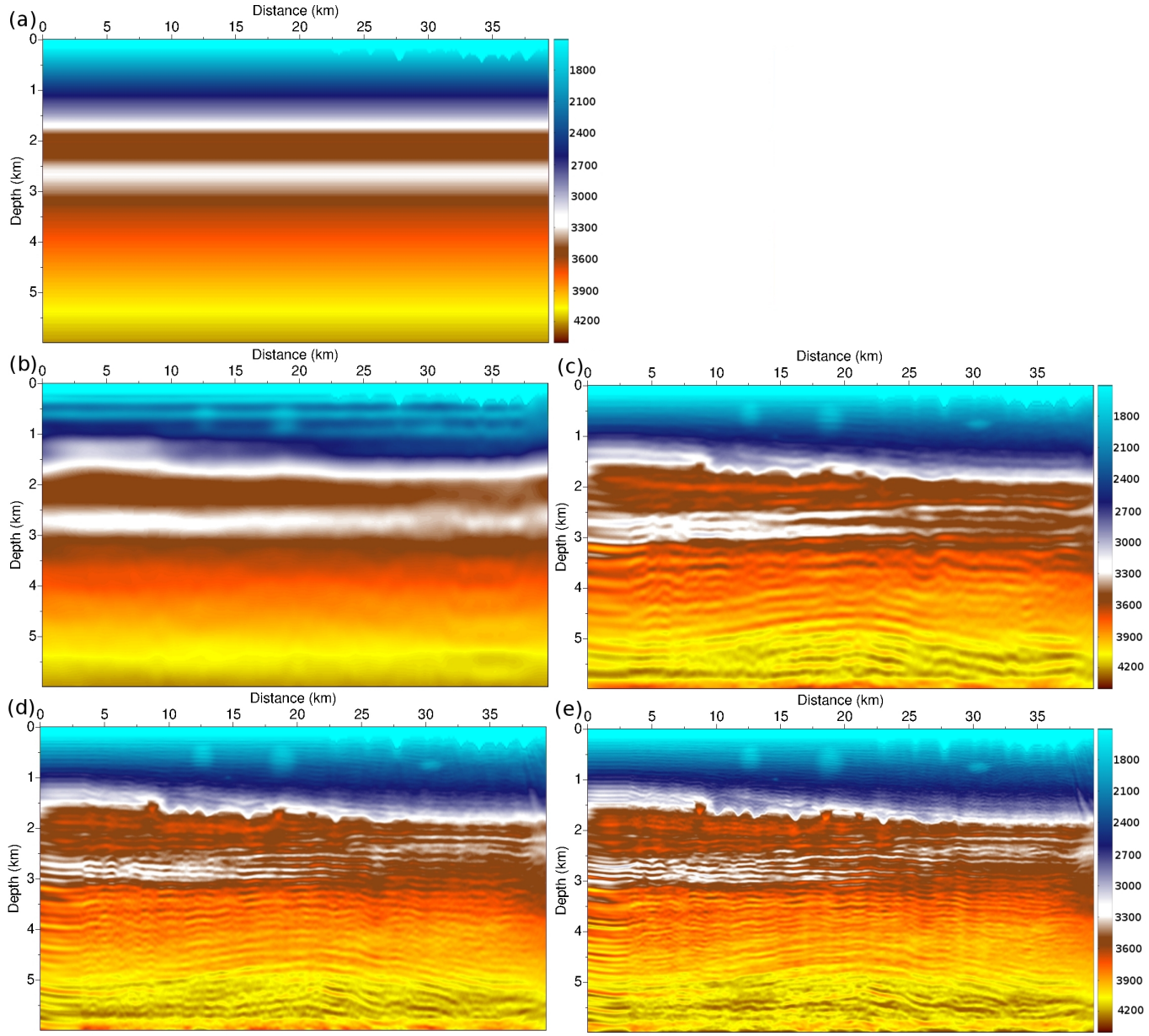

Figure 4.15 – Chevron 2014 starting P-wave velocity model (a). Estimated P-wave velocity model at 4 Hz (b), 10 Hz (c), 16 Hz (d), 25 Hz (e).

on the conventional L^2 distance. The results obtained after the first frequency band and the 8th frequency band are compared to the results obtained when the optimal transport distance is used in Figure 4.18. As can be seen, the L^2 distance based FWI converges to a local minimum. Already after the first frequency band, the shallow part of the P-wave velocity estimation seems incorrect as a strong, flat reflector is introduced at the depth $z = 500$ m. Note that for this simple comparison, no data-windowing strategy is used. As previous experiments in our group indicate, better results using the L^2 distance can be obtained for the reconstruction of the shallow part of the model by designing a hierarchical workflow based on the interpretation of transmitted energy first.

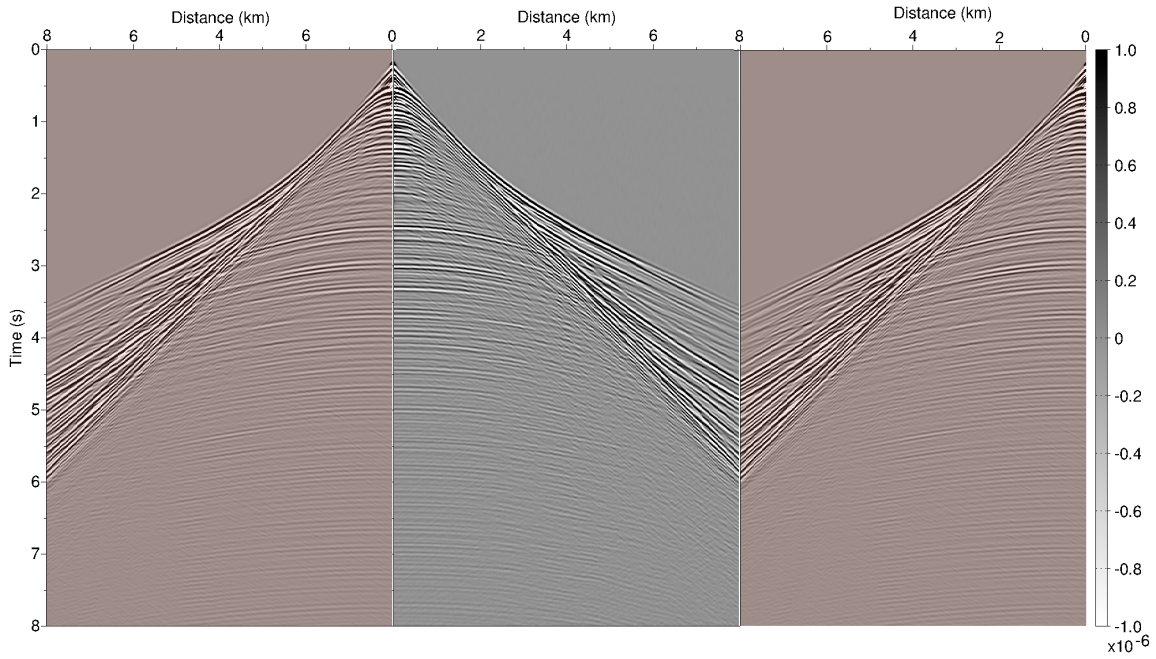

Figure 4.16 – Exact common shot-gather for the left most source at 25 Hz, compared to the corresponding synthetic in the final model at 25 Hz (orange panels). The synthetic data is mirrored and placed on both sides of the real data to better compare the match of the different phases.

4.4.3 3D FWI

In this experiment, the shallow left part of the 3D SEG/EAGE overthrust model is considered (Fig. 4.19). A 250 m deep water layer is added on top. This model covers a surface of 10×10 km^2 and is 2.5 km deep. A fixed spread surface acquisition is used, with $8 \times 8 = 64$ sources (respectively $97 \times 97 = 9409$ receivers) regularly located each 1.2 km (respectively 100 m) in both x and y directions, and at 50 m depth. The synthetic dataset is generated using a Ricker source band-pass filtered between 3 Hz and 7.5 Hz (Fig. 4.20). The spatial discretization leads to a representation of the P-wave velocity model with $201 \times 201 \times 51$ discrete points with a discretization step h equal to 50 m. The time step is chosen equal to 0.004 s to respect the CFL condition. The recording time for one seismogram is fixed to 4 s (1000 discrete time steps). Each seismogram thus corresponds to a data cube of $97 \times 97 \times 1000 \simeq 10^7$ discrete points.

The purpose of this experiment is to focus on cycle skipping problems in a 3D context and compare the results obtained with the L^2 distance and the KR distance. Cycle skipping is mostly observed on diving waves, which sample the shallowest part of the model. For this reason, the initial model is chosen to poorly represent the exact model, especially in its shallow part. Slices of the exact and initial models at constant $y = 5$ km and constant depth $z = 1.5$ km, $z = 2$ km are presented in Figure 4.21. The initial model is an almost constant velocity model around 3000 m.s^{-1}, while the velocity of the exact model reaches 3500 m.s^{-1} already at $z = 1$ km depth. For this reason, the kinematic of the diving waves is not correctly predicted by the initial model (Fig. 4.22). The data is dominated by the direct arrival propagating in the water layer and the strong reflection coming from the interface between the water layer and the see bottom. The relative complexity of the signal is related to the source signature: the Butterworth filters applied to the Ricker wavelet yield a complicated wavelet with a large time support (Fig. 4.20). As the source and the water layer are considered to be known, the initial model correctly reproduces the direct arrivals. However, a time shift of at least 0.3 s can be observed for the diving waves recorded by the farthest receivers. Conventional FWI using the L^2 distance is thus likely to produce

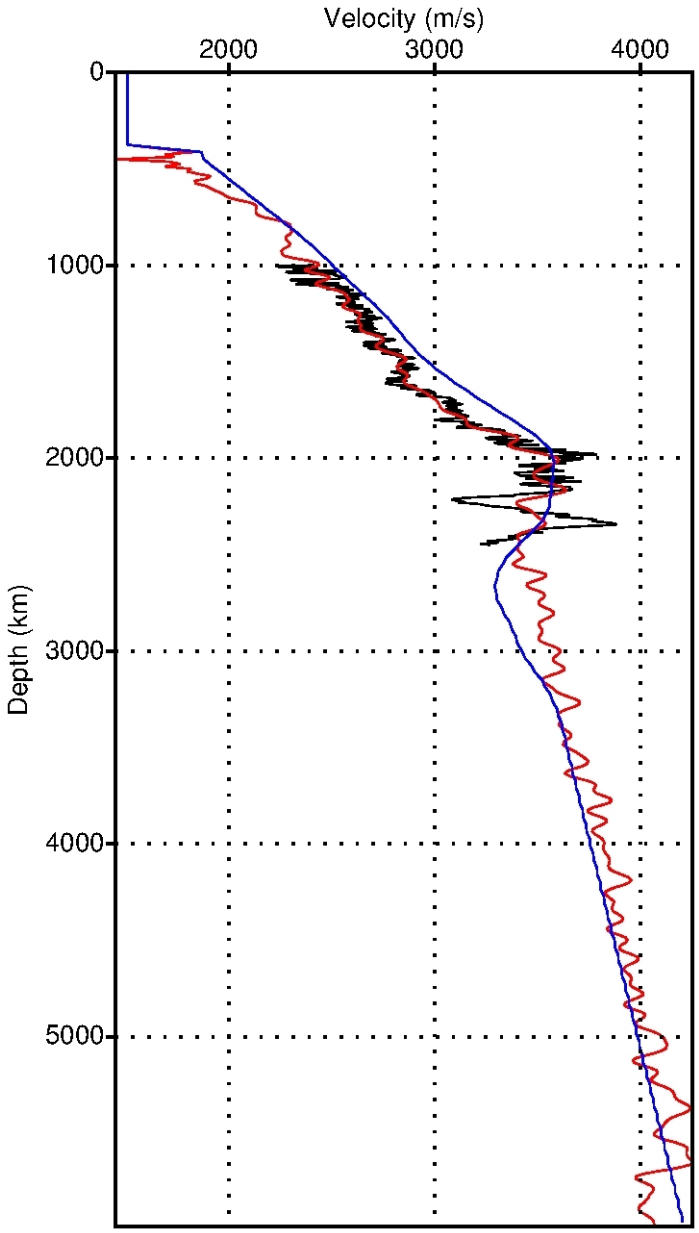

Figure 4.17 – Vertical P-wave velocity log taken at $x = 39,375$ km. Initial model (blue), exact model (black), estimation at 25 Hz (red).

inaccurate results in this configuration.

In Figure 4.23, the L^2 and KR residuals in the initial model are presented. The three panels correspond to slices in the residual volume at constant $x = 5$ km, constant $y = 5$ km, and constant $t = 3$ s., similarly as for the data in Figure 4.22. The energy of the seismic events in the KR residuals is balanced so that the amplitude of each event is comparable, while the L^2 residuals are dominated by short-to-intermediate offset missing events. Interestingly, this balance can be observed in the three panels, which testifies that the solution of the optimal transport problem is performed in the 3D volume without privileging one dimension over the two others.

The results are presented in Figure 4.24. Obvious signs of cycle skipping are visible in the estimation

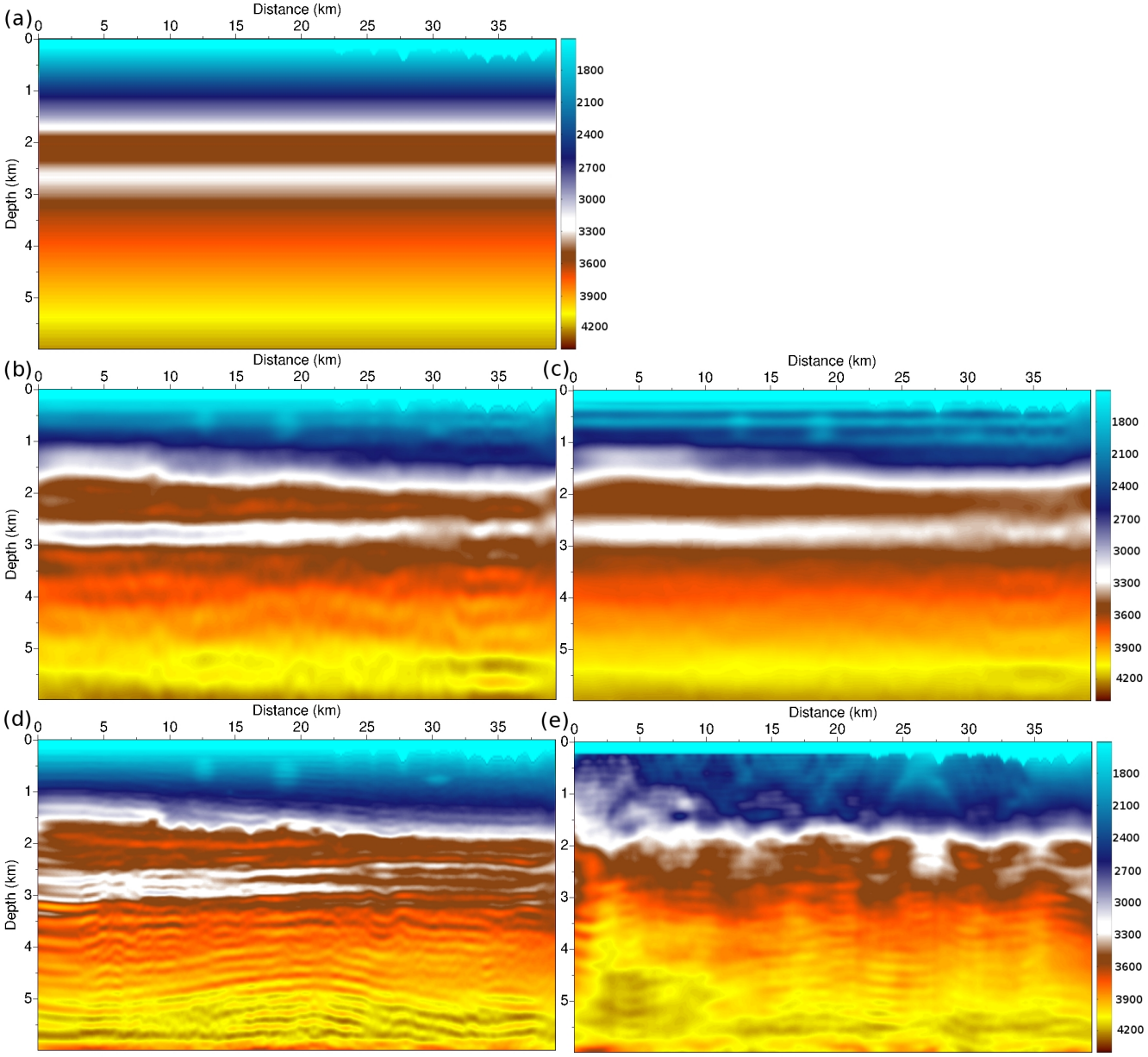

Figure 4.18 – Chevron 2014 starting P-wave velocity model (a). Estimated P-wave velocity model at 4 Hz with the optimal transport distance (b), with the L^2 distance (c). Estimated P-wave velocity model at 10 Hz with the optimal transport distance (d), with the L^2 distance (e).

obtained with the L^2 distance. In the constant y sections (Fig. 4.24a), low velocity artifacts can be observed at 1 km depth, in zones where the velocity update should be positive. Conversely, the KR distance provides a more reliable result in the shallow part of the model, until $z > 1.5$ km (Fig. 4.24b). This difference between the L^2 and KR distance is emphasized by the constant z cross-sections presented in Figure 4.24c-f. At $z = 1.5$ km, the cross-section of the KR estimation presents the main structures of the exact model (Fig. 4.24d). Conversely, the L^2 estimation does not exhibit these structures and present low velocity artifacts caused by cycle skipping (Fig. 4.24c). At $z = 2$ km, the KR estimation still provides some relevant information on the exact model, for instance in the zone 6 km $< y <$ 8 km,

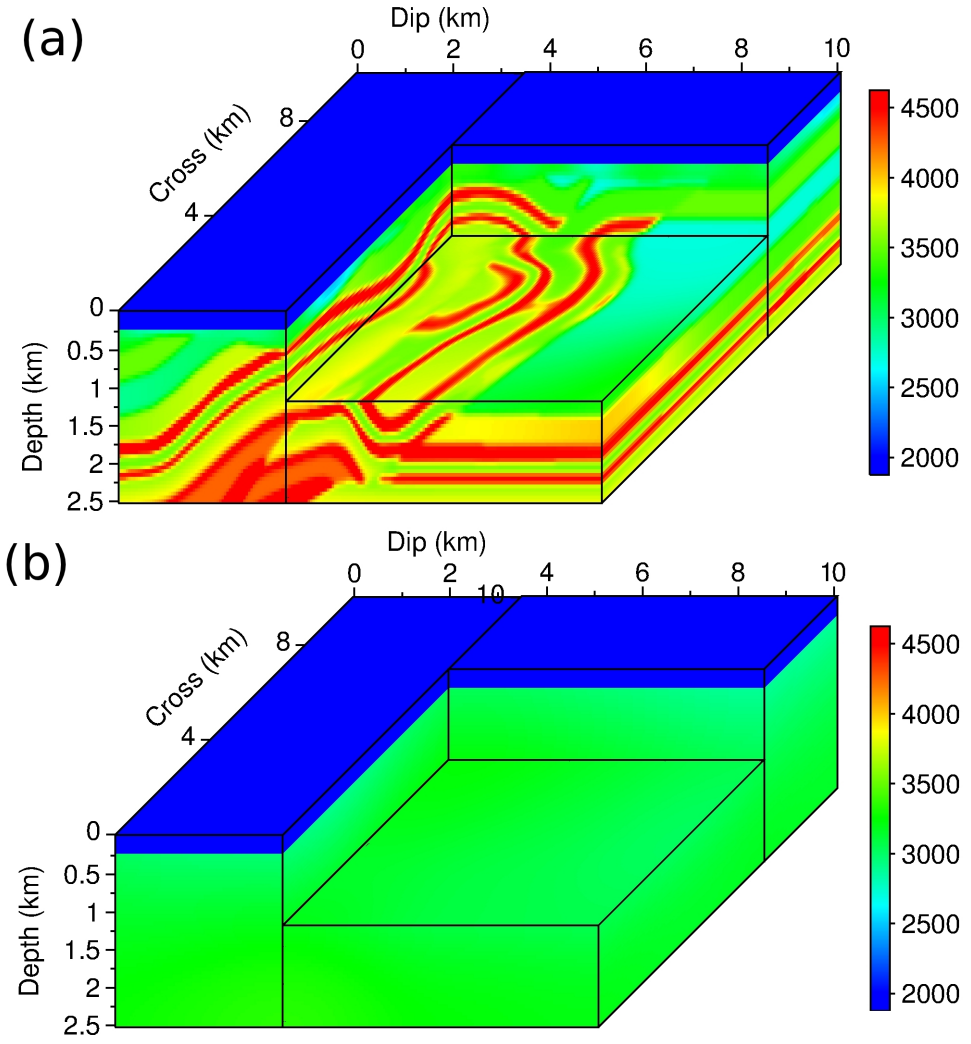

Figure 4.19 – Exact (a) and initial (b) models used for the 3D SEG/EAGE overthrust case study.

