

UNIVERSITE DE NICE-SOPHIA ANTIPOLIS - UFR SCIENCES Ecole Doctorale
Sciences Fondamentales et Appliquées

THESE

pour obtenir le titre de

Docteur en Sciences
de l'Université de Nice-Sophia Antipolis

Spécialité : Sciences de l'Univers - Géophysique

Speed-up and regularization techniques for seismic full waveform inversion

Clara CASTELLANOS LOPEZ

Thèse dirigée par **Stéphane GAFFET** et **Stéphane OPERTO** préparée au
laboratoire Géoazur, Sophia-Antipolis

soutenue le 18 Avril 2014 devant le Jury

Hervé CHAURIS
Stéphane GAFFET
Josselin GARNIER
Ludovic MÉTIVIER
Guust NOLET
Stéphane OPERTO
René-Edouard PLESSIX
Jean VIRIEUX

Professeur MINES ParisTech
Directeur de Recherche CNRS
Professeur Paris VII
Chargé de Recherche CNRS
Professeur Université Nice-Sophia Antipolis
Chargé de Recherche CNRS
Chercheur Sénior SHELL
Professeur Université Joseph Fourier

Rapporteur
Directeur de Thèse
Examineur
Invité
Examineur
Co-directeur de Thèse
Rapporteur
Examineur

RÉSUMÉ

Les méthodes d'imagerie sismique ont pour objet d'estimer des propriétés du sous-sol à partir des ondes élastiques enregistrées au voisinage de la surface de la Terre. Parmi les méthodes sismiques, l'inversion des formes d'ondes complètes (FWI) est un problème d'optimisation numérique local visant à exploiter la totalité de l'information contenue dans les formes d'onde. Bien que la FWI fasse aujourd'hui partie intégrante des séquences de traitement en exploration sismique pétrolière, cette technologie pose encore de nombreux défis, en particulier en termes de résolution, de coût calcul et d'optimisation non-linéaire. Cette thèse explore plusieurs méthodes visant à réduire le coût de calcul et à améliorer les techniques de régularisation interfacées avec le problème d'optimisation.

Actuellement, le principal obstacle à la mise en oeuvre de la FWI élastique en trois dimensions sur des cas d'étude réalistes réside dans le coût de calcul associé aux tâches de modélisation sismique. Pour surmonter cette difficulté, je propose deux contributions. Tout d'abord, je propose de calculer le gradient de la fonctionnelle avec la méthode de l'état adjoint à partir d'une forme symétrisée des équations de l'élastodynamique formulées sous forme d'un système du premier ordre en vitesse-contrainte. Cette formulation auto-adjointe des équations de l'élastodynamique permet de calculer les champs incidents et adjoints intervenant dans l'expression du gradient avec un seul opérateur de modélisation numérique. Le gradient ainsi calculé facilite également l'interfaçage de plusieurs outils de modélisation avec l'algorithme d'inversion.

Le coût de calcul de la FWI est proportionnel au nombre de sources de l'expérience sismique ($> 10^3$ en 3D). Une approche possible pour réduire ce coût consiste à effectuer la modélisation sismique pour une combinaison linéaire des sources plutôt que de traiter chaque source indépendamment. Les facteurs aléatoires utilisés dans cette combinaison linéaire ont pour fonction de réduire au cours des itérations les artefacts créés par cet assemblage artificiel. L'encodage de sources a été abondamment utilisé avec des méthodes d'optimisation du premier ordre tels que des algorithmes de plus grande pente ou de gradients conjugués. J'explore dans cette thèse dans quelle mesure son utilisation dans des algorithmes d'optimisation du second-ordre de quasi-Newton et de Newton tronqué permettait de réduire encore le coût de la FWI. J'ai évalué cette méthodologie à l'aide de cas d'études synthétiques et réels 2D en utilisant une méthode de modélisation sismique en domaine fréquentiel fondée sur l'utilisation de solveurs directs. Néanmoins, les résultats de mon étude peuvent s'extrapoler à des modélisations sismiques 3D en domaine fréquentiel effectuées avec des solveurs itératifs.

Le problème d'optimisation associé à la FWI est mal posé au sens où les données ne permettent pas de déterminer un modèle du sous-sol unique, nécessitant ainsi d'ajouter des contraintes de régularisation à la fonctionnelle à minimiser. Le type de régularisation généralement utilisée en imagerie sismique consiste à chercher le modèle le plus lisse possible permettant d'expliquer les données. Je montre ici comment une régularisation fondée sur la variation totale du modèle fournissait une représentation adéquate des modèles du sous-sol en préservant le caractère discontinu des interfaces lithologiques. Pour améliorer la qualité des images du sous-sol, je propose également un algorithme de débruitage fondé sur une variation totale locale au sein duquel j'incorpore l'information structurale fournie par une image migrée pour préserver les structures de faible dimension dans l'image.

Mots clés : inversion de forme d'onde, FWI, encodages des sources, optimisation numérique non linéaire, gradient stochastique, régularisation, variation totale, débruitage, méthodes de Newton.

ABSTRACT

Seismic imaging methods allow to reconstruct the earth’s subsurface parameters based on partial measurements of elastic waves at, or near, the surface. Full waveform inversion (FWI) is a numerical optimization problem that uses the whole waveform information of all arrivals to update the subsurface parameters that govern seismic wave propagation. FWI has shown to provide high quality images and now constitutes a production imaging tool in reservoir exploration. However, FWI still faces many challenges, specially in the topics of the image resolution, computational cost and non-linear optimization. This thesis concerns some aspects to reduce the computational cost and the use of regularization techniques in the optimization problem.

Currently, the main limitation to perform 3D elastic full waveform inversion on a production level is the computational cost it represents. With this in mind, we provide two contributions. First, we develop a self adjoint formulation of the isotropic first order velocity-stress elastic equations that allow to implement only one forward modeling operator in the gradient computation. Second, solving the forward problem is the most computationally expensive part of FWI, and its cost is proportional to the number of sources ($> 10^3$ in 3D