2 km $< x <$ 8 km (Fig. 4.24f). Conversely, the L^2 estimation at this depth is completely cycle skipped exhibiting low velocity structures at the location where high velocity updates would be expected (Fig. 4.24g).

An additional illustration of the cycle skipping in the shallow part of the model using the L^2 distance is provided in Figure 4.25. The slices at constant $y = 5$ km of the estimated models are compared with the exact one and the zone below $z = 1.25$ km is shaded. As can be seen, the curved reflector at 1 km depth is replaced with a low velocity anomaly in the L^2 estimation, while its structure is arising in the KR estimation.

An analysis of the residuals in the final estimations is provided in Figure 4.26. As can be seen in constant x and constant y panels (Fig. 4.26b,e), the residuals corresponding to the diving waves (arrival between $t = 2$ s and $t = 2.5$ for farthest offset receivers) are strong in the L^2 final model. Comparatively, these residuals are strongly attenuated in the KR final model (Fig. 4.26c,f). This observation is confirmed by the residual panel at constant $t = 3$ s, where the fringes associated to mismatched event are considerably reduced for receivers between $x = 2$km, $y = 2$ km and $x = 8$ km, $y = 8$ km (Fig. 4.26h,i).

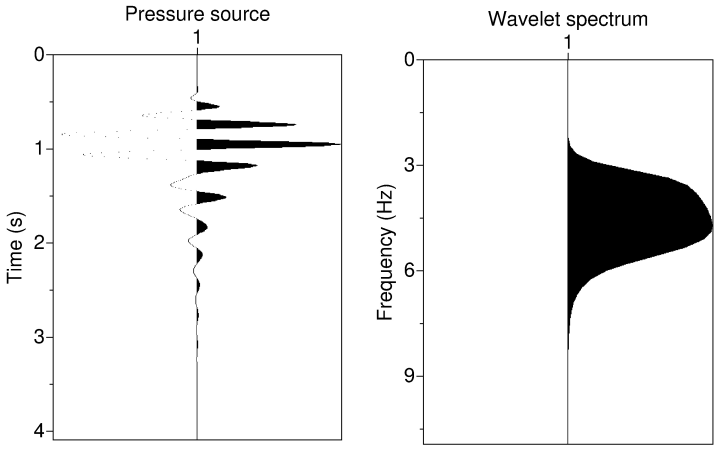

Figure 4.20 – Source wavelet profile and spectrum for the 3D overthrust experiment.

4.5 Short conclusion

This work exhibits promising properties of the optimal transport distance used as a measure of the discrepancy between seismic datasets. The main difficulty to overcome is related to the oscillatory nature of the seismic signal which leads to the comparison of signed measures, for which no definition of optimal transport exists. The second difficulty is related to the size of the data sets for realistic applications, which requires efficient approximation algorithms to be implemented.

A first implementation based on the dual form of the Wasserstein-1 distance and a ℓ_1 ground distance has been presented here. This dual problem provides a natural extension of the optimal transport framework to the comparison of signed measures through its relation with the Kantorovich-Rubinstein norm in the space of Radon's measures. The choice of the ℓ_1 ground distance yields a natural simplification of the underlying linear programming to be solved. Reformulating the problem as a convex non smooth problem, an efficient iterative scheme has been designed, where the computational complexity of each iteration is at most quasi-linear with respect to the number of discrete data.

Promising results are obtained, as the new misfit function appears to be less sensitive to the choice of the initial model, the FWI scheme being able to produce meaningful subsurface properties estimation while a standard least-squares approach fails. However, the simple 1D test performed on shifted Ricker signal clearly illustrates the non convexity with respect to the time shifts of the distance which is employed here. This loss of convexity is due to the use of the dual version of the Wasserstein-1 distance to compare signed quantities. The convexity is ensured only when the positivity assumption is valid.

This leaves the question opened how to better/more appropriately use the optimal transport to compare seismic data. Is it possible to find extension of the optimal transport distance which remain convex with respect to shifts? Is it possible to find nonlinear transformations of the data into positive quantities that would allow to use conventional optimal transport distances while preserving all the information within the signal? These questions are developed in more details in my research project.

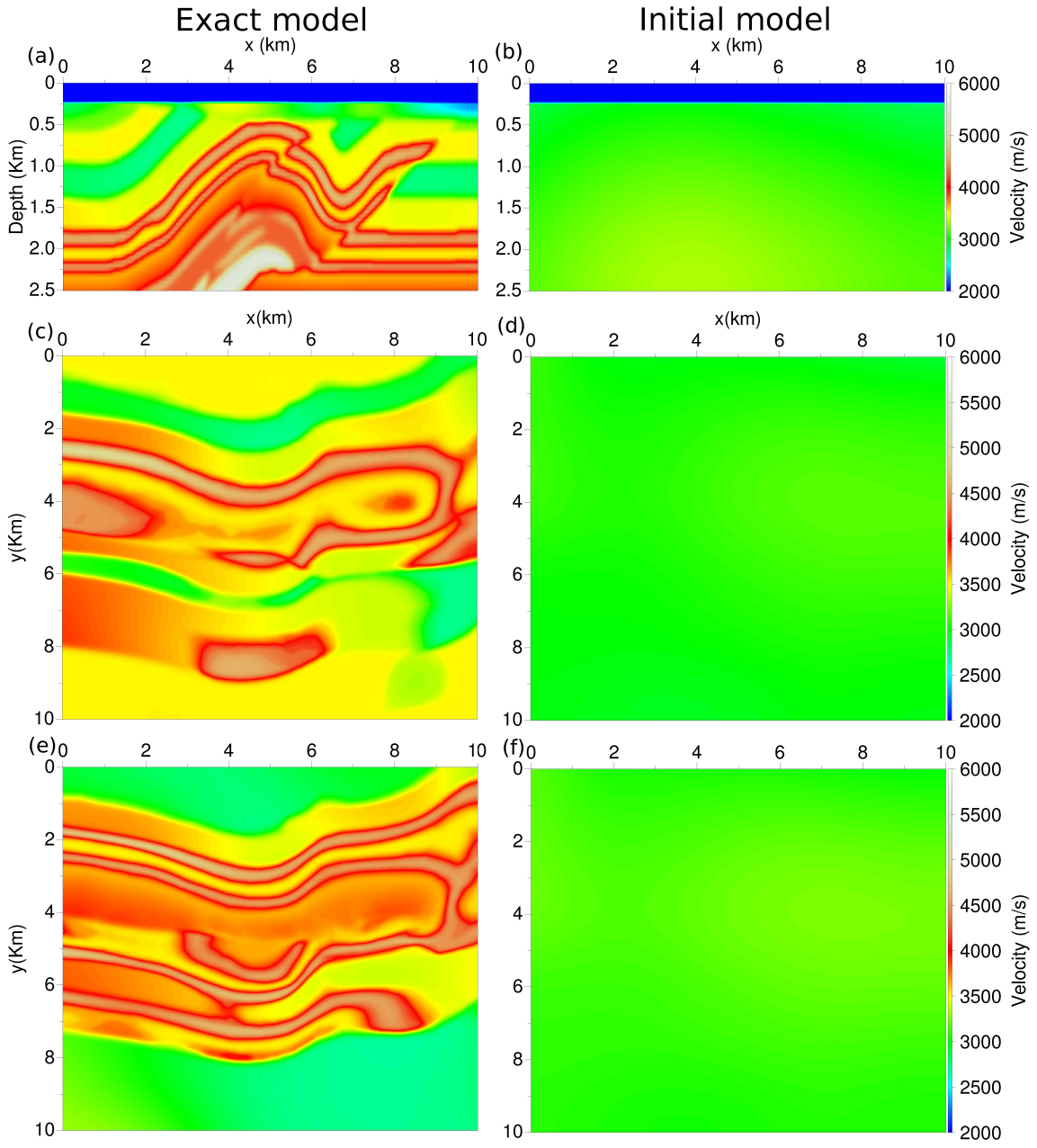

Figure 4.21 – Exact and initial model overthrust model cross-sections. Cross-section at constant $y = 5$ km for the exact model (a), initial model (b). Cross-section at constant $z = 1.5$ km for the exact model (c), initial model (d). Cross-section at constant $z = 2$ km for the exact model (e), initial model (f).

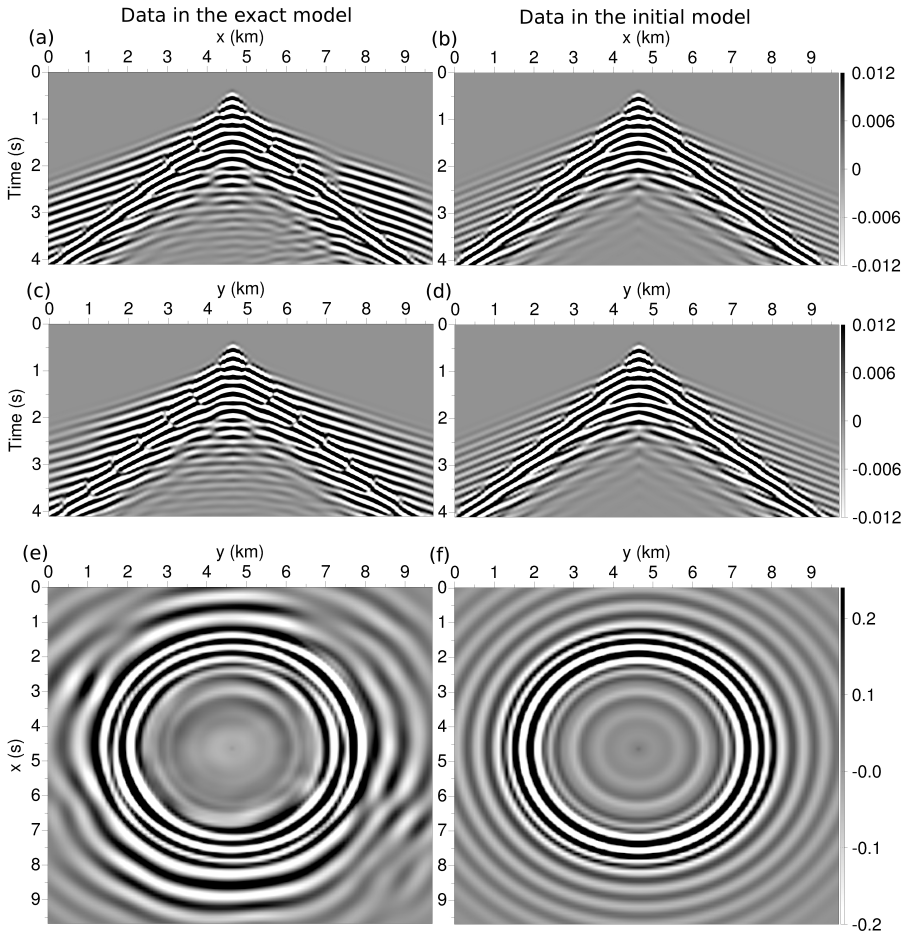

Figure 4.22 – Seismograms in the exact (left column) and initial (right column) models. Seismogram cross-section at constant $y = 5$ km in the exact (a) and initial (b) models. Seismogram cross-section at constant $x = 5$ km in the exact (a) and initial (b) models. Seismogram cross-section at constant $t = 3$ s in the exact (e) and initial (f) models.

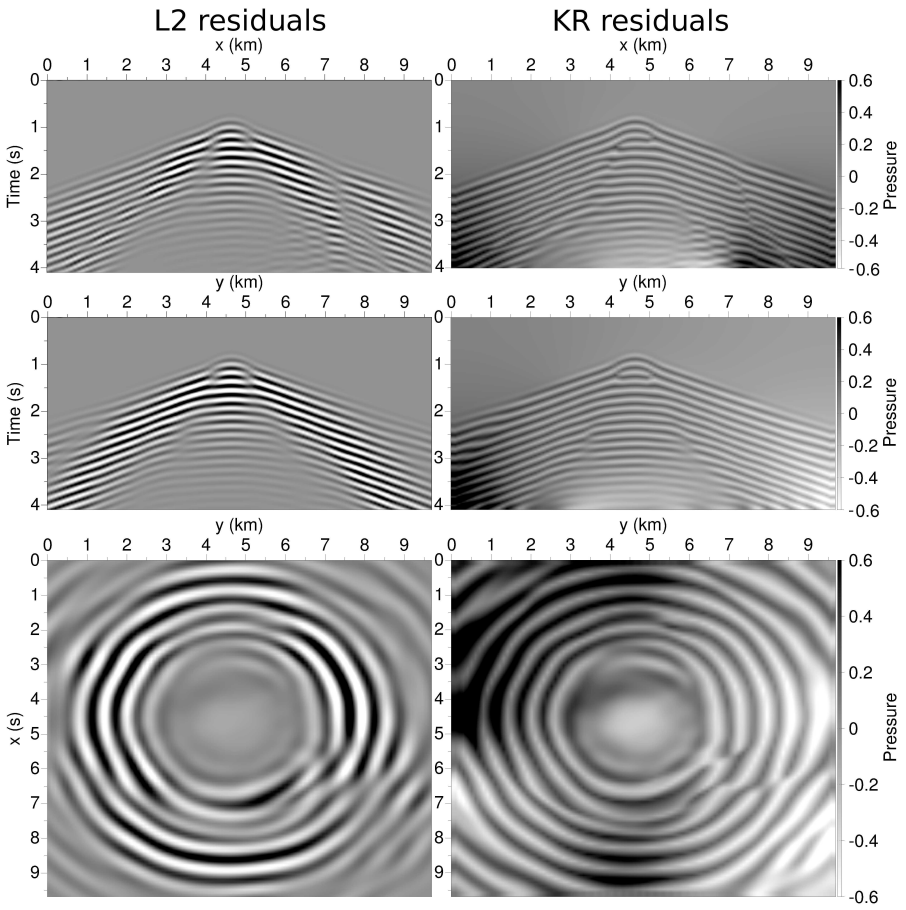

Figure 4.23 – Residuals in the initial model for the L^2 distance (first row), for the KR distance (second row). Cross-section for constant $y = 5$ km (a,b). Cross-section for constant $x = 5$ km (c,d). Cross-section for constant $t = 3$ s (e,f).

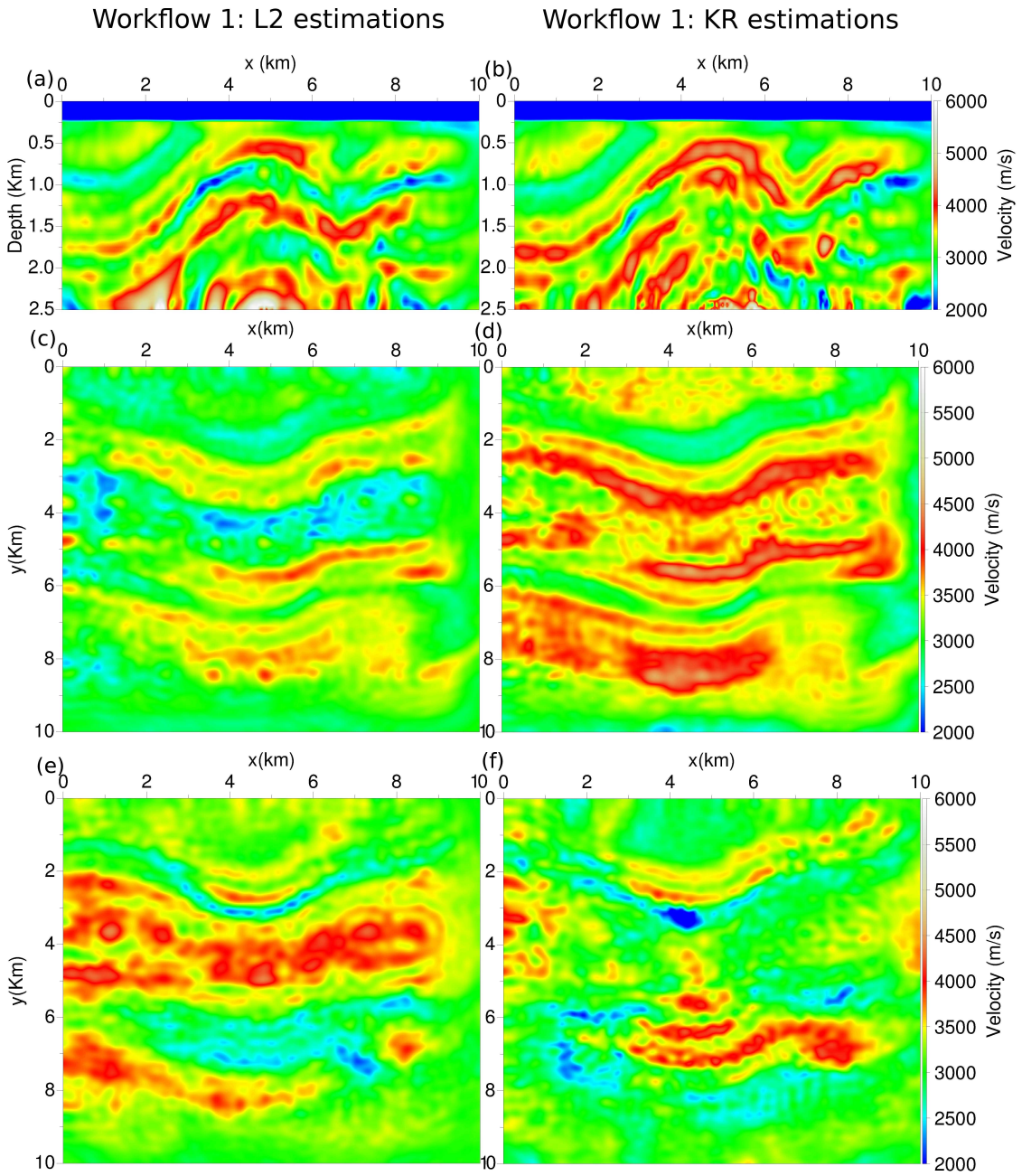

Figure 4.24 – Cross-sections of the L^2 and KR estimations. Cross-section at constant $y = 5$ km for the L^2 estimation (a), KR estimation (b). Cross-section at constant $z = 1.5$ km for the L^2 estimation (c), KR estimation (d). Cross-section at constant $z = 2$ km for the L^2 estimation(e), KR estimation (f).

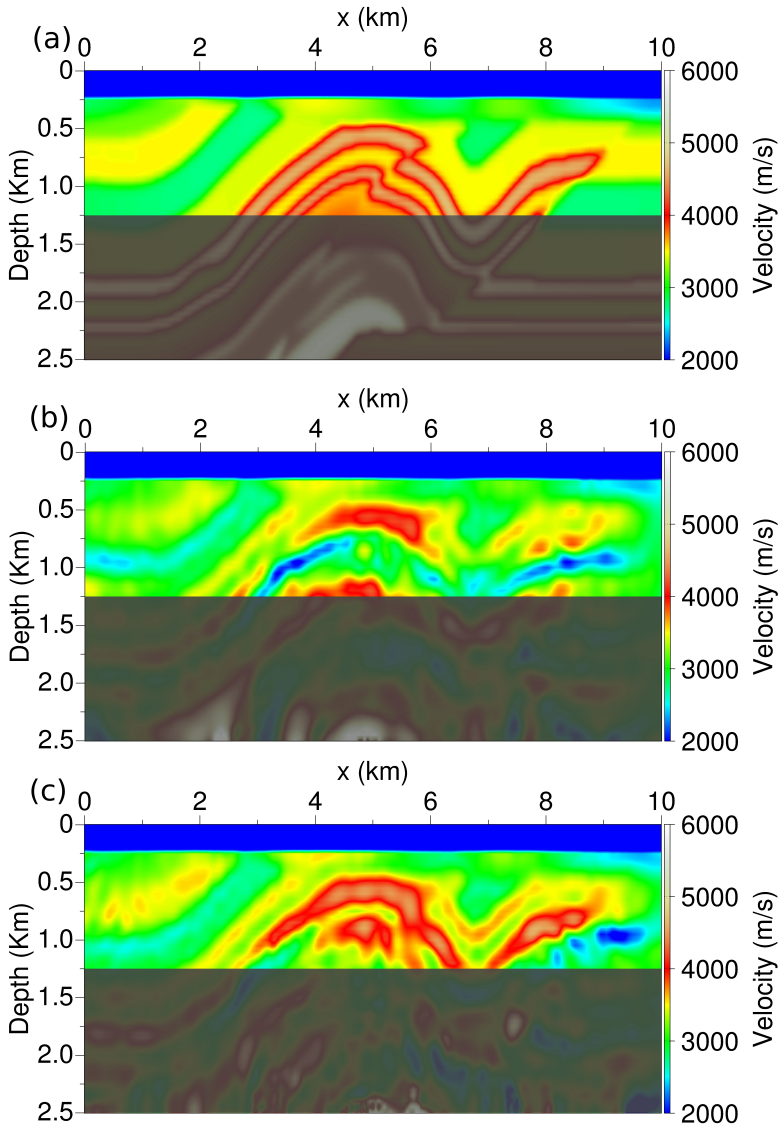

Figure 4.25 – Focus on the shallow reconstruction. Constant $y = 5$ km cross-section for the exact model (a), L^2 estimation (b), KR estimation (c).

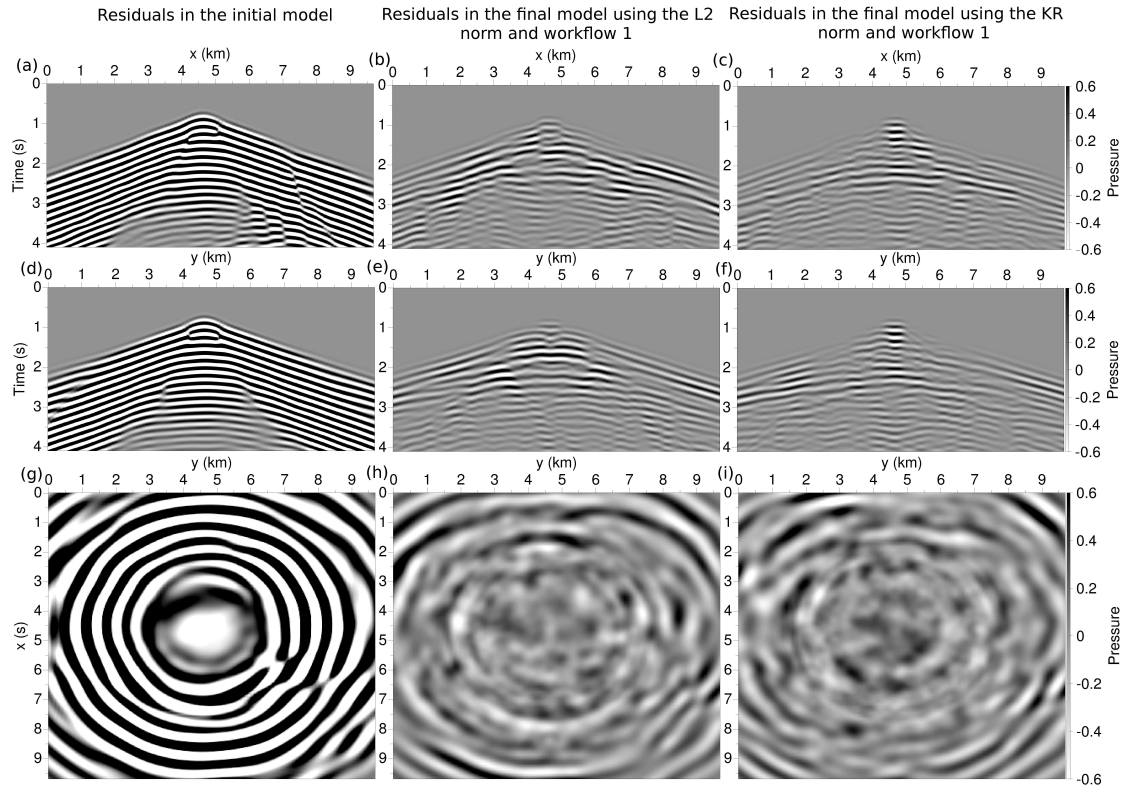

Figure 4.26 – L^2 residuals in the initial model (left column), in the final model obtained using the L^2 (middle column) and KR (right column) distances. Cross-section for constant $y = 5$ km in the initial model (a), in the final model obtained using the L^2 (b) and KR (c) distances. Cross-section for constant $x = 5$ km in the initial model (d), in the final model obtained using the L^2 (e) and KR (f) distances. Cross-section for constant $t = 3$ s in the initial model (g), in the final model obtained using the L^2 (h) and KR (i) distances.

5 Other contributions

In this Section, I summarize my other scientific contributions in the field of seismic imaging with FWI.

5.1 Iterative solvers for frequency-domain visco-elastic wave modeling

Frequency-domain full waveform inversion is an appealing strategy when the redundancy of the data allows to extract few discrete frequencies to invert rather than the whole time series. This redundancy is met at the exploration scale, when dense acquisition systems involving large number of sources and receivers can be used. The extraction of few discrete frequencies leads to significant computation cost reductions.

However, the dynamic wave propagation problem is transformed into a time-harmonic problem generalizing the Helmholtz equation, which requires the solution of a poorly conditioned linear system (Ernst & Gander, 2012). In addition, this linear system has to be solved for a large number of right hand sides, corresponding to the number of seismic sources used during the acquisition.

State-of-the-art techniques consist in factorizing this linear system into a product of lower and upper triangular matrices through a Gaussian elimination approach (LU decomposition). Once factorized, the system can be efficiently solved through forward and backward substitution for any right hand side. This is the direct solver approach.

This strategy has a drawback: the memory consumption. Starting from a banded sparse matrix, its LU factorization yields triangular matrices which are dense on the bandwidth of the initial matrix. Using specifically designed compact stencils (Operto et al., 2014b) and a specialized parallel direct solver (MUMPS team, 2015) based on a multi-frontal approach (Duff & Reid, 1983; Amestoy et al., 2000) and a block low rank reduction (Amestoy et al., 2015c, 2016b), we have been able to push the limit of this strategy to the inversion of a moderate target at the oil exploration scale (Valhall field) in the 3D visco-acoustic approximation (propagation of approximately 100 wavelengths in each direction) (Amestoy et al., 2015b, 2016b,a).

Working in the frequency-domain 3D visco-elastic approximation requires the solution of one to two order larger linear system. The use of direct solvers thus seems still out of reach for these problems, even using modern computational architectures. This is the reason why we are still interested in the potential of iterative solvers to solve these large and ill conditioned linear systems.

The state-of-the-art techniques for the solution of the frequency-domain wave equation using iterative solvers are mainly dedicated to the solution of the acoustic equations. We thus have developed a special interest to what is possible to perform with iterative solvers for frequency-domain elastic wave equations.

The first step has been the identification of the CARP-CG solver as a good candidate for the solution of the corresponding linear systems (Gordon & Gordon, 2010; van Leeuwen et al., 2012). This solver derives from the Kaczmarz technique (Kaczmarz, 1937), one of the oldest algorithm for the solution of linear systems. It can be seen as a conjugate gradient acceleration of the Kaczmarz strategy applied with double sweeps to the original linear system. The main virtue of the CARP-CG strategy is its robustness. We have emphasized this property for the solution of 2D and 3D frequency-domain elastic wave equation in realistic media, focusing on the impact of the complexity of the physics (free surface condition and propagation of surface waves, high Poisson' ratio representing sharp contrasts between S-wave velocities and P-wave velocities) on the convergence of the solver. In all cases, CARP-CG appears as more robust than its alternatives, GMRES, BiCGSTAB, or conjugate gradient applied on normal equations. This work has been performed during the 2 years stay of Y. Li as an exchange PhD student, and has been published in Li et al. (2014a,b, 2016, 2015b).

The second step of this work has consisted in designing a preconditioning approach compatible with the CARP-CG strategy. This has been the main objective of the PhD work of O. Hamitou (Hamitou, 2016). We have been able to generalize an approach based on the approximation of the wave propagation operator through a damped version of this same operator. A sparse preconditioner for this damped operator can be efficiently computed. It is used as a preconditioner for the linear system before applying the CARP-CG strategy. We have been able to show that this strategy exhibits good acceleration of the convergence rate for the computation of the solution in 2D and 3D realistic visco-

elastic media. This work has been presented at the 2015 SEG conference (Hamitou et al., 2015) and a publication is now in preparation.

5.2 Asymptotic approximation of the Hessian operator

In the framework of the studies led on the approximation of the Hessian operator, I have also investigated how asymptotic approximation of the wave equation can lead to design approximations of this operator and its inverse. This work has been presented in (Métivier et al., 2015c). The starting point is the linearized inverse problem approach promoted in particular by Beylkin (1985). This linearized inverse problem can be written as the reconstruction of the subsurface parameter m as $m_0 + \delta m$ such that

$$J(m_0)\delta m = d_{obs} - d_{cal}(m_0). \tag{5.1}$$

where $J(m_0)$ is the Jacobian operator

$$J(m_0) = \frac{\partial d_{cal}}{\partial_m} \tag{5.2}$$

The strategy proposed by Beylkin for solving this problem includes three steps.

- Use the first-order Born approximation to derive an analytic expression for $J(m_0)$ depending on the source and receiver Green functions $G^0(x, \omega, x_s)$ and $G^0(x, \omega, x_r)$ representing the wavefield recorded at point x from an impulse source located at x_s (respectively x_r).
- Compute an asymptotic approximation of $J(m_0)$, denoted by $\mathscr{J}(m_0)$, through the asymptotic approximation of the Green functions $G^0(x, \omega, x_s)$ and $G^0(x, \omega, x_r)$.
- Compute an approximate left inverse of $\mathscr{J}(m_0)$, denoted by $\mathscr{B}(m_0)$. The latter operator is roughly equivalent to a preserved-amplitude migration operator.

Based on this approach we have proposed to consider the misfit function $g(m)$ defined as

$$g(m) = \frac{1}{2} \left\| \mathscr{B}(m_0) \left(d_{cal}(m) - d_{obs} \right) \right\|^2_{\mathcal{M}}, \tag{5.3}$$

as an extension of the work of Jin (1992). The expression for the gradient of $g(m)$ is

$$\nabla g(m) = J^*(m) \mathscr{B}(m_0)^* \mathscr{B}(m_0) \left(d_{cal}(m) - d_{obs} \right). \tag{5.4}$$

Compared to the gradient of the standard misfit function $f(m)$ (equation (3.23)), the difference induced by the integration of the migration operator relies on the multiplication of the residuals by $\mathscr{B}(m_0)$ and its adjoint $\mathscr{B}(m_0)^*$. These two multiplications can be interpreted as migration (from data to image domain) and demigration (from image to data domain) acting on the residuals. This mapping from the data space to the model space is performed following the geodesics defined by rays connecting sources and receivers to the scattering point of the investigated domain.

The Gauss-Newton approximation of the Hessian of $g(m)$, denoted by $H^g_{GN}(m)$, is

$$H^g_{GN}(m) = J^*(m) \mathscr{B}(m_0)^* \mathscr{B}(m_0) J(m) = \left(\mathscr{B}(m_0) J(m) \right)^* \left(\mathscr{B}(m_0) J(m) \right). \tag{5.5}$$

From equation (5.5), it is clear that the operator $H^g_{GN}(m)$ is the normal operator associated with the operator $\mathscr{B}(m_0)J(m)$. In the asymptotic approximation,

$$\mathscr{B}(m_0) J(m) \simeq \mathscr{B}(m_0) \mathscr{J}(m). \tag{5.6}$$

In addition at the first iteration

$$\mathscr{B}(m_0) \mathscr{J}(m) = \mathscr{B}(m_0) \mathscr{J}(m_0). \tag{5.7}$$

Following the work of (Beylkin, 1985), the operator $\mathscr{B}(m_0)\mathscr{J}(m_0)$ is proved to be a pseudo-differential operator which tends towards the identity as the illumination of the subsurface increases. As a pseudo-differential operator, this property holds for its adjoint. In addition, the composition of two pseudo-differential operators which tends to the identity tends also towards the identity. Therefore, $H^g_{GN}(m_0)$ can be approximated as

$$H^g_{GN}(m_0) \simeq \left(\mathscr{B}(m_0)\mathscr{J}(m_0)\right)^* \left(\mathscr{B}(m_0)\mathscr{J}(m_0)\right) \simeq I + \sum_i T_i, \tag{5.8}$$

where the operators T_i belong to classes of increasingly smooth operators. The introduction of the integral operator $\mathscr{B}(m_0)$ in the misfit function thus yields a better conditioned inverse problem, as the Hessian operator associated with the new misfit function should be closer to the identity.

Because the modified misfit function $g(m)$ possesses a Hessian operator which tends asymptotically towards the identity operator, the trade-offs between discrete parameters are reduced in the inversion scheme. Results on 2D synthetic case studies demonstrate the fast convergence of this inversion method in a migration regime. From an accurate estimation of the initial velocity, three and five iterations only are required to generate high resolution P-wave velocity estimations models on the Marmousi 2 and synthetic Valhall case studies.

5.3 Source encoding and second-order optimization

The PhD of C. Castellanos was dedicated to the investigation of source encoding technique in FWI. Source encoding is a strategy used to reduce the computational cost of FWI. When dense acquisition systems are used, involving large number of seismic sources (up to thousands), each iteration of FWI requires the solution of the same number of PDE. The solution of these PDE is embarrassingly parallel, however, performing simultaneously thousand of 3D acoustic or elastic simulations might require a computation power not always available (or one would like to distribute the computational tasks in a different way to increase the efficiency of the overall inversion).

While it is straightforward to decimate the data volume to reduce the computational cost, either by extracting few discrete frequencies, as in the frequency-domain approach, or by sub-sampling the number of seismic sources, another possibility to decrease this computational cost is to take advantage of the linearity of the wave propagation problem with respect to the source and to use random linear combinations of the sources at each inversion iteration rather than all the sources treated independently. Several choices for these linear combinations exist: deterministic stacking of all sources (Gao et al., 2010; Habashy et al., 2011), random stacking of all sources (Romero et al., 2000; Haber et al., 2012), or random linear combinations of smaller ensemble of sources (Krebs et al., 2009; Schuster et al., 2011; Ben Hadj Ali et al., 2011; Baumstein et al., 2011). The latter have been introduced as source encoding strategies within the FWI community. One can prove that if the random linear combination satisfies specific stochastic properties, provided a sufficient number of inversion iterations is performed, the "noise" artificially introduced by the linear combination of the data can be arbitrarily decrease below a given level.

A trade-off is therefore required to be found between the reduction of the computational cost associated with the decrease of the number of PDEs to be solved and the increase of computational cost due to the slower convergence of the minimization method. Based on this, the use of second-order l-BFGS and truncated Newton strategies has been investigated to evaluate if a better trade-off can be obtained with these methods compared to first-order methods such as steepest descent or nonlinear conjugate gradient (Castellanos et al., 2015). The conclusion show that l-BFGS with adequate restarting strategies is probably the best optimization algorithm in this configuration.

5.4 FWI of GPR data

Ground-penetrating radar (GPR) is a subsurface prospecting technique used in civil engineering based on the propagation of electromagnetic waves, dedicated to the characterization of the near surface, from few hundred meters to few centimeters depth. GPR imaging is therefore a technique widely

used in geology, glaciology, hydrology, and civil engineering, with applications such as: active faults imaging (McClymont et al., 2008), rock fractures characterization (Deparis & Garambois, 2009), ice sheet thickness measurement (Walford, 1964; Bailey et al., 1964; Bentley, 1964), water content evaluation, salinity and/or pollution contamination of the soil (Lambot et al., 2006; Deeds & Bradford, 2002), anthropic structures detection and road pavement inspection (Evans et al., 2008) or land mine detections (Daniels, 2004).

GPR data is sensitive to the electromagnetic properties of the subsurface, mainly the permittivity, the conductivity and the magnetic permeability. Imaging from GPR data shares many similarities with seismic imaging techniques. While the first applications are mainly based on common offset measurement and standard imaging techniques based on tomography or migration, multi-offset measurements have been deployed in the last decade, yielding an interest for this community towards using FWI techniques, with the objective to quantitatively estimate the electromagnetic properties of the subsurface with high resolution (Ernst et al., 2007; Meles et al., 2010; Klotzsche et al., 2013).

The first applications have been proposed in cross-hole configurations, where sources and receivers are deployed in boreholes. In this configuration, the data is dominated by transmitted energy, which is favorable for the use of FWI. The results which are obtained were encouraging and prompted the investigation of FWI applied to surface GPR data. This was the main objective of the PhD of F. Lavoué (Lavoué, 2014).

Based on the analogy between 2D acoustic equations and 2D Maxwell's equation, a 2D frequency-domain FWI code has been designed for the inversion of GPR data and the reconstruction of permittivity, conductivity, and magnetic permeability. In the transverse electric (TE) mode, permittivity is an analogous of the P-wave velocity, conductivity an analogous of a quality factor (attenuation modeling) inducing attenuation effects, and the permeability is an analogous of the density. The GPR data is mainly sensitive to permittivity and conductivity. Compared to seismic data, the imprint of the quality factor cannot be neglected, and a simultaneous reconstruction of both conductivity and permittivity is required, leading naturally to a multi-parameter inversion with a strong trade off between permittivity and conductivity, the radiation patterns being isotropic for both parameters. Another difficulty for GPR applications is related to the strong contrasts in permittivity that can be found in the subsurface, with ratios reaching factors as large as 30, while at the exploration scale a ratio reaching 3 is already considered as challenging (salt structures for instance).

A first strategy based on an ad-hoc scaling of the parameters and the l-BFGS algorithm has been designed based on synthetic experiments (Lavoué et al., 2014), and successfully applied to experimental data of the Fresnel institute (Lavoué et al., 2015).

Elaborating on these first results, a second PhD project has been launched in 2013 with H. Pinard, to assess how to better decouple the permittivity and conductivity parameters, for instance using truncated Newton strategies, and to apply the method to real data. From this work we conclude that the truncated Newton without an adequate preconditioning strategy does not really mitigate the trade-off between parameters, which is consistent with what is observed with seismic data: only the combination of the truncated Newton strategy and a multi-parameter preconditioner provides an efficient decoupling. However, the results obtained on cross-hole GPR data acquired at the LSBB laboratory (Gaffet et al., 2003) are quite encouraging, as a high resolution estimation of the permittivity, consistent with local geological information, was reconstructed.

5.5 Kinematic source inversion

Another topic I have the opportunity to contribute to is related to the application of full waveform inversion procedure to earthquake source reconstruction. These reconstructions are primordial to unveil earthquake rupture. They are used on ground motion predictions, and there results shed light on seismic cycle for better tectonic understanding. They are not yet used for quasi-real-time analysis, and this is the direction towards which the PhD project of H. Sanchez intends to go.

Most of the current techniques for earthquake source imaging are posterior procedures once seismograms are available. Incorporating source parameters estimation into early warning systems would require to update the source build-up while the data is recorded. In order to go towards this dy-

namic estimation, we developed a kinematic source inversion formulated in the time-domain, for which seismograms are linearly related to the slip distribution on the fault through convolutions with Green functions previously estimated and stored. These convolutions are performed in the time-domain as we progressively increase the time window of records at each station specifically. The unknowns are the spatio-temporal slip-rate distribution on the fault to keep the linearity of the forward problem with respect to unknowns, as promoted by Liu & Archuleta (2004) and Fan et al. (2014) among others. Through the spatial extension of the expected rupture zone, we progressively build-up the slip-rate when adding new data by assuming rupture causality.

This formulation is based on the adjoint-state method for efficiency (Plessix, 2006). The inverse problem is non-unique and, in most cases, underdetermined. The key is thus to determined adequate regularization strategies to reduce the null space. We use standard Tikhonov regularization terms for stabilizing the inversion, introducing regularity and prior information on the source. We avoid strategies based on parameter reduction leading to an unwanted non-linear relationship between parameters and seismograms. Satisfactory results have been obtained on a synthetic example proposed by the Source Inversion Validation project (Mai et al., 2016). A real case application on the Kumamoto earthquake in Japan (Asano & Iwata, 2016; Uchide et al., 2016) yields also promising results. Our specific formulation combined with simple prior information, as well as numerical results obtained so far, yields interesting perspectives for a real-time implementation.

5.6 The SEISCOPE optimization toolbox

The idea of designing an open-source optimization toolbox came from the need to compare different optimization schemes, namely steepest descent, nonlinear conjugate gradient, l-BFGS, and truncated Newton methods. When it comes to comparison, the question of the linesearch algorithm is crucial. Therefore, the four optimization schemes implemented in our toolbox use the same linesearch strategy. It is based on a bracketing algorithm and the satisfaction of the Wolfe conditions. This impacted the particular choice of the nonlinear conjugate gradient algorithm we made (Dai & Yuan, 1999), as not all the nonlinear conjugate variant are compatible with the Wolfe conditions (most require the strong Wolfe conditions to be enforced to ensure convergence, for instance the Polak Ribière or the Fletcher-Reeves ones). Also, each method can be combined with any user-defined preconditioner.

From an implementation point of view, we opted for a reverse communication strategy (Dongarra et al., 1995), so as to make easy the coupling with any code where nonlinear optimization problem has to be solved. The principle is the following: the computation of the solution of the optimization problem is performed in a specific routine of the code. This routine is organized as a minimization loop. At each iteration of the loop, the minimization routine chosen by the user is called. This routine communicates what is the quantity required at the current iteration: objective function, gradient, or Hessian-vector product. These quantities are computed by the user in specific routines, external to the minimization loop. The process continues until the convergence is reached. This implementation paradigm yields a complete separation between the code related to the physics of the problem and the code related to the solution of the minimization problem. This ensures a greater versatility as one can easily modify one of these parts while keeping the other unchanged.

This work has been published in the Algorithms & Softwares Section of Geophysics (Métivier & Brossier, 2016a) together with the `FORTRAN 90` source code in which it is implemented.

Chapter 3

Research project

Contents

1 Introduction

Full waveform inversion is dedicated to the quantitative characterization, with high resolution, of the subsurface physical properties. This characterization is performed from local, incomplete measurements of the seismic wavefields, possibly complemented with prior information on these physical properties (upper and lower bound, *in situ* measurement, structural information for instance).

The physics of wave propagation can be represented with reasonable accuracy through the use of partial differential equations. In this representation, the physical properties of the medium in which the waves propagate are spatially dependent coefficients of these equations. This understanding of the physics of wave propagation is the reason why full waveform inversion is designed as a PDE parameter identification problem: the estimation of the physical properties are computed as PDE coefficients yielding the minimum misfit between observed and synthetic measurements. Note that this is not always possible: there are research areas where the physical phenomenon which are considered cannot be simply represented through the use of partial differential equations.

This first step in the formulation of FWI might however already be questionable: how can we relate the estimation of the coefficients of the PDE to the real physical properties we are interested in recovering? A key concept to answer to this question is the concept of homogenization. The information we can recover strongly depends on the way we illuminate our target. Waves propagate with a finite

frequency content, and their sensitivity to heterogeneities with small scales (much smaller than the smallest propagated wavelength) is different from the macro and meso-scale heterogeneities (up to the order of the smallest wavelength). The small heterogeneities are seen as a larger scale, smooth anisotropy of the medium by the waves. The theory of homogenization tells us that it is possible to compute, from a set of different media with different local, small heterogeneities, an "equivalent" homogenized smooth (no small scale heterogeneities) anisotropic media, for which the wave propagation, in the frequency range associated with the signal we use to illuminate our target, is the same. This already tells us that the best estimation we can make of the PDE parameters only carries information about a homogenized version of the real physical parameters we are interested in recovering. It also gives a hint on how ill posed the problem can be, as a wide variety of parameters can yield approximately the same wavefield.

Another fundamental question in the design of the FWI problem relies on the way the data is interpreted. For now on, this interpretation is mostly based on the L^p differences of some observables computed from the data. As such it is based on local, point-to-point comparisons, where the geometrical information carried by the natural organization of the data, depending on the space in which it is represented, is lost. The recent development of new similarity measures between images, in the framework of image processing (some of which being based on the optimal transport theory) might represent an opportunity to significantly improve the solution of inverse problems based on the estimation of PDE parameters, such as FWI. The similarity measures should be able to capture geometrical, global features of the data, allowing to automatically incorporate this information in the inverse problem solution.

Finally, even if the tools which are used to measure the similarity between observed and synthetic data are improved, we know that the solution of such an inverse problem will remain highly non-unique. It is therefore primordial to move from a deterministic point of view to a statistical interpretation of the inverse problem solution. This means that the computation of uncertainty estimation on the solution will become mandatory in the forthcoming years. This is an old dream (Tarantola, 2006), however, both the methodology developed in data assimilation communities and the design of new exascale HPC devices might allow to step forward in this direction.

The research project I intend to lead is built along these lines. The first axis of my project is related to the use of optimal transport distances as a new similarity measurements between datasets. Given the impact the use of these distances already had in image processing, with applications in pattern recognition and learning strategies, their use in the context of geophysical application could be decisive. The two main difficulties to overcome are the extension of these distances to the comparison of signed measures, and the design of approximation strategies making these methods amenable to the comparison of large scale discrete datasets involving millions to billions of discrete parameters.

The second axis of my project is the investigation of methodological developments aiming at reducing the size of the solution space (through homogenization and/or regularization), accessing to the uncertainty of the solution (through data assimilation techniques), and improving our modeling capabilities (through the design of efficient solvers for elastodynamics frequency-domain equations, and the use of particle methods for wave propagation as an open, more speculative project).

The third axis of my project is related to the application of the methodologies we develop, in particular multi-parameter FWI, to real data-sets. Exploration scale targets should be investigated through interactions with the sponsors of the SEISCOPE consortium, as well as larger scale targets such as subduction zones, through a collaboration with the Japan-Agency for Marine-Earth Science (JAMSTEC). A project on the interpretation of passive seismic data through noise correlation techniques combined with FWI should also allow to aim at larger scale crustal targets. This should be performed through a collaboration with the team of M. Campillo in ISTerre. I also intend to pursue the collaboration with S. Garambois on the application of FWI to GPR data, and I would like to promote the use of FWI for ultra-sound imaging, through collaboration within ISTerre on non destructive testing (L. Moreau), and possibly for medical imaging (this is a longer term project). What we have learned within the SEISCOPE project is that application to real data is the best way to promote and reveal the potential of FWI as a versatile and generic high resolution imaging method, and that the problems to overcome when facing the application to real data are always the source for new methodological developments.

The fourth axis of my project is transverse to the three first ones, and relies on the objective to

provide, along with developments and applications, easy-to-use, versatile, open-source implementations, however adapted to large scale computing, of the strategies which are designed, to better benefit the community.

2 Making the most of the data we record: can optimal transport help?

The constant progresses in the design of acquisition systems leads to the collection of increasing amounts of data, in various fields of experimental science. This is particularly true in geophysics: denser and denser acquisition systems are deployed, either at the exploration scale (active seismic) or for lithospheric and global imaging (permanent networks). These data are rich of information, however, in most of the applications, they are carefully pre-processed so that the interpretation focus on few pieces of observables selected before hand. This questions the robustness of the tools which are used to compare these datasets and extract the information from the data.

I think the use of optimal transport distances might significantly improve this aspect of the problem in geosciences, and, to a larger extent, for PDE parameter identification problems. The main reason for that is its ability to perform global comparisons as opposed to the local, sample by sample comparisons performed by L^p distances. This induces that any geometrical coherence in the space where the data is represented is really taken into account when the data is compared through optimal transport distances, while this information is discarded when using L^p distances. In most applications, the first step of data pre-processing consists in finding a suitable representation space in which the physics of the studied phenomenon expresses itself as recognizable patterns. In this respect, optimal transport distances provide a way to automatize the incorporation of an information which is, for now, used only at a human expertise level, for a pre-processing step.

An example is the representation of seismograms as 2D or 3D shot-gathers, where the vertical axis is the time-axis and the horizontal axis correspond to the spatial location of the receivers at the surface. This representation allows to visualize precisely the organization of waves through their propagation, as surface waves, direct waves, diving waves, reflected and converted waves. Currently this information is used only to display the data and pre-process it through the definition of specific filters. However within the inversion, the misfit function is based only on local comparison of observed and predicted data, which does not account for this spatial coherency. Optimal transport offers a way to account for it automatically within the inversion.

The first applications I have lead so far for FWI lead me to the conclusion that the main difficulty to be overcome to apply optimal transport distance in this general framework is the non-positivity of the data. The optimal transport problem is defined for positive measures, and there is no straightforward extension to signed measures which ensures that the properties of the optimal transport problem are preserved.

Several transforms of the data can be applied to the data to recover positive quantities. Engquist & Froese (2014) for instance propose to separate the positive and negative part of the data and transport them separately. Other approaches could consist in rescaling and shifting the data by the opposite of its minimum value (proposed by (Yang et al., 2016), taking the absolute value of the signal, or finally the envelope of the signal, following approaches already used in seismology (Fichtner et al., 2008; Bozdağ et al., 2011). All these strategies are straightforward, however not really satisfactory. Important physical information are lost through these transformations: artificially de-correlating the positive and negative part of the signal is difficult to justify, absolute value and envelope lose the sensitivity of phase. In addition, nothing ensures that the mass conservation is satisfied by the transformed signal. Shifting the data with a given amount of mass also modifies the optimal transport process, as it creates artificial mass everywhere which can be moved locally to match the observed data. The convexity of the resulting distance with shifted patterns is thus lost.

The question how to extend optimal transport distances to signed measures is actually investigated in the work of Mainini (2012) and Ambrosio et al. (2011) from a more theoretical point of view. Given

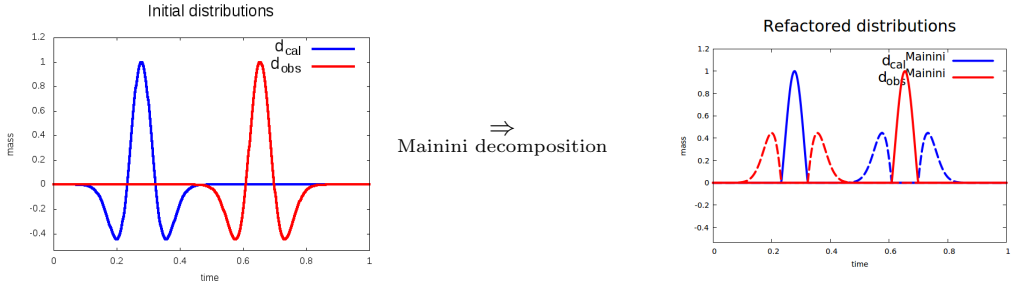

Figure 2.1 – Mainini decomposition.

two signed functions $f(x)$ and $g(x)$ satisfying the mass conservation assumption

$$\int f(x)dx = \int g(x)dx, \tag{2.1}$$

a generalization of the Wasserstein distance is proposed as

$$\widetilde{W}_p(f,g) = W_p(f^+ + g^-, g^+ + f^-), \tag{2.2}$$

where the positive and negative parts of a function h are defined as

$$h^+ = \frac{h+|h|}{2}, \quad h^- = \frac{|h-|h}{2}, \quad h = h^+ - h^-. \tag{2.3}$$

An illustration of such a decomposition is provided in Figure 2 for two time-shifted Ricker signals. Interestingly, provided the mass conservation assumption (2.1) is satisfied, the mass conservation is also satisfied by $f^+ + g^-$ and $f^- + g^+$. These two quantities are also positive, therefore the optimal transport distance between them can be effectively computed.

In the particular case of the Wasserstein-1 distance we have used so far, following the dual formulation of W_1 , such a decomposition yields

$$\begin{aligned} W_1(f^+ + g^-, g^+ + f^-) &= \max_{\varphi \in \mathrm{Lip}_1} \int \varphi \left(f^+ + g^- - (g^+ + f^-) \right) dx \\ &= \max_{\varphi \in \mathrm{Lip}_1} \int \varphi \left(f^+ - f^- - (g^+ - g^-) \right) dx \\ &= \max_{\varphi \in \mathrm{Lip}_1} \int \varphi \left(f - g \right) dx \\ &= W_1(f,g). \end{aligned} \tag{2.4}$$

From the identity (2.4), we see that the strategy we have used, based on the dual form of the W_1 distance and its extension as the Kantorovich-Rubinstein norm, corresponds actually to the decomposition proposed by (Mainini, 2012). This brings an answer to the loss of convexity with respect to the time shifts observed on the simple comparison of Ricker signals. The decomposition does not ensure this convexity as it amounts to compute the optimal transport between the positive and the negative part of the residuals $f - g$.

This leads to two open questions which will be at the heart of my short term research activity.

1. The W_1 distance already provided interesting results for FWI, despite the loss of the convexity property which could be expected from optimal transport. What would be the results using the same decomposition with the W_2 distance (A. Allain PhD project) ?

2. Are there other ways of considering the data to recover the convexity property of the optimal transport without losing the physical information in the signal?

These questions will be investigated conjointly with the design of efficient numerical strategies for computing the optimal transport solution for large scale problems (several tens of millions to billions of unknowns). For such large scale problems, exact algorithms cannot be used, therefore we will rely on approximation schemes either based on

- the solution of the Monge-Ampère equation (Benamou et al., 2014) (W_2 only);
- the solution of the Benamou-Brenier equation (Benamou & Brenier, 2000) (W_2 only);
- semi-discrete multiscale approaches (Mérigot, 2011) (W_2 only);
- the solution through entropic regularization (Benamou et al., 2015) (general W_p distances).

Regarding the second question, one possibility which I want to investigate is based on an idea proposed by Thorpe et al. (2016). Instead of considering the optimal transport of functions $f(x)$ and $g(x)$ in a space of dimension d, one could consider transporting the graph $(x, f(x))$ $(x, g(x))$ in a space of dimension $d+1$. This higher dimensional space induces an increase of the computational complexity, however, after discretization, the two signals are described as point clouds in the graph space. The optimal transport problem is brought back to the comparison of these point clouds, yielding naturally the mass conservation and positivity properties required, without destroying the physical information carried by the signal. This strategy is thus appealing: the main difficulty will rely on the design of an efficient enough numerical strategy for the computation of the optimal transport solution in a space of dimension $d+1$.

3 Additional methodological developments

3.1 Combining FWI with homogenization

Seismic waves propagate within the subsurface with a given frequency content. They can bring structural information only on heterogeneities with scales to the order of the shortest propagated wavelength (at best a fraction of this shortest wavelength). Heterogeneities at smaller scales are seen by these waves as a local anisotropy of the subsurface. For a given small scale model (containing or not anisotropy), its homogenized counterpart (for a given frequency content) is a fully anisotropic medium which does not contain any heterogeneities smaller than the reference wavelength. For a wave propagating within this frequency band, the propagation both in the small scale and the homogenized model is strictly equivalent. First results on homogenization for seismic waves go back to Backus (1962) for 1D tabular models. Recent generalization to 2D non-periodic elastic media have been proposed by Capdeville et al. (2010); Capdeville & Marigo (2013); Capdeville & Cance (2015), where the homogenized model is computed through the solution of an elliptical equation.

In the frame of a French national research project (ANR project) with Y. Capdeville as principal investigator, we want to investigate how homogenization can help to mitigate the non-uniqueness of the FWI problem. In other words, while for a given FWI problem, there is an infinite number of solutions satisfying the data at a given accuracy, it might be possible that in the homogenized model space, these models correspond to the homogenized version of the "true" model (Capdeville, 2016). This is a conjecture which needs to be investigated (is it true? to what extent? what are the limitations?), but which makes sense from a physical point of view: homogenization takes into account the effective sensitivity of the recorded waves to the subsurface heterogeneities. Trying to recover information below this sensitivity is equivalent to explore the null space and is thus subjected to the introduction of spurious artifacts in the reconstruction. While this issue is usually mitigated through the use of appropriate regularization technique (see for instance Trinh et al. (2017b)), homogenization might provide a systematic approach to perform it.

3.2 Uncertainty quantification in FWI: a data assimilation strategy?

One very important question which has so far not really been addressed in the framework of FWI is related to uncertainty quantification. The indetermination of the FWI problem (due to the presence

of noise, inaccurate modeling, lack of illumination, cycle skipping) leads to the non-uniqueness of the subsurface parameters being able to provide the same level of misfit. This non-uniqueness prompts us to investigate strategies to compute the uncertainty associated with the reconstructed parameters.

In a Bayesian inversion framework, this information is carried out by the posterior covariance matrix, which corresponds, in the least-squares formalism, to the inverse Hessian operator. The size of the Hessian matrix for realistic scale application makes it impossible to compute and invert with reasonable computational resources. We want to exploit to this end uncertainty estimation strategies which have been at the heart of research lead in the data assimilation community for weather forecasting, known as extended Kalman filter techniques (Evensen, 2003, 2009).

These techniques are based on the assumption of a unimodal function with a potentially very flat valley of attraction, leading to strong indetermination of the parameter reconstruction. This is the case for FWI when the initial velocity model is sufficiently accurate to avoid cycle skipping. The presence of cycle skipping would lead to a multi-modal misfit function.

Data assimilation problems consist in determining the initial state of a dynamical system (for instance salinity, or temperature of the ocean, or density, pressure, and flow velocity vector field of the atmosphere) given a modeling operator for this dynamical system and some local, sparse observations. The uncertainty on the reconstruction is often crucial (for instance for weather forecasting). The basis of extended Kalman filter is to consider the evolution of a population of initial states of the system, with a given statistical repartition (covariance operator). Each member of the population is forecast in time (forecasting step) following the dynamic modeling operator available, and then corrected to account for the observed data (analysis step). The statistic of the ensemble is also forecast and analyzed. This iterative process yields an estimation of the initial state of the system as the mean of the member of the ensemble. The spreading of the members in the model space also yields an information on the posterior-covariance operator, under the form of a low-rank approximation of this operator. The rank of approximation of this operator should be equal to the number of members in the ensemble.

Being able to compute this low rank approximation of the posterior-covariance matrix in the context of FWI is appealing. However, this is not straightforward because the theory of the extended Kalman filter is based on the assumption that an information on the initial state of a dynamical system is computed, while in FWI we are interested in the reconstruction of PDE parameters (a static problem).

We have proposed in a PhD research project to mimic a dynamic evolution through the hierarchical approaches which are usually considered in FWI. In this framework, the ensemble is a set of subsurface models built around a reasonable initial model. The forecasting step is the solution of a FWI problem itself, for a subset of the data (data filtered in a particular band, subset of sources and/or receivers, specific time-window for instance). The analysis step consists in the comparison of modeled data with observed data for the next subset of data. In this framework, the final ensemble yields a low-rank estimation of the posterior covariance operator, from which we can extract the variance map of the reconstructed parameters (diagonal of the operator). Resolution and trade-offs between parameters can be extracted from the analysis of the rows of this operator.

The first results we have obtained for 2D frequency-domain acoustic FWI are encouraging (Thurin et al., 2017a). The system dynamic is mimicked with the conventional multi-scale hierarchical frequency approach (interpretation from low to high frequency). On the 2D Marmousi model with a surface acquisition, the variance which is reconstructed increases smoothly with depth and with the horizontal distance, which is related to the illumination of the subsurface: the uncertainty is higher for less illuminated parameters. The variance also increases along the interfaces of the model, where the solution is not enough constrained by the band-limited data.

These results open many questions which need to be investigated.

- Is it possible to go beyond a qualitative interpretation of the variance map and infer some quantitative estimate of the uncertainty on the parameter reconstruction?
- How to design the initial covariance matrix of the ensemble? This estimation should incorporate the prior information available on the solution. Also, finding efficient numerical techniques to ensure the initial population satisfies this initial covariance is not straightforward.
- How will behave other hierarchical strategies to mimic the dynamic evolution: sources/receivers

subsets or random sampling, time-windowing?

- How can we control the accuracy of the low-rank approximation of the posterior covariance matrix depending on the number of members in the population?

Furthermore, we think that the method would be extremely valuable in multi-parameter contexts, where the trade-offs between parameters increases the uncertainty in the reconstruction. It should also allow to better handle the incorporation of prior information through the natural incorporation of prior covariance matrices in the framework of data assimilation. The method could also be used to assess the sensitivity of a FWI application with respect to empirically chosen parameters such as regularization weights, and other pre-processing parameters that are needed to be set up by hand for practical applications.

3.3 Asymptotic approaches for travel-time computation

Asymptotic approximation of the wave equation is a less expensive alternative (in terms of computational cost and specifically in terms of memory request) to the computation of the full wavefield. An ansatz of the solution, composed as the product of a smoothly varying amplitude term and rapidly oscillating term, is plugged into the wave equation. Assuming the variation of the medium are large compared to the wavelength of the propagated signal, the asymptotic approximation yields a set of equations related to the first arrival travel-time (Eikonal equation) and the amplitude term (transport equation).

The asymptotic approximation of the wave propagation was used in seismic imaging already by (Clayton & Stolt, 1981) under the name WKBJ development, and later on by Beylkin (1985) and Bleistein (1987) for designing a seismic migration operator for reflection data.

It was also used in the so-called Ray+Born inversion scheme (Lambaré et al., 1992; Forgues & Lambaré, 1997; Lambaré et al., 2003). The computation of first-arrival traveltimes can also be used directly for tomography methods dedicated to the reconstruction of smoothly varying velocity models (Nolet, 2008). In Métivier et al. (2015c) we have shown that the asymptotic approximation can provide an asymptotic version of the Hessian operator that can be used to precondition inversion methods based on the full wavefield, such as FWI, and possibly with more impact, reverse time migration (RTM) algorithms.

The possibility to compute the travel-times and their gradient in a given velocity model is therefore appealing. Conventional strategies are either based on ray tracing approaches (Lagrangian approach), or on finite-difference solvers for the Eikonal equation (Eulerian approach).

Ray-tracing approaches are limited because of multipathing and triplication phenomena. Several rays may cross at a given point, and sometimes at an infinity of them, where are located caustics, yielding different traveltime values for one couple of points. In addition, there exist some source-receiver couples for which no classical ray can be found (existence of shadow zones).

The most efficient finite-difference techniques for the solution of the Eikonal equation are coupled with what is called Fast Sweeping Method (FSM) (Tsai et al., 2003; Noble et al., 2014). FSM are multi-pass algorithms relying on global orderings of the nodes. All the nodes are updated at each Gauss-Seidel iteration (sweep), following alternative orderings (see Fomel et al. (2009); Waheed et al. (2015) for instance). However, these solvers are limited to the Eikonal equation coming from simple acoustic equations. Accounting for anisotropy of the subsurface is not straightforward.

In an attempt to overcome the difficulties of these two techniques, we have investigated a more generic approach. The Eikonal equation can be rewritten under the form of a Hamilton-Jacobi equation. Each wave equation gives rise to a new Hamiltonian function, however the Hamilton-Jacobi formalism remains valid. This class of equations has extensively been studied in the mathematical and numerical analysis literature. A fundamental result is the existence and uniqueness of a class of solutions named "viscosity solutions" (Crandall & Lions, 1983). Recently, a discontinuous Galerkin scheme has been proposed with a particular definition of the fluxes, adapted to the solution of Hamilton-Jacobi equation (Cheng & Wang, 2014). The investigation of the efficiency of this scheme for 2D acoustic anisotropic modeling is the main topic of the PhD project of P. Le Bouteiller started at the end of 2015.

A 2D solver which follows this approach has already been developed, and tested on realistic subsurface

models, yielding promising results (Le Bouteiller et al., 2017). The next step for this work is to combine this approach with fast sweeping techniques, to improve the computational efficiency of the method. The extension to a 3D framework should also be straightforward, at least from a methodological point of view.

More interestingly, this research project should provide an efficient Hamilton-Jacobi Discontinuous Galerkin solver for the computation of the viscosity solution. This might open the way towards the computation of not only first-arrival travel-times but any travel-times, working in the phase space (potentially a 7D space for 3D geometries). Being able to compute the travel times of all the seismic events would be a significant step towards an automatic interpretation of the seismic data. This would yield the possibility to detect particular phases traveling in zones of interest: for instance for sub-salt imaging.

Interestingly, the design of a Hamilton-Jacobi solver also make a link with the optimal transport topic as following the approach proposed by (Benamou & Brenier, 2000), the computation of the Wasserstein-2 distance amounts to the solution of a Hamilton-Jacobi equation for the computation of a pressure-less potential flow. This might provide an alternative numerical approach to the one proposed by these authors, which is based on convex optimization techniques and the Uzawa relaxation algorithm.

3.4 Regularization strategies: what can we learn from image processing techniques?

Regularization strategies take an important part in the success of FWI applications. While a systematic approach might rely on an homogenization approach (see 3.1) more pragmatic approaches rely on smoothing techniques (Gaussian filter, Laplace filter, more recently Bessel filters (Trinh et al., 2017b)), either applied to the subsurface model or to the descent direction within the nonlinear optimization scheme. A standard challenge is the ability to remove oscillatory components associated with noise (scale below the smallest wavelength), while being able to preserve the main interfaces/discontinuities. The Bessel filter we have introduce is flexible enough to incorporate local directions and coherent lengths.

However, a number of methods have been designed in the image processing community to overcome the same difficulties. Of particular interest are the methods based on nonlinear anisotropic diffusion (Perona & Malik, 1990; Weickert, 1998; Cottet & Germain, 1993). The smoothing is performed through the solution of the equation

$$\frac{\partial u}{\partial t} = \nabla(g(|\nabla u|)u), \tag{3.1}$$

where the diffusivity $g(|x|)$ is a non-negative decreasing function tending to 0 at infinity. The function g controls the smoothing process by admitting strong diffusion if the gradient ∇u is small (possibly caused by noise) and by slowing down (or even stopping) the smoothing for large gradients. As an example of diffusivity function, the Perona-Malik function is given by

$$g(|x|) := \frac{1}{1 + \frac{x^2}{\kappa^2}}. \tag{3.2}$$

The strength of the filter thus increases in the region where the gradient of the model is small, while it decreases near discontinuities where the gradient of the model is large. These methods can be combined with curvelet/wavelet denoising operator in a very efficient manner (Plonka & Ma, 2008). Recently, strategies have also been proposed to determine the parameter of these filters through machine learning strategies (Chen et al., 2015).

These techniques could be applied as regularization tool for FWI, and also for seismic data denoising and inpainting. There are always inoperative sensor during a seismic acquisition, leading to empty traces on 2D common shot gathers, which need to be filled in, more often through interpolation techniques. The nonlinear diffusion technique has been proven to be quite efficient for inpainting problems, it would be interesting to assess the interest of this strategy in the context of seismic data.

An alternative to anisotropic nonlinear diffusion in the seismic imaging community rely on Total Variation (TV) regularization with a L^1 measure of the misfit between the original image and the denoised image. In Lellmann et al. (2014), the link between the L^1 norm and the Kantorovich-Rubinstein

norm we have used for modifying the FWI misfit function is presented. It is shown that this norm can be seen as a generalization of the L^1 norm, and that it shares some similarities with the G-Meyer's norm used for cartoon-texture decomposition. It is thus suggested that the KR-TV problem might be a good candidate for efficient denoising algorithms. Assessing the interest of this approach for seismic data denoising, FWI regularization, and seismic imaging through migration could be an interesting topic to investigate in the future.

3.5 Towards 3D elastic multi-parameter FWI in the frequency-domain

The successful results obtained on the 3D Valhall data-set using acoustic frequency-domain FWI (Operto et al., 2015b) lead us towards extending these results to frequency-domain elastic inversion. Frequency-domain strategies bring substantial computational savings when the redundancy of the data is sufficient, which is the case for dense marine Ocean Bottom Cable (OBC) surveys. Going towards elastic inversion is a step further that should give complementary information on the structure of the subsurface through information on shear wave velocities and possibly density and anisotropy parameters.

In this perspective, we need to design a feasible strategy for computing frequency-domain solution of the elastic wave equation. This strategy must handle acoustic/elastic interfaces as the main target is 3D OBC data. We thus have started a post-doc project which should assess and compare different strategies, using either iterative or direct solver approaches.

As spectral element methods are not well suited to handle acoustic/elastic interfaces, the first step of this project is to design a 3D optimized finite difference compact stencil following the approach of Gosselin-Cliche & Giroux (2014).

For the solution of the corresponding linear system, we will investigate the efficiency of the pre-conditioned CARP-CG strategy we have designed (Li et al., 2015b; Hamitou, 2016), combined with multiple right hand sides acceleration such as block conjugate gradient approaches (O'Leary, 1980) and/or Galerkin projection approaches (Chan & Ng, 1999).

We will compare this approach with direct solvers developed by different groups: the MUMPS solver (MUMPS team, 2015), the PASTIX solver (Hénon et al., 2002), and the IBM WSMP solver (Gupta et al., 2009). All of these solvers start implementing recently introduced low rank compression strategies to reduce the memory imprint of the LU factorization (Amestoy et al., 2015c). The question for us is to assess which linear system size these methods can handle on common HPC resources available for academics (local, national and European centers), which have limited memory per node.

Alternatively, we will investigate the feasibility of other sparse factorizations based on low rank compression to serve as preconditioner for the solution through Krylov-based iterative solvers. This is an approach promoted by Engquist & Ying (2011) and Gatto & Hesthaven (2017), where the low rank factorization is obtained through a low-rank compression of the Schur complement. No application to visco-elastic equation of this strategy has been reported yet to the best of our knowledge.

3.6 Particles methods for the wave equation and Hamilton-Jacobi equations

Finally, another (more speculative) topic that could be investigated is related to the application of particle methods for the modeling of wave propagation.

The particle strategies have been developed mainly in the framework of fluid dynamics equations: Navier-Stokes equations, plasma equations (Bergdorf et al., 2005; Cottet, G.-H. et al., 2014). They are mesh free: instead of solving the equation on a given mesh, a Lagrangian approach is used. They have shown very impressive results in terms of computational efficiency, making an efficient use of GPU architectures. This efficiency is due to the fact that no CFL condition has to be satisfied by the time step (this is an inherent advantage of mesh-free technique). Also, directional splitting make the treatment of high dimension space very efficient.

These methods are adapted to the solution of first-order advection equations. Since the wave propagation equation can be cast as a first-order hyperbolic system, the particle filter technique could be applied to the solution of these equations, considering the wave equation as a generalized advection system. Some attempts in this direction have already been proposed in the framework of Smoothed

Particle Hydrodynamics (SPH) methods (Li et al., 2015a; Zhang et al., 2016) and Lattice-Boltzmann strategies (Salomons et al., 2016).

This technique could also be applied to the solution of Hamilton-Jacobi equations. We have seen in Section 3.3 that the computation of asymptotic solution to the wave equation amounts to the solution of Hamilton-Jacobi equations. In particular, in the framework of the computation of the asymptotic solution in the phase space, for retrieving not only the first-arrival travel-times but all the travel-times (any phase), these kind of techniques might be helpful to mitigate the strong increase of the computational cost due to the high dimension of the problem (up to 7D for 3D applications).

This is a topic that could be investigated through a strong interaction with experts from the LJK: C. Picard, G.H. Cottet.

4 Multiparameter FWI: application to real data

4.1 Application to exploration scale and crustal scale seismic data

The third main objective of my research project is related to the application of 3D multi-parameter FWI strategies to real data. Investigating large scale FWI problems currently requires time-domain based algorithms. Frequency-domain approaches are still limited by the memory imprint of direct solvers for large scale problems. For this reason, the implementation of 3D visco-acoustic (using finite-differences (Virieux, 1986; Levander, 1988; Fornberg, 1998)) and 3D visco-elastic (using spectral elements (Komatitsch & Tromp, 2002a; Capdeville et al., 2003)) time-domain codes has been recently performed within the SEISCOPE project.

We have introduced attenuation through standard linear solid models in both codes (Yang et al., 2016c). A specific strategy has been designed for the computation of gradient through the adjoint strategy in the time-domain. The correlation of the incident and adjoint wavefield requires to access the two wavefields at the same time while the incident wavefields is propagating forward from an initial condition and the adjoint is propagating backward from a final condition. Our strategy is based on an extension of checkpointing strategies and makes possible the efficient computation of the gradient with limited memory resources, working only in-core (no storage on disks) (Yang et al., 2016a). Based on this strategy, the truncated Newton strategy has been implemented in the visco-acoustic time domain FWI code (Yang et al., 2017b). It should be soon implemented in its visco-elastic counterpart.

Applying these codes to the inversion of real marine and land data will be the objective of the next years. These applications will be conducted through the definition of PhD projects. Several targets have already been identified and some of these PhD projects have already started. At the exploration scale, we shall have access, through the sponsors of the SEISCOPE project, to several industrial data-sets. The first is a land data-set with complex topography. The main difficulty will be associated with the correct modeling and interpretation of surface waves. We might have access to another land data set acquired in Mongolia where very low frequencies (down to 1.5 Hz) have been recorded. The topography is flat for this case study, but the problem size is challenging for elastic FWI because of low shear velocities.

A 4D marine data set acquired in the North sea in a shallow water environment might also be made available to SEISCOPE. These data are acquired with a dense receiver coverage using ocean bottom cables (OBC). The term 4D refers to the fact that the acquisition has been repeated at different times to monitor the evolution in time of the subsurface (Lumley, 2001). This yields very interesting challenges and perspectives both in terms of applications and methodology. Specific FWI techniques are designed for 4D imaging, to try to account at best for the changes between the two acquisition times (Asnaashari et al., 2015). The inversion should be performed first in the visco-acoustic approximation. Apart from the imaging challenge, optimal transport could also reveal as an important tool to directly process the data. The localization of the time shifts between two data-sets acquired at different times could be performed through the computation and the analysis of the transport plan.

A collaboration with the Japan-Agency for Marine-Earth Science (JAMSTEC) `http://www.jamstec.go.jp/e/` is also starting, which should give us access to ultra long offset marine seismic data acquired for imaging deep crustal targets, such as subduction zones. Improving the resolution of the subsurface parameters estimation in these zones is crucial to better understand the geodynamic processes in

progress which are responsible for earthquakes. One of the first successful application of FWI on real data was performed on the Nankai trough (Dessa et al., 2004; Operto et al., 2006; Gorszczyk et al., 2017). Since then the quality of the data and the coverage have substantially increased, yielding interesting perspectives in terms of imaging, with new acquisition vessels deployed by JAMSTEC in 2018.

4.2 Application to GPR and electric data: towards multi-physics inversion

We have started investigating how FWI can be applied to GPR data through two PhD projects. From the applications to real data, two main limitations have been emphasized.

- The inaccurate modeling of electromagnetic wave propagation, due to the 2D assumption.
- The strong indetermination on the conductivity parameter and the important coupling between the conductivity and the resistivity.

Further exploring the application of FWI to GPR data would thus require to overcome these limitations.

The design of a 3D time-domain Maxwell's equations solver adapted to the GPR data modeling would be a first step. State-of-the-art numerical methods for the solution of the Maxwell's equations either rely on finite-differences schemes (FDTD) such as the open-source solver gprMAX (Warren et al., 2016), or finite-element discontinuous Galerkin methods (Hesthaven & Warburton, 2002). An alternative would be to consider a spectral element discretization of these equations, similar as the one we use for 3D visco-elastic time-domain modeling, to ensure we properly account for the free surface topography, while mitigating the high computational cost associated with discontinuous Galerkin strategies.

The combination of GPR data with other electric measurement such as Transient Electromagnetic data (TEM) and Electrical Resistivity Tomography (ERT) data would be the second step, to better constrain the conductivity reconstruction. These data bring low resolution information on the conductivity that may help mitigate the trade-offs between conductivity and permittivity in the GPR data.

The question how to combine these different type of data would be crucial. Besides step by step inversion, the combination of all the data in a single inversion problem would be a significant progress. This would require an adaptation of the modeling code for Maxwell's equation to the diffusion regime (TEM) and the static regime (ERT) which might not be a difficult task. The difficulty for TEM and ERT modeling code relies more on an accurate source implementation, which is usually treated through local mesh refinement techniques. Another option might be to use source injection technique (scattering problem in electromagnetism).

The question how to combine different types of data would also yield important perspectives for exploration geophysics. The FWI community working at the seismic exploration scale has been interested for some time in the combination of seismic data with gravity measurement (equivalent of ERT data for electromagnetic waves) to better constrain the reconstruction of the density (see for instance the work of Wehner et al. (2015)). The discrepancy of the wavenumber information between the two types of data (high resolution information from seismic, very low resolution information for gravity) make this combination hard when trying to go beyond simple step by step separate reconstructions. This work on electromagnetic data could be a first step towards providing answers to this question.

4.3 Application to noise correlation data for lithospheric targets

On a longer term, the feasibility of applying FWI algorithms to noise-correlation data will be investigated through an interaction between the SEISCOPE group and geophysicists from ISTerre laboratory, working in the team of M. Campillo. Noise-correlation is a passive seismic imaging technique relying on mathematical and statistical studies, and applied to geophysics data since the work of Campillo & Paul (2003); Shapiro & Campillo (2004); Shapiro et al. (2005). The leading idea is that the impulse seismic response between a couple of stations can be inferred through the correlation of the ambient noise which has been recorded, provided the recording duration is long enough and that the noise source is isotropically spread around the couple of stations (Bardos et al., 2008). During the last decades, this strategy has made increasing progresses towards the reconstruction of these impulse responses, from the reconstruction of surface waves to the detection of volumetric waves. The method seems mature enough

now so that the waveform reconstructed through ambient noise correlation might serve as observables that could be modeled and interpreted through FWI. The presence of experts in noise correlation in ISTerre is a very nice opportunity to investigate these questions I would like to benefit.

4.4 Application to ultrasound data

Another application I would like also to develop on a longer term is related to ultrasound data imaging through FWI. Already two types of applications could be investigated.

The first is the detection of defaults in industrial materials, in the field of non destructive testing. L. Moreau at ISTerre laboratory is currently working in this topic, and we have started interacting to go towards application of FWI to this type of data. The objective is to detect defaults in welds in metallic pieces. Compared to exploration seismic, the ability to record data both in transmission and reflection modes make us think it as a very promising field of application for FWI.

The second kind of application I would be interested in developing with ultrasound data is medical imaging. The techniques developed in FWI can be applied to breast cancer detection with ultrasound(Pratt et al., 2007). Taking into account the frequency of the signal (up to 1 MHz), the sound speed in human tissues (1500 $m.s^{-1}$), the average wavelength of the signal reaches 1.5 mm, for target size of approximately 250 mm, yielding a number of wavelength to propagate to the order of 200. This corresponds to medium to large size problems for exploration scale studies. These applications imply strong trade-offs between the wave velocity and the attenuation parameter, for which the methods we develop for multi-parameter inversions should be helpful. A strong advantage of medical imaging over seismic applications is again the target illumination: contrary to the seismic case, sources and receivers can be located all around the target, offering an optimal illumination of the target.

Going a step beyond, accounting for the elastic effect of the propagation at this scale might also yield the possibility to retrieve even finer details on the characteristics and the sanity of the tissues. This would of course imply significant computational cost because of low shear wave velocities.

Besides this extension, quasi real time inversion would be needed for medical imaging applications, which would require to design faster and more efficient algorithms compared to what is used for exploration seismic. For these applications, the computation of the results can take several days, a delay which seems less reasonable in a medical imaging framework. One possibility to initiate this research project would be to start interacting with S. Catheline in Univ. Claude Bernad (Lyon), who is a specialist in medical imaging with ultrasonic and elastic fields.

5 Development of open source toolboxes

Finally, an important point for the diffusion of research dedicated to the design of numerical strategies is the ability to provide small, versatile, open-source codes, however adapted to large scale computing. This is something I will try to enforce within my research activity, based on the positive experience of the design of the SEISCOPE toolbox (Métivier & Brossier, 2016b). This nonlinear smooth optimization toolbox is used in all the codes developed in the SEISCOPE research group, and starts being used outside.

The next toolbox that should be designed is related to the optimal transport strategy we have implemented for the computation of the Wasserstein-1 distance. It should be a `FORTRAN90` package linked with the `FISHPACK` `https://www2.cisl.ucar.edu/resources/legacy/fishpack` and the `MUDPACK` `https://www2.cisl.ucar.edu/resources/legacy/mudpack` libraries for the solution of the Poisson equation which has to be solved at each iteration of the SDMM algorithm. The input should two 1D to 3D datasets, and the output should give the Wasserstein-1 distance between these two datasets, as well as the corresponding dual variable at optimum.

Another toolbox that should be developed is associated with the PhD project of A. Pladys started within SEISCOPE. One aim of this project is to perform an exhaustive review of all the misfit functions which have been proposed to mitigate cycle skipping (non-convexity) in FWI. This work should yield a benchmark 2D acoustic FWI `FORTRAN90` code implementing these misfit functions. In addition, it will offer the possibility for implementing any other misfit functions through very simple modifications.

This is a benefit from the adjoint state formulation for FWI: the gradient of the misfit function is given by a weighted correlation between the incident and the adjoint wavefield. Modifying the misfit function implies only to modify the adjoint source implementation to find its gradient. In terms of computer implementation, this means that modifying the misfit function can be done by modifying only two specific routines: one for the misfit function itself, one for the corresponding adjoint source. This would yield a very flexible tool to test already published misfit functions, and compare them with new propositions.

References

Abarbanel, S., Gottlied, D., & Hesthaven, J. S., 1999. Well-posed perfectly matched layers for advective acoustics, *Journal of Computational Physics*, **154**(2), 266–283.

Alkhalifah, T., 1998. Acoustic approximations for processing in transversely isotropic media, *Geophysics*, **63**, 623–631.

Alkhalifah, T., 2000. An acoustic wave equation for anisotropic media, *Geophysics*, **65**, 1239–1250.

Ambrosio, L., Gigli, N., & Savaré, G., 2008. *Gradient flows: in metric spaces and in the space of probability measures*, Springer Science & Business Media.

Ambrosio, L., Mainini, E., & Serfaty, S., 2011. Gradient flow of the Chapman–Rubinstein–Schatzman model for signed vortices, *Annales de l'Institut Henri Poincaré (C) Non Linear Analysis*, **28**(2), 217–246.

Amestoy, P., Duff, I. S., & L'Excellent, J. Y., 2000. Multifrontal parallel distributed symmetric and unsymmetric solvers, *Computer Methods in Applied Mechanics and Engineering*, **184**(2-4), 501–520.

Amestoy, P., , Brossier, R., Buttari, A., L'Excellent, J.-Y., Mary, T., Métivier, L., Miniussi, A., Operto, S., Virieux, J., & Weisbecker, C., 2015a. 3D frequency-domain seismic modeling with a parallel BLR multifrontal direct solver, in *Expanded Abstracts, 85^{th} Annual SEG Meeting (New Orleans)*, pp. 3606–3611, SEG.

Amestoy, P., Brossier, R., Buttari, A., L'Excellent, J.-Y., Mary, T., Métivier, L., Miniussi, A., Operto, S., Ribodetti, A., Virieux, J., & Weisbecker, C., 2015b. Efficient 3D frequency-domain full-waveform inversion of ocean-bottom cable data with sparse block low-rank direct solver: a real data case study from the North Sea, in *Expanded Abstracts, 85^{th} Annual SEG Meeting (New Orleans)*, pp. 1303–1308, SEG.

Amestoy, P., Brossier, R., Buttari, A., L'Excellent, J.-Y., Mary, T., Métivier, L., Miniussi, A., & Operto, S., 2016a. Fast 3d frequency-domain full waveform inversion with a parallel Block Low-Rank multifrontal direct solver: application to OBC data from the North Sea, *Geophysics*, **81**(6), R363 – R383.

Amestoy, P., Buttari, A., L'Excellent, J., & Mary, T., 2016b. Sparse direct solvers towards seismic imaging of large 3D domains, in *Methods and Challenges of Seismic Wave Modelling for Seismic Imaging Workshop, 78^{th} Annual EAGE Meeting (Vienna)*.

Amestoy, P. R., Ashcraft, C., Boiteau, O., Buttari, A., L'Excellent, J.-Y., & Weisbecker, C., 2015c. Improving multifrontal methods by means of block low-rank representations, *SIAM Journal on Scientific Computing*, **37**(3), 1451–1474.

Appelö, D. & Kreiss, G., 2006. A new absorbing layer for elastic waves, *Journal of Computational Physics*, **215**(2), 642 – 660.

Asano, K. & Iwata, T., 2016. Source rupture processes of the foreshock and mainshock in the 2016 kumamoto earthquake sequence estimated from the kinematic waveform inversion of strong motion data, *Earth, Planets and Space*, **68**(1), 147.

Asnaashari, A., Brossier, R., Castellanos, C., Dupuy, B., Etienne, V., Gholami, Y., Hu, G., Métivier, L., Operto, S., Pageot, D., Prieux, V., Ribodetti, A., Roques, A., & Virieux, J., 2012. Hierarchical approach of seismic full waveform inversion, *Numerical Analysis and Applications*, **5**(2), 99–108.

Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P., & Virieux, J., 2015. Time-lapse seismic imaging using regularized full waveform inversion with prior model: which strategy?, *Geophysical Prospecting*, **63**(1), 78–98.

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

Bailey, J. T., Evans, S., & Robin, G. Q., 1964. Radio echo sounding of polar ice sheets, *Nature*, **204**(4957), 420–421.

Bardos, C., Garnier, J., & Papanicolaou, G., 2008. Identification of Green's functions singularities by cross correlation of noisy signals, *Inverse Problems*, **24**(1), 015011.

Barucq, H., Boillot, L., Diaz, J., & Ettouati, S., 2016. Performance analysis of discontinuous galerkin methods for seismic wave propagation, in *Methods and Challenges of Seismic Wave Modelling for Seismic Imaging Workshop, 78th Annual EAGE Meeting (Vienna)*.

Basu, U. & Chopra, A. K., 2003. Perfectly matched layers for time-harmonic elastodynamics of unbounded domains: theory and finite-element implementation, *Computer Methods in Applied Mechanics Engineering*, **192**, 1337–375.

Baumstein, A., Ross, W., & Lee, S., 2011. Simultaneous source elastic inversion of surface waves, in *Expanded Abstracts*, p. C040, European Association of Geoscientists and Engineers.

Becache, E., Fauqueux, S., & Joly, P., 2003. Stability of Perfectly Matched Layers, Group Velocities and Anisotropic Waves, *Journal of Computational Physics*, **188**, 399–433.

Beller, S., Monteiller, V., Operto, S., Nolet, G., Combe, L., Métivier, L., Virieux, J., Nissen-Meyer, T., Paul, A., & Zhao, L., 2015. 3d elastic full waveform inversion of teleseismic data for high-resolution ospheric imaging, in *77th EAGE Conference and Exhibition 2015*.

Ben Hadj Ali, H., Operto, S., & Virieux, J., 2011. An efficient frequency-domain full waveform inversion method using simultaneous encoded sources, *Geophysics*, **76**(4), R109.

Benamou, J.-D. & Brenier, Y., 2000. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem, *Numerische Mathematik*.

Benamou, J. D., Froese, B. D., & Oberman, A. M., 2014. Numerical solution of the Optimal Transportation problem using the Monge-Ampère equation, *Journal of Computational Physics*, **260**, 107–126.

Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L., & Peyré, G., 2015. Iterative Bregman Projections for Regularized Transportation Problems, *SIAM Journal on Scientific Computing*, **37**(2), A1111–A1138.

Bentley, C. R., 1964. The structure of Antarctica and its ice cover, in *Research in Geophysics, v.2 Solid Earth and Interface Phenomena*, pp. 335–389, Technology Press of Massachusetts Institute of Technology.

Bérenger, J.-P., 1994. A perfectly matched layer for absorption of electromagnetic waves, *Journal of Computational Physics*, **114**, 185–200.

Bérenger, J. P., 1996. Three-dimensional perfectly matched layer for the absorption of electromagnetic waves, *Journal of Computational Physics*, **127**(2), 363–379.

Bergdorf, M., Cottet, G.-H., & Koumoutsakos, P., 2005. Multilevel Adaptive Particle Methods for Convection-Diffusion Equations, *Multiscale Modeling & Simulation*, **4**(1), 328–357.

Bermúdez, A., Hervella-Nieto, L., Prieto, A., & Rodríguez, R., 2007. An optimal perfectly matched layer with unbounded absorbing function for time-harmonic acoustic scattering problems, *Journal of Computational Physics*, **223**(2), 469–488.

Beylkin, G., 1985. Imaging of discontinuities in the inverse scaterring problem by inversion of a causal generalized Radon transform, *Journal of Mathematical Physics*, **26**, 99–108.

Billette, F. J. & Brandsberg-Dahl, S., 2004. The 2004 BP velocity benchmark, in *Extended Abstracts, 67^{th} Annual EAGE Conference & Exhibition, Madrid, Spain*, p. B035.

Biondi, B. & Almomin, A., 2013. Tomographic full waveform inversion (TFWI) by combining FWI and wave-equation migration velocity analysis, *The Leading Edge*, **September, special section: full waveform inversion**, 1074–1080.

Biondi, B. & Symes, W., 2004. Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging, *Geophysics*, **69(5)**, 1283–1298.

Bleistein, N., 1987. On the imaging of reflectors in the Earth, *Geophysics*, **52**(7), 931–942.

Bogachev, V. I., 2007. *Measure Theory*, no. vol. I,II in Measure Theory, Springer Berlin Heidelberg.

Bonnans, J. F., Gilbert, J. C., Lemaréchal, C., & Sagastizábal, C. A., 2006. *Numerical Optimization, Theoretical and Practical Aspects*, Springer series, Universitext.

Bouteiller, P. L., Jemaa, M. B., Chauris, H., Métivier, L., Tavakoli, B., Noble, M., & Virieux, J., 2017. Discontinuous galerkin method for tti eikonal equation, in *Proceedings of the 79^{th} EAGE Conference & Exhibition*, EAGE.

Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N., & Pugmire, D., 2016. Global adjoint tomography: first-generation model, *Geophysical Journal International*, **207**(3), 1739–1766.

Bozdağ, E., Trampert, J., & Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *Geophysical Journal International*, **185**(2), 845–870.

Brandt, A., 1977. Multi-level adaptive solutions to boundary-value problems, *Mathematics of Computation*, **31**, 333–390.

Bretaudeau, F., Brossier, R., Métivier, L., & Virieux, J., 2014. First-arrival delayed tomography using 1st and 2nd order adjoint-state method, in *Expanded Abstracts*, pp. 4757–4762, Society of Exploration Geophysics.

Brossier, R., Operto, S., & Virieux, J., 2009. Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion, *Geophysics*, **74**(6), WCC105–WCC118.

Brossier, R., Gholami, Y., Virieux, J., & Operto, S., 2010. 2D frequency-domain seismic wave modeling in VTI media based on a hp-adaptive discontinuous galerkin method, in *Expanded Abstracts, 72^{th} Annual International Meeting, EAGE*, p. C046.

Brossier, R., Pajot, B., Combe, L., Operto, S., Métivier, L., & Virieux, J., 2014. Time and frequency-domain FWI implementations based on time solver: analysis of computational complexities, in *Expanded Abstracts, 76^{th} Annual EAGE Meeting (Amsterdam)*.

Bunks, C., Salek, F. M., Zaleski, S., & Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics*, **60**(5), 1457–1473.

Burridge, R., 1996. Elastic waves in anisotopric media, Schlumberger-Doll Research.

Busch, S., van der Kruk, J., Bikowski, J., & Vereecken, H., 2012. Quantitative conductivity and permittivity estimation using full-waveform inversion of on-ground GPR data, *Geophysics*, **77**(6), H79–H91.

Campillo, M. & Paul, A., 2003. Long Range Correlations in the Diffuse Seismic Coda, *Science*, **299**, 547.

Capdeville, Y., 2016. Non-periodic homogeneization for seismic forward and inverse problems, in *Methods and Challenges of Seismic Wave Modelling for Seismic Imaging Workshop, 78th Annual EAGE Meeting (Vienna)*.

Capdeville, Y. & Cance, P., 2015. Residual homogenization for elastic wave propagation in complex media, *Geophysical Journal International*, **200**(2), 986.

Capdeville, Y. & Marigo, J.-J., 2013. A non-periodic two scale asymptotic method to take account of rough topographies for 2D elastic wave propagation, *Geophysical Journal International*, **192**, 163–189.

Capdeville, Y., Chaljub, E., Vilotte, J., & Montagner, J., 2003. Coupling the spectral element method with a modal solution for elastic wave propagation in global earth models, *Geophysical Journal International*, **152**, 34–67.

Capdeville, Y., Guillot, L., & Marigo, J.-J., 2010. 2-D non-periodic homogenization to upscale elastic media for P-SV waves, *Geophysical Journal International*, **182**, 903–922.

Castellanos, C., Métivier, L., Operto, S., Brossier, R., & Virieux, J., 2015. Fast full waveform inversion with source encoding and second-order optimization methods, *Geophysical Journal International*, **200(2)**, 720–744.

Castellanos-Lopez, C., Métivier, L., Operto, S., & Brossier, R., 2013. Fast full waveform inversion with source encoding and second order optimization methods, in *75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013*, p. We 11 10.

Cerjan, C., Kosloff, D., Kosloff, R., & Reshef, M., 1985. A nonreflecting boundary condition for discrete acoustic and elastic wave equations, *Geophysics*, **50**(4), 2117–2131.

Chan, T. F. & Ng, M. K., 1999. Galerkin projection methods for solving multiple linear systems, *SIAM Journal on Scientific Computing*, **21**(3), 836–850.

Chavent, G., 1971. *Analyse fonctionnelle et identification de coefficients répartis dans les équations aux dérivées partielles*, Ph.D. thesis, Université de Paris.

Chen, Y., Yu, W., & Pock, T., 2015. On learning optimized reaction diffusion processes for effective image restoration, in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Cheng, Y. & Wang, Z., 2014. A new discontinuous Galerkin finite element method for directly solving the Hamilton–Jacobi equations, *Journal of Computational Physics*, **268**, 134–153.

Choi, Y. & 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.

Clayton, R. & Engquist, B., 1977. Absorbing boundary conditions for acoustic and elastic wave equations, *Bulletin of the Seismological Society of America*, **67**, 1529–1540.

Clayton, R. W. & Stolt, R. H., 1981. A Born-WKBJ inversion method for acoustic reflection data, *Geophysics*, **46**, 1558–1565.

Cocher, E. & Chauris, H., 2014. Iterative migration to remove the imprint of multiples on the reflectivity, in *SEG/KOC Workshop: Seismic multiples – Are they signal or noise?*, Kuwait City.

Cocher, E., Chauris, H., & Lameloise, C. A., 2015. Imaging with surface-related multiples in the subsurface-offset domain, in *77th EAGE Conference & Exhibition*, Madrid.

Collino, F., 1993. High order absorbing boundary conditions for wave propagation models: Straight line boundary and corner cases, in *Proceedings of the Second International Conference on Mathematical and Numerical Aspects of Wave Propagation*, pp. 161–171, SIAM, Delaware.

Collino, F. & Tsogka, C., 2001. Application of the perfectly matched absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media, *Geophysics*, **66**, 294–307.

Combettes, P. L. & Pesquet, J.-C., 2011. Proximal splitting methods in signal processing, in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, vol. 49 of **Springer Optimization and Its Applications**, pp. 185–212, eds Bauschke, H. H., Burachik, R. S., Combettes, P. L., Elser, V., Luke, D. R., & Wolkowicz, H., Springer New York.

Cottet, G.-H. & Germain, L., 1993. Image processing through reaction combined with nonlinear diffusion, *Mathematics of Computation*, **61**(204), 659–673.

Cottet, G.-H., Etancelin, J.-M., Perignon, F., & Picard, C., 2014. High order semi-Lagrangian particle methods for transport equations: numerical analysis and implementation issues, *ESAIM: M2AN*, **48**(4), 1029–1060.

Crandall, M. G. & Lions, P. L., 1983. Viscosity solutions of Hamilton–Jacobi equations, *Transactions of the American mathematical society*, **277(1)**, 1–42.

Dahlen, F. A., Hung, S. H., & Nolet, G., 2000. Fréchet kernels for finite-difference traveltimes - I. theory, *Geophysical Journal International*, **141**, 157–174.

Dai, Y. & Yuan, Y., 1999. A nonlinear conjugate gradient method with a strong global convergence property, *SIAM Journal on Optimization*, **10**, 177–182.

ed. Daniels, D. J., 2004. *Ground Penetrating Radar*, Institution of Electrical Engineers, London, UK, 2nd edn.

Deeds, J. & Bradford, J., 2002. Characterization of an aquitard and direct detection of LNAPL at Hill Air Force base using GPR AVO and migration velocity analyses, in *9th International Conference on Ground Penetrating Radar (GPR 2002), Santa Barbara, California (USA)*, vol. 4758 of **SPIE proceedings series**, pp. 323–329.

Delprat-jannaud, F., Lailly, P., & Metivier, L., 2010. Method for imaging a target area of the subsoil using walkaway data, EP Patent 2,253,970.

Deparis, J. & Garambois, S., 2009. On the use of dispersive APVO GPR curves for thin-bed properties estimation: Theory and application to fracture characterization, *Geophysics*, **74**(1), J1–J12.

Dessa, J. X., Operto, S., Kodaira, S., Nakanishi, A., Pascal, G., Virieux, J., & Kaneda, Y., 2004. Multiscale seismic imaging of the eastern nankai trough by full waveform inversion, *Geophysical Research Letters*, **31**(L18606), doi:10.1029/2004GL020453.

Devaney, A., 1984. Geophysical diffraction tomography, *Geoscience and Remote Sensing, IEEE Transactions on*, **GE-22**(1), 3–13.

Diaz, J. & Joly, P., 2006. A time domain analysis of pml models in acoustics, *Computer Methods in Applied Mechanics and Engineering*, **195**(29-32), 3820–3853.

Dmitriev, M. M. & Lisitsa, V. V., 2011. Application of m-pml reflectionless boundary conditions to the numerical simulation of wave propagation in anisotropic media. part i: Reflectivity, *Numerical Analysis and Applications*, **4**(4), 271–280.

Dongarra, J., Eijkhout, V., & Kalhan, A., 1995. Reverse communication interface for linear algebra templates for iterative methods, Tech. rep., University of Tennessee.

Duff, I. S. & Reid, J. K., 1983. The multifrontal solution of indefinite sparse symmetric linear systems, *ACM Transactions on Mathematical Software*, **9**, 302–325.

Dupuy, B., Asnaashari, A., Brossier, R., Garambois, S., Métivier, L., Ribodetti, A., & Virieux, J., 2016. A downscaling strategy from FWI to microscale reservoir properties from high-resolution images, *The Leading Edge*, **35**, 1146–150.

Duveneck, E. & Bakker, P. M., 2011. Stable P-wave modeling for reverse-time migration in tilted TI media, *Geophysics*, **76**(2), S65–S75.

Eisenstat, S. C. & Walker, H. F., 1994. Choosing the forcing terms in an inexact Newton method, *SIAM Journal on Scientific Computing*, **17**, 16–32.

Engquist, B. & Froese, B. D., 2014. Application of the wasserstein metric to seismic signals, *Communications in Mathematical Science*, **12**(5), 979–988.

Engquist, B. & Majda, A., 1977. Absorbing boundary conditions of the numerical simulation of waves, *Mathematics of Computation*, **31**, 629–651.

Engquist, B. & Ying, L., 2011. Sweeping preconditioner for the Helmholtz equation: hierarchical matrix representation, *Communications on pure and applied mathematics*, **64**(5), 697–735.

Ernst, J. R., Green, A. G., Maurer, H., & Holliger, K., 2007. Application of a new 2D time-domain full-waveform inversion scheme to crosshole radar data, *Geophysics*, **72**(5), J53–J64.

Ernst, O. G. & Gander, M. J., 2012. Why it is Difficult to Solve Helmholtz Problems with Classical Iterative Methods, in *Numerical Analysis of Multiscale Problems*, vol. 83 of **Lecture Notes in Computational Science and Engineering**, pp. 325–363, eds Graham, I. G., Hou, T. Y., Lakkis, O., & Scheichl, R., Springer Berlin Heidelberg.

Etienne, V., Chaljub, E., Virieux, J., & Glinsky, N., 2010. An hp-adaptive discontinuous Galerkin finite-element method for 3D elastic wave modelling, *Geophysical Journal International*, **183**(2), 941–962.

Evans, R., Frost, M., Stonecliffe-Jones, M., & Dixon, N., 2008. A Review of Pavement Assessment Using Ground Penetrating Radar (GPR), in *12th International Conference on Ground Penetrating Radar (GPR 2008), Birmingham (UK)*.

Evensen, G., 2003. The ensemble kalman filter: theoretical formulation and practical implementation, *Ocean Dynamics*, **53**, 343–367.

Evensen, G., 2009. *Data assimilation : The ensemble Kalman filter*, Springer.

Fan, W., Shearer, P. M., & Gerstoft, P., 2014. Kinematic earthquake rupture inversion in the frequency domain, *Geophysical Journal International*, **199**(2), 1138.

Faye, J. P. & Jeannot, J. P., 1986. Prestack migration velocities from focusing depth analysis, in *Expanded Abstracts*, pp. 438–440, Soc. Expl. Geophys.

Fichtner, A., Kennett, B. L. N., Igel, H., & Bunge, H. P., 2008. Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain, *Geophysical Journal International*, **175**, 665–685.

Fichtner, A., Kennett, B. L. N., Igel, H., & Bunge, H. P., 2009. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophysical Journal International*, **179**(3), 1703–1725.

Fomel, S., Luo, S., & Zhao, H.-K., 2009. Fast sweeping method for the factored eikonal equation, *Journal of Computational Physics*, **228**, 6440–6455.

Forgues, E. & Lambaré, G., 1997. Parameterization study for acoustic and elastic ray+born inversion, *Journal of Seismic Exploration*, **6**, 253–278.

Fornberg, B., 1998. Classroom note: Calculation of weights in finite difference formulas, *SIAM review*, **40**(3), 685–691.

Gaffet, S., Guglielmi, Y., Virieux, J., Waysand, G., Chwala, A., Stolz, R., Emblanch, C., Auguste, M., Boyer, D., & Cavaillou, A., 2003. Simultaneous seismic and magnetic measurements in the low noise underground laboratory (LSBB) of rustrel, france, during the 2001, 26th January Indian earthquake, *Geophysical Journal International*, **155**, 977–986.

Gao, F., Atle, A., & Williamson, P., 2010. Full waveform inversion using deterministic source encoding, *SEG Technical Program Expanded Abstracts*, **29**(1), 1013–1017.

Gatto, P. & Hesthaven, J. S., 2017. A preconditioner based on low-rank approximation of Schur complements, *Journal of Scientific Computing*, p. in press.

Gauthier, O., Virieux, J., & Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: numerical results, *Geophysics*, **51**(7), 1387–1403.

Givoli, D., 2001. High-order non-reflecting boundary conditions without high-order derivatives, *Journal of Computational Physics*, **170**(2), 8.

Gordon, D. & Gordon, R., 2010. CARP-CG: A robust and efficient parallel solver for linear systems, applied to strongly convection dominated PDEs, *Parallel Computing*, **36**, 495–515.

Gorszczyk, A., Operto, S., & Malinowski, M., 2017. Full-waveform inversion of ocean-bottom seismometer for deep crustal imaging: the eastern Nankai Trough revisited, *Journal of Geophysical Research: Solid Earth*, **in-press**.

Gosselin-Cliche, B. & Giroux, B., 2014. 3D frequency-domain finite-difference viscoelastic-wave modeling using weighted average 27-point operators with optimal coefficients, *Geophysics*, **79**(3), T169–T188.

Griewank, A. & Walther, A., 2000. Algorithm 799: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation, *ACM Trans. Math. Software*, **26**, 19–45.

Groos, L., Schäfer, M., Forbriger, T., & Bohlen, T., 2014. The role of attenuation in 2D full-waveform inversion of shallow-seismic body and Rayleigh waves, *Geophysics*, **79**(6), R247–R261.

Gupta, A., Koric, S., & George, T., 2009. Sparse matrix factorization on massively parallel computers, in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, SC '09, pp. 1:1–1:12, ACM, New York, NY, USA.

Habashy, T. M., Abubakar, A., Pan, G., & Belani, A., 2011. Source-receiver compression scheme for full-waveform seismic inversion, *Geophysics*, **76**(4), R95–R108.

Haber, E., Chung, M., & Herrmann, F., 2012. An effective method for parameter estimation with PDE constraints with Multiple Right-Hand Sides, *SIAM Journal on Optimization*, **22**(3), 739–757.

Hale, D., 2013. Dynamic warping of seismic images, *Geophysics*, **78**(2), S105–S115.

Halpern, L., Petit-Bergez, S., & Rauch, J., 2011. The Analysis of Matched Layers, *Confluentes Mathematici*, **3**(2), 159–236.

Hamitou, O., 2016. *Efficient preconditioning method for the CARP-CG iterative solver for the solution of the frequency-domain visco-elastic wave equation*, Ph.D. thesis, Université Grenoble Alpes.

Hamitou, O., Métivier, L., Labbé, S., Brossier, R., & Virieux, J., 2015. Preconditioning and multiple-right hand sides strategies for the solution of the frequency-domain wave propagation problem using the CGMN method, in *SEG Technical Program Expanded Abstracts 2015*, pp. 3612–3616.

Hastings, F. D., Schneider, J. B., & Broschat, S. L., 1996. Application of the perfectly matched layer (PML) absorbing boundary condition to elastic wave propagation, *Journal of Acoustical Society of America*, **100**, 3061–3069.

Hénon, P., Ramet, P., & Roman, J., 2002. PaStiX: A High-Performance Parallel Direct Solver for Sparse Symmetric Definite Systems, *Parallel Computing*, **28**(2), 301–321.

Hesthaven, J., 1998. On the analysis and construction of perfectly matched layers for the linearized euler equations, *Journal of Computational Physics*, **142**, 129–147.

Hesthaven, J. & Warburton, T., 2002. Nodal High-Order Methods on Unstructured Grids, *Journal of Computational Physics*, **181**(1), 186 – 221.

Hu, F. Q., 1996. On absorbing boundary conditions for linearized euler equations by a perfectly matched layer, *Journal of Computational Physics*, **129**(1), 201–219.

Hustedt, B., Operto, S., & Virieux, J., 2004. Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling, *Geophysical Journal International*, **157**, 1269–1296.

Innanen, K. A., 2014. Seismic AVO and the inverse Hessian in precritical reflection full waveform inversion, *Geophysical Journal International*, **199**(2), 717–734.

Israeli, M. & Orszag, S. A., 1981. Approximation of radiation boundary conditions, *Journal of Computational Physics*, **41**, 115–135.

Jannane, M., Beydoun, W., Crase, E., Cao, D., Koren, Z., Landa, E., Mendes, M., Pica, A., Noble, M., Roeth, G., Singh, S., Snieder, R., Tarantola, A., & Trezeguet, D., 1989. Wavelengths of Earth structures that can be resolved from seismic reflection data, *Geophysics*, **54**(7), 906–910.

Jin, S., 1992. *Inversion de données de sismique pétrolière : séparation des paramètres élastiques et détermination des vitesse de référence.*, Ph.D. thesis, Université Paris 7.

Kaczmarz, S., 1937. Angenäherte Auflösung von Systemen linearer Gleichungen (English translation by Jason Stockmann), *Bulletin International de l'Académie Polonaise des Sciences et des Lettres*, **35**, 355–357.

Kalogeropoulos, A., van der Kruk, J., Hugenschmidt, J., Bikowski, J., & Brühwiler, E., 2013. Full-waveform GPR inversion to assess chloride gradients in concrete, *NDT&E International*, **57**, 74–84.

Kaltenbacher, B., Neubauer, A., & Scherzer, O., 2008. *Iterative Regularization Methods for Nonlinear Problems*, de Gruyter, Berlin, New York.

Kantorovich, L., 1942. On the transfer of masses, *Dokl. Acad. Nauk. USSR*, **37**, 7–8.

Klotzsche, A., van der Kruk, J., Linde, N., Doetsch, J., & Vereecken, H., 2013. 3-D characterization of high-permeability zones in a gravel aquifer using 2-D crosshole GPR full-waveform inversion and waveguide detection, *Geophysical Journal International*, **195**(2), 932–944.

Kolsky, H., 1956. The propagation of stress pulses in viscoelastic solids, *Philosophical Magazine*, **1**, 693–710.

Komatitsch, D. & Martin, R., 2007. An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation, *Geophysics*, **72**(5), SM155–SM167.

Komatitsch, D. & Tromp, J., 2002a. Spectral-element simulations of global seismic wave propagation, II. 3D models, oceans, rotation, and self-gravitation, *Geophysical Journal International*, **150**, 303–318.

Komatitsch, D. & Tromp, J., 2002b. Spectral-element simulations of global seismic wave propagation—i. validation, *Geophysical Journal International*, **149**(2), 390–412.

Komatitsch, D. & Tromp, J., 2003. A Perfectly Matched Layer absorbing boundary condition for the second-order seismic wave equation, *Geophysical Journal International*, **154**(1), 146–153.

Komatitsch, D., Ritsema, J., & Tromp, J., 2002. The spectral-element method, Beowulf computing, and global seismology, *Science*, **298**(5599), 1737–1742.

Korta, N., Fichtner, A., & Sallarčs, V., 2013. Block-diagonal approximate hessian for preconditioning in full waveform inversion, in *Expanged Abstracts, 75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013*.

Krebs, J., Anderson, J., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A., & Lacasse, M. D., 2009. Fast full-wavefield seismic inversion using encoded sources, *Geophysics*, **74(6)**, WCC105–WCC116.

Lailly, P., 1983a. The seismic inverse problem as a sequence of before stack migrations, in *Conference on Inverse Scattering, Theory and application, Society for Industrial and Applied Mathematics, Philadelphia*, pp. 206–220.

Lailly, P., 1983b. The seismic problem as a sequence of before-stack migrations, in *Conference on Inverse Scattering: Theory and Applications*, SIAM, Philadelphia.

Lambaré, G., 2002. The use of locally coherent events in depth processing : a state of the art., in *Extended Abstracts, 72nd annual meeting , (6-10 October 2002, Salt Lake City)*, pp. 2261–2264, Society of Exploration Geophysics.

Lambaré, G., Virieux, J., Madariaga, R., & Jin, S., 1992. Iterative asymptotic inversion in the acoustic approximation, *Geophysics*, **57**, 1138–1154.

Lambaré, G., Operto, S., Podvin, P., Thierry, P., & Noble, M., 2003. 3-D ray+Born migration/inversion - part 1: theory, *Geophysics*, **68**, 1348–1356.

Lambot, S., Weihermüller, L., Huisman, J. A., Vereecken, H., Vanclooster, M., & Slob, E. C., 2006. Analysis of air-launched ground-penetrating radar techniques to measure the soil surface water content, *Water Resources Research*, **42**, W11403.

Langouët, H., Métivier, L., Sinoquet, D., & Tran, Q.-H., 2008. Optimization for engine calibration, in *ENGOPT International conference on engineering optimization, Rio de Janeiro, Brazil*, pp. 1–5.

Langouët, H., Métivier, L., Sinoquet, D., & Tran, Q.-H., 2011. Engine calibration: multi-objective constrained optimization of engine maps, *Optimization and Engineering*, **12**(3), 407–424.

Lavoué, F., 2014. *2D Full waveform inversion of ground penetrating radar data: Towards multiparameter imaging from surface data*, Ph.D. thesis, Univ. Grenoble Alpes.

Lavoué, F., Brossier, R., Métivier, L., Garambois, S., & Virieux, J., 2013. 2D full waveform inversion of GPR surface data: permittivity and conductivity imaging, in *7th International Workshop on Advanced Ground-Penetrating Radar (IWAGPR 2013), Nantes (France)*.

Lavoué, F., Brossier, R., Métivier, L., Garambois, S., & Virieux, J., 2014. Two-dimensional permittivity and conductivity imaging by full waveform inversion of multioffset GPR data: a frequency-domain quasi-Newton approach, *Geophysical Journal International*, **197**(1), 248–268.

Lavoué, F., Brossier, R., Métivier, L., Operto, S., Garambois, S., & Virieux, J., 2015. Frequency-domain modelling and inversion of electromagnetic data for 2d permittivity and conductivity imaging: An application to the institut fresnel experimental data, *Near Surface Geophysics*, **13**(3), 227–241.

Le Bouteiller, P., Benjemaa, M., Métivier, L., & Virieux, J., 2017. An accurate discontinuous Galerkin method for solving point–source Eikonal equation in 2D heterogeneous anisotropic media, *Submitted to Journal of Computational Physics*.

Le Dimet, F. & Talagrand, O., 1986. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects, *Tellus A*, **38A**(2), 97–110.

Lellmann, J., Lorenz, D., Schönlieb, C., & Valkonen, T., 2014. Imaging with kantorovich–rubinstein discrepancy, *SIAM Journal on Imaging Sciences*, **7**(4), 2833–2859.

Levander, A. R., 1988. Fourth-order finite-difference P-SV seismograms, *Geophysics*, **53**(11), 1425–1436.

Li, X., Zhang, T., & Zhang, Y. O., 2015a. Time domain simulation of sound waves using smoothed particle hydrodynamics algorithm with artificial viscosity, *Algorithms*, **8**(2), 321–335.

Li, Y., Métivier, L., Brossier, R., Han, B., & Virieux, J., 2014a. A robust parallel iterative solver for frequency-domain elastic wave modeling, in *Expanded Abstracts, 76^{th} Annual EAGE Meeting (Amsterdam)*.

Li, Y., Métivier, L., Brossier, R., Han, B., & Virieux, J., 2014b. CARP-CG: A robust parallel iterative solver for frequency-domain elastic wave modeling, application to the Marmousi2 model, in *Expanded Abstracts, 84^{th} Annual SEG Meeting (Denver)*.

Li, Y., Métivier, L., Brossier, R., Han, B., & Virieux, J., 2015b. 2D and 3D frequency-domain elastic wave modeling in complex media with a parallel iterative solver, *Geophysics*, **80(3)**, T101–T118.

Li, Y., Han, B., Métivier, L., & Brossier, R., 2016. Optimal fourth-order staggered-grid finite-difference scheme for 3D frequency-domain viscoelastic wave modeling, *Journal of Computational Physics*, **321**, 1055–1078.

Lions, J. L., 1968. *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*, Dunod, Paris.

Liu, P. & Archuleta, R. J., 2004. A new nonlinear finite fault inversion with three-dimensional green's functions: Application to the 1989 loma prieta, california, earthquake, *Journal of Geophysical Research: Solid Earth*, **109**(B2).

Liu, Q. & Gu, Y., 2012. Seismic imaging: from classical to adjoint tomography, *Tectonophysics*, **566**, 31–66.

Lumley, D. E., 2001. Time-lapse seismic reservoir monitoring, *Geophysics*, **66**(1), 50–53.

Luo, J. & Wu, R.-S., 2015. Seismic envelope inversion: reduction of local minima and noise resistance, *Geophysical Prospecting*, **63**(3), 597–614.

Luo, S. & Sava, P., 2011. A deconvolution-based objective function for wave-equation inversion, *SEG Technical Program Expanded Abstracts*, **30**(1), 2788–2792.

Luo, Y. & Schuster, G. T., 1991. Wave-equation traveltime inversion, *Geophysics*, **56**(5), 645–653.

Ma, Y. & Hale, D., 2013. Wave-equation reflection traveltime inversion with dynamic warping and full waveform inversion, *Geophysics*, **78**(6), R223–R233.

Maggi, A., Tape, C., Chen, M., Chao, D., & Tromp, J., 2009. An automated time-window selection algorithm for seismic tomography, *Geophysical Journal International*, **178**, 257–281.

Mai, P. M., Schorlemmer, D., Page, M., Ampuero, J.-P., Asano, K., Causse, M., Custodio, S., Fan, W., Festa, G., Galis, M., et al., 2016. The earthquake-source inversion validation (siv) project, *Seismological Research Letters*.

Mainini, E., 2012. A description of transport cost for signed measures, *Journal of Mathematical Sciences*, **181**(6), 837–855.

Masoni, I., Brossier, R., Boelle, J.-L., Macquet, M., & Virieux, J., 2014a. Robust full waveform inversion of surface waves, *Seismic Technology*, **11**(4), 19.

Masoni, I., Zhou, W., Brossier, R., Métivier, L., Operto, S., & Virieux, J., 2014b. Near-surface full waveform inversion using surface waves and reflected waves, in *Expanded Abstracts, 76th Annual EAGE Meeting (Amsterdam)*.

McClymont, A. F., Green, A. G., Streich, R., Horstmeyer, H., Tronicke, J., Nobes, D. C., Pettinga, J., Campbell, J., & Langridge, R., 2008. Visualization of active faults using geometric attributes of 3D GPR data: An example from the Alpine Fault Zone, New Zealand, *Geophysics*, **73**(2), B11–B23.

Meles, G. A., van der Kruk, J., Greenhalgh, S. A., Ernst, J. R., Maurer, H., & Green, A. G., 2010. A new vector waveform inversion algorithm for simultaneous updating of conductivity and permittivity parameters from combination crosshole/borehole-to-surface GPR data, *IEEE Transactions on Geoscience and Remote Sensing*, **48**, 3391–3407.

Meles, G. A., Greenhalgh, S. A., Green, A. G., Maurer, H., & van der Kruk, J., 2012. GPR full-waveform sensitivity and resolution analysis using an FDTD adjoint method, *IEEE Transactions on Geoscience and Remote Sensing*, **50**(5), 1881–1896.

Mérigot, Q., 2011. A multiscale approach to optimal transport, *Computer Graphics Forum*, **30**(5), 1583–1592.

Métivier, L., 2009. Utilisation des équations Euler-PML en milieu hétérogène borné pour la résolution d'un problème inverse en géophysique, *ESAIM: Proc.*, **27**, 156–170.

Métivier, L., 2011. Interlocked optimization and fast gradient algorithm for a seismic inverse problem, *Journal of Computational Physics*, **230**(19), 7502–7518.

Métivier, L. & Brossier, R., 2016a. The seiscope optimization toolbox: A large-scale nonlinear optimization library based on reverse communication, *Geophysics*, **81**(2), F11–F25.

Métivier, L. & Brossier, R., 2016b. The SEISCOPE optimization toolbox: A large-scale nonlinear optimization library based on reverse communication, *Geophysics*, **81**(2), F11–F25.

Métivier, L. & Montarnal, P., 2012. Strategies for solving index one dae with non-negative constraints: Application to liquid–liquid extraction, *Journal of Computational Physics*, **231**(7), 2945–2962.

Métivier, L. & Roussel, H., 2012. Accounting robustly for instantaneous chemical equilibriums in reactive transport: a numerical method and its application to liquid–liquid extraction modeling, *Computers & Chemical Engineering*, **45**, 50–61.

Métivier, L., Delprat-Jannaud, F., Halpern, L., & Lailly, P., 2009. 2d nonlinear inversion of walkaway data, in *SEG Technical Program Expanded Abstracts 2009*, pp. 2342–2346, Society of Exploration Geophysicists.

Métivier, L., Lailly, P., Delprat-Jannaud, F., & Halpern, L., 2011. A 2D nonlinear inversion of well-seismic data, *Inverse Problems*, **27**(5), 055005.

Métivier, L., Brossier, R., Virieux, J., & Operto, S., 2012a. Optimization schemes in FWI: the truncated Newton method, in *2012 SEG Abstracts*.

Métivier, L., Brossier, R., Virieux, J., & Operto, S., 2012b. Toward Gauss-Newton and exact Newton optimization for full waveform inversion, in *EAGE, 74th Conference and Exhibition*.

Métivier, L., Brossier, R., Virieux, J., & Operto, S., 2013. Full Waveform Inversion and the truncated Newton method, *SIAM Journal On Scientific Computing*, **35(2)**, B401–B437.

Métivier, L., Bretaudeau, F., Brossier, R., Operto, S., & Virieux, J., 2014a. Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures, *Geophysical Prospecting*, **62**, 1353–1375.

Métivier, L., Brossier, R., Labbé, S., Operto, S., & Virieux, J., 2014b. A robust absorbing layer for anisotropic seismic wave modeling, *Journal of Computational Physics*, **279**, 218–240.

Métivier, L., Brossier, R., Labbé, S., Operto, S., & Virieux, J., 2014c. Smart: dissipative absorbing layer technique for general elastodynamics equations. application as s-waves filter in acoustic ti media., *Seismic Technology*, **11**(4), 14.

Métivier, L., Brossier, R., Operto, S., & Virieux, J., 2014d. A robust absorbing layer method for seismic wave simulation in anisotropic media, in *Expanded Abstracts, 76th Annual EAGE Meeting (Amsterdam)*.

Métivier, L., Brossier, R., Operto, S., & Virieux, J., 2014e. Multi-parameter FWI - an illustration of the Hessian operator role for mitigating trade-offs between parameter classes, in *Expanded Abstracts, 6th EAGE St-Petersbourg International Conference & Exhibition*.

Métivier, L., Brossier, R., Operto, S., & Virieux, J., 2014f. Smart: Robust absorbing layer and s-waves filtering for acoustic anisotropic wave simulation, in *Expanded Abstracts, 84th Annual SEG Meeting (Denver)*.

Métivier, L., Brossier, R., Operto, S., & Virieux, J., 2015a. Acoustic multi-parameter FWI for the reconstruction of P-wave velocity, density and attenuation: preconditioned truncated newton approach, in *Expanded Abstracts, 85th Annual Meeting*, pp. 1198–1203, SEG.

Métivier, L., Brossier, R., Operto, S., & Virieux, J., 2015b. Acoustic multi-parameter FWI for the reconstruction of P-wave velocity, density and attenuation: preconditioned truncated Newton approach, in *Expanded Abstracts, 85th Annual SEG Meeting (New Orleans)*.

Métivier, L., Brossier, R., & Virieux, J., 2015c. Combining asymptotic linearized inversion and full waveform inversion, *Geophysical Journal International*, **201**(3), 1682–1703.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J., 2016a. Overcoming cycle skipping in FWI: An optimal transport approach, in *Expanded Abstracts, 78th Annual EAGE Meeting (Vienna)*.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J., 2016b. An optimal transport approach for seismic tomography: Application to 3D full waveform inversion, *Inverse Problems*, **32**(11), 115008.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J., 2016c. Increasing the robustness and applicability of full waveform inversion: an optimal transport distance strategy, *The Leading Edge*, **35**(12), 1060–1067.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J., 2016d. Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion, *Geophysical Journal International*, **205**, 345–377.

Métivier, L., Brossier, R., Oudet, E., Mérigot, Q., & Virieux, J., 2016e. An optimal transport distance for full-waveform inversion: Application to the 2014 chevron benchmark data set, in *SEG Technical Program Expanded Abstracts 2016*, pp. 1278–1283.

Métivier, L., Brossier, R., Operto, S., & J., V., 2017. Full waveform inversion and the truncated Newton method, *SIAM Review*, **59**(1), 153–195.

Meza-Fajardo, K. C. & Papageorgiou, A. S., 2008. A Nonconvolutional, Split-Field, Perfectly Matched Layer for Wave Propagation in Isotropic and Anisotropic Elastic Media: Stability Analysis, *Bulletin of the Seismological Society of America*, **98**(4), 1811–1836.

Montelli, R., Nolet, G., Dahlen, F. A., Masters, G., Engdahl, E. R., & Hung, S. H., 2004. Finite-frequency tomography reveals a variety of plumes in the mantle, *Science*, **303**, 338–343.

Mulder, W. & Plessix, R. E., 2008. Exploring some issues in acoustic full waveform inversion, *Geophysical Prospecting*, **56**(6), 827–841.

MUMPS team, 2015. *MUMPS - MUltifrontal Massively Parallel Solver users' guide - version 5.0.1 (July 23, 2015)*, ENSEEIHT-ENS Lyon, http://www.enseeiht.fr/apo/MUMPS/ or http://graal.ens-lyon.fr/MUMPS.

Nash, S. G., 2000. A survey of truncated Newton methods, *Journal of Computational and Applied Mathematics*, **124**, 45–59.

Noble, M., Gesret, A., & Belayouni, N., 2014. Accurate 3-d finite difference computation of travel time in strongly hetrogeneous media, *Geophysical Journal International*, **199**, 1572–1585.

Nocedal, J., 1980. Updating Quasi-Newton Matrices With Limited Storage, *Mathematics of Computation*, **35**(151), 773–782.

Nocedal, J. & Wright, S. J., 2006. *Numerical Optimization*, Springer, 2nd edn.

Nolet, G., 2008. *A Breviary of Seismic Tomography*, Cambridge University Press, Cambridge, UK.

O'Leary, D. P., 1980. The block conjugate gradient algorithm and related methods, *Linear algebra and its applications*, **29**, 293–322.

Operto, S., Ravaut, C., Improta, L., Virieux, J., Herrero, A., & Dell'Aversana, P., 2004. Quantitative imaging of complex structures from dense wide-aperture seismic data by multiscale traveltime and waveform inversions: a case study, *Geophysical Prospecting*, **52**, 625–651.

Operto, S., Virieux, J., Dessa, J. X., & Pascal, G., 2006. Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: application to the eastern Nankai trough, *Journal of Geophysical Research*, **111**(B09306), doi:10.1029/2005JB003835.

Operto, S., Virieux, J., Ribodetti, A., & Anderson, J. E., 2009. Finite-difference frequency-domain modeling of visco-acoustic wave propagation in two-dimensional TTI media, *Geophysics*, **74 (5)**, T75–T95.

Operto, S., Brossier, R., Gholami, Y., Métivier, L., Prieux, V., Ribodetti, A., & Virieux, J., 2013. A guided tour of multiparameter full waveform inversion for multicomponent data: from theory to practice, *The Leading Edge*, **Special section Full Waveform Inversion**(September), 1040–1054.

Operto, S., Brossier, R., Combe, L., Métivier, L., Ribodetti, A., & Virieux, J., 2014a. A frequency-domain seismic modeling engine for 3d visco-acoustic vti full waveform inversion of fixed- spread data, in *Expanded Abstracts, 76th Annual EAGE Meeting (Amsterdam)*.

Operto, S., Brossier, R., Combe, L., Métivier, L., Ribodetti, A., & Virieux, J., 2014b. Computationally-efficient three-dimensional visco-acoustic finite-difference frequency-domain seismic modeling in vertical transversely isotropic media with sparse direct solver, *Geophysics*, **79(5)**, T257–T275.

Operto, S., Miniussi, A., Brossier, R., Combe, L., Haller, N., Kjos, E., Métivier, L., Milne, R., Ribodetti, A., Song, Z., Virieux, J., & Zheng, Y., 2015a. Efficient 3d frequency-domain full-waveform inversion of ocean-bottom cable data: application to valhall in the visco-acoustic vti approximation, in *Expanded Abstracts, 77th Annual EAGE Meeting (Madrid)*.

Operto, S., Miniussi, A., Brossier, R., Combe, L., Métivier, L., Monteiller, V., Ribodetti, A., & Virieux, J., 2015b. Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic approximation, *Geophysical Journal International*, **202**(2), 1362–1391.

Pajot, B., Li, Y., Berthoumieux, V., Weisbecker, C., Brossier, R., Métivier, L., Thierry, P., Operto, S., & Virieux, J., 2014. A review of recent forward problem developments used for frequency-domain fwi, in *Expanded Abstracts, 76th Annual EAGE Meeting (Amsterdam)*.

Patera, A. T., 1984. A spectral element method for fluid dynamics: laminar flow in a channel expansion, *Journal of Computational Physics*, **54**, 468–488.

Pérez Solano, C., Donno, D., & Chauris, H., 2014. Alternative waveform inversion for surface wave analysis in 2-d media, *Geophysical Journal international*, **198**, 1359–1372.

Perona, P. & Malik, J., 1990. Scale-space and edge detection using anisotropic diffusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(7), 629–639.

Pinard, H., Dietrich, M., Garambois, S., Lavoué, F., Métivier, L., & Virieux, J., 2015. Simultaneous gpr reconstruction of electrical conductivity and permittivity, in *77th EAGE Conference and Exhibition-Workshops*.

Pinard, H., Garambois, S., Métivier, L., Dietrich, M., Sénéchal, G., & Rousset, D., 2016. Full-waveform inversion of gpr data acquired between boreholes in rustrel carbonates, in *Inter-Disciplinary Underground Science & Technology (i-DUST 2016), Avignon (France)*, vol. 12, E3S Web of Conferences.

Plessix, R. E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophysical Journal International*, **167**(2), 495–503.

Plonka, G. & Ma, J., 2008. Nonlinear regularized reaction-diffusion filters for denoising of images with textures, *IEEE Transactions on Image Processing*, **17**(8), 1283–1294.

Pratt, R. G., 1990. Inverse theory applied to multi-source cross-hole tomography. part II : elastic wave-equation method, *Geophysical Prospecting*, **38**, 311–330.

Pratt, R. G., 1999. Seismic waveform inversion in the frequency domain, part I : theory and verification in a physical scale model, *Geophysics*, **64**, 888–901.

Pratt, R. G. & Worthington, M. H., 1990. Inverse theory applied to multi-source cross-hole tomography. Part I: acoustic wave-equation method, *Geophysical Prospecting*, **38**, 287–310.

Pratt, R. G., Shin, C., & Hicks, G. J., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophysical Journal International*, **133**, 341–362.

Pratt, R. G., Huang, L., Duric, N., & Littrup, P., 2007. Sound-speed and attenuation imaging of breast tissue using waveform tomography of transmission ultrasound data, *Proc. SPIE*, **6510**, 65104S–65104S–12.

Prieux, V., Brossier, R., Gholami, Y., Operto, S., Virieux, J., Barkved, O., & Kommedal, J., 2011. On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study, *Geophysical Journal International*, **187**, 1495–1515.

Qi, Q. & Geers, T. L., 1998. Evaluation of the perfectly matched layer for computational acoustics, *Journal of Computational Physics*, **139**(1), 166–183.

Romero, L. A., Ghiglia, D. C., Ober, C. C., & Morton, S. A., 2000. Phase encoding of shot records in prestack migration, *Geophysics*, **65, (2)**, 426–436.

Roy, O., Zuberi, M. A. H., Pratt, R. G., & Duric, N., 2016. Ultrasound breast imaging using frequency domain reverse time migration, *Proc. SPIE*, **9790**, 97900B–97900B–9.

Saad, Y., 2003. *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia.

Salomons, E. M., Lohman, W. J. A., & Zhou, H., 2016. Simulation of sound waves using the lattice boltzmann method for fluid flow: Benchmark cases for outdoor sound propagation, *PLOS ONE*, **11**(1), 1–19.

Santambrogio, F., 2015. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, Progress in Nonlinear Differential Equations and Their Applications, Springer International Publishing.

Sava, P. & Biondi, B., 2004a. Wave-equation migration velocity analysis. i. theory, *Geophysical Prospecting*, **52**(6), 593–606.

Sava, P. & Biondi, B., 2004b. Wave-equation migration velocity analysis. ii. subsalt imaging examples, *Geophysical Prospecting*, **52**(6), 607–623.

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

Schäfer, M., Groos, L., Forbriger, T., & Bohlen, T., 2014. Line-source simulation for shallow seismic data. part 2: full-waveform inversion - a synthetic 2-D case study, *Geophysical Journal International*, **198**(3), 1405–1418.

Schuster, G. T., Wang, X., Huang, Y., Dai, W., & C., B., 2011. Theory of multisource crosstalk reduction by phase encoded statics, *Geophysical Journal International*, **184**, 1289–303.

Shapiro, N. & Campillo, M., 2004. Emergence of broadband rayleigh waves from correlations of the ambient seismic noise, *Geophysical Research Letters*, **31**.

Shapiro, N. M., Campillo, M., Stehly, L., & Ritzwoller, M. H., 2005. High-Resolution Surface-Wave Tomography from Ambient Seismic Noise, *Science*, **307**(5715), 1615–1618.

Shin, C., Jang, S., & Min, D. J., 2001. Improved amplitude preservation for prestack depth migration by inverse scattering theory, *Geophysical Prospecting*, **49**, 592–606.

Shipp, R. M. & Singh, S. C., 2002. Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data, *Geophysical Journal International*, **151**, 325–344.

Sirgue, L. & Pratt, R. G., 2004. Efficient waveform inversion and imaging : a strategy for selecting temporal frequencies, *Geophysics*, **69**(1), 231–248.

Socco, L. V., Foti, S., & Boiero, D., 2010. Surface-wave analysis for building near-surface velocity models — established approaches and new perspectives, *Geophysics*, **75**(5), 75A83–75A102.

Swarztrauber, P. N., 1974. A Direct Method for the Discrete Solution of Separable Elliptic Equations, *SIAM Journal on Numerical Analysis*, **11**(6), 1136–1150.

Symes, W. & Kern, M., 1994. Inversion of reflection seismograms by differential semblance analysis: algorithm structure and synthetic examples, *Geophysical Prospecting*, **42**, 565–614.

Symes, W. W., 2008. Migration velocity analysis and waveform inversion, *Geophysical Prospecting*, **56**, 765–790.

Tago, J., Métivier, L., & Virieux, J., 2014. SMART layers: a simple and robust alternative to PML approaches for elastodynamics, *Geophysical Journal International*, **199**(2), 700–706.

Tape, C., Liu, Q., Maggi, A., & Tromp, J., 2009. Adjoint tomography of the southern california crust, *Science*, **325**, 988–992.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**(8), 1259–1266.

Tarantola, A., 2006. Popper, bayes and the inverse problem, *Nature Physics*, **2**, 492–494.

Thomsen, L. A., 1986. Weak elastic anisotropy, *Geophysics*, **51**, 1954–1966.

Thorpe, M., Park, S., Kolouri, S., Rohde, G. K., & Slepčev, D., 2016. A transportation l^p distance for signal analysis, *ArXiv e-prints*.

Thurin, J., Brossier, R., & Métivier, L., 2017a. Ensemble-based uncertainty estimation in full waveform inversion, in *79th EAGE Conference and Exhibition 2017, Paris*, p. Tu P1 07.

Thurin, J., Brossier, R., & Métivier, L., 2017b. An ensemble-transform kalman filter - full waveform inversion scheme for uncertainty estimation, in *87th SEG Conference and Exhibition 2017, Houston*.

Trinh, P. T., Brossier, R., Métivier, L., Tavard, L., & Virieux, J., 2017a. Efficient 3d elastic FWI using a spectral-element method, in *87th SEG Conference and Exhibition 2017, Houston*.

Trinh, P. T., Brossier, R., Métivier, L., Virieux, J., & Wellington, P., 2017b. Bessel smoothing filter for spectral element mesh, *Geophysical Journal International*, **209**(3), 1489–1512.

Trinh, P. T., Brossier, R., Métivier, L., Virieux, J., & Wellington, P., 2017c. Structure-smoothing Bessel filter for finite element mesh: Application on 3D elastic FWI, in *79th EAGE Conference and Exhibition 2017, Paris*.

Tromp, J., Tape, C., & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophysical Journal International*, **160**, 195–216.

Tsai, Y.-H. R., Chen, L.-T., Osher, S., & Zhao, H.-K., 2003. Fast sweeping algorithms for a class of Hamilton–Jacobi equations, *SIAM Journal on Numerical Analysis*, **41**(2), 673–694.

Uchide, T., Horikawa, H., Nakai, M., Matsushita, R., Shigematsu, N., Ando, R., & Imanishi, K., 2016. The 2016 kumamoto–oita earthquake sequence: aftershock seismicity gap and dynamic triggering in volcanic areas, *Earth, Planets and Space*, **68**(1), 180.

van Leeuwen, T., Gordon, D., Gordon, R., & Herrmann, F., 2012. Preconditioning the Helmholtz equations via row projections, in *Expanded Abstracts*, p. A002, EAGE.

Villani, C., 2003. *Topics in optimal transportation*, Graduate Studies In Mathematics, Vol. 50, AMS.

Villani, C., 2008. *Optimal transport : old and new*, Grundlehren der mathematischen Wissenschaften, Springer, Berlin.

Virieux, J., 1986. P-SV wave propagation in heterogeneous media: Velocity-stress finite difference method, *Geophysics*, **51**, 889–901.

Virieux, J. & Operto, S., 2009. An overview of full waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.

Virieux, J., Brossier, R., Métivier, L., Operto, S., & Ribodetti, A., 2016. Direct and indirect inversions, *Journal of Seismology*.

Virieux, J., Asnaashari, A., Brossier, R., Métivier, L., Ribodetti, A., & Zhou, W., 2017. An introduction to Full Waveform Inversion, in *Encyclopedia of Exploration Geophysics*, pp. R1–1–R1–40, eds Grechka, V. & Wapenaar, K., Society of Exploration Geophysics.

Waheed, U. B., Yarman, C. E., & Flagg, G., 2015. An iterative, fast-sweeping-based eikonal solver for 3D tilted anisotropic media, *Geophysics*, **80**, C49–C58.

Walford, M. E. R., 1964. Radio echo sounding through an ice shelf, *Nature*, **204**(4956), 317–319.

Wang, Y. & Rao, Y., 2009. Reflection seismic waveform tomography, *Journal of Geophysical Research*, **114**(B3), 1978–2012.

Warner, M. & Guasch, L., 2014. Adaptative waveform inversion - FWI without cycle skipping - theory, in *76th EAGE Conference and Exhibition 2014*, p. We E106 13.

Warren, C., Giannopoulos, A., & Giannakis, I., 2016. gprmax: Open source software to simulate electromagnetic wave propagation for ground penetrating radar, *Computer Physics Communications*, **209**, 163 – 170.

Wehner, D., Köhn, D., De Nil, D., Schmidt, S., al Hagrey, S., & Rabbel, W., 2015. A combined elastic waveform and gravity inversion for improved density model resolution applied to the Marmousi-II model, in *Proceedings of the* 77^{th} *EAGE Conference & Exhibition*, EAGE.

Weickert, J., 1998. *Anisotropic diffusion in image processing, Treubner Verlag, Stuttgart*, Treubner Verlag.

Yang, P., Brossier, R., Métivier, L., & Virieux, J., 2016a. Wavefield reconstruction in attenuating media: A checkpointing-assisted reverse-forward simulation method, *Geophysics*, **81**(6), R349–R362.

Yang, P., Brossier, R., Métivier, L., & Virieux, J., 2016b. Checkpointing-assisted reverse forward simulation: an optimal recomputation method for fwi and rtm, in *SEG Technical Program Expanded Abstracts 2016*, pp. 1089–1093.

Yang, P., Brossier, R., Métivier, L., & Virieux, J., 2016c. A review on the systematic formulation of 3D multiparameter full waveform inversion in viscoelastic medium, *Geophysical Journal International*, **207**(1), 129–149.

Yang, P., Brossier, R., Métivier, L., Virieux, J., & Zhou, W., 2017a. A second-order adjoint truncated newton approach to time-domain multiparameter full waveform inversion in viscoacoustic medium, in *Proceedings of the* 79^{th} *EAGE Conference & Exhibition*, EAGE.

Yang, P., Brossier, R., Métivier, L., Virieux, J., & Zhou, W., 2017b. A second-order adjoint truncated Newton approach to time-domain multiparameter full waveform inversion in viscoacoustic medium, *SIAM Journal on Scientific Computing*, p. submitted.

Yang, Y., Engquist, B., Sun, J., & Froese, B. D., 2016. Application of optimal transport and the quadratic wasserstein metric to full-waveform inversion, *ArXiv e-prints*.

Zhang, Y., Zhang, H., & Zhang, G., 2011. A stable TTI reverse time migration and its implementation, *Geophysics*, **76**(3), WA3–WA11.

Zhang, Y., Zhang, T., Ouyang, H., & Li, T., 2016. Efficient SPH simulation of time-domain acoustic wave propagation, *Engineering Analysis with Boundary Elements*, **62**, 112 – 122.

Zhu, H., Bozdağ, E., Peter, D., & Tromp, J., 2012. Structure of the european upper mantle revealed by adjoint tomography, *Nature Geoscience*, **5**, 493–498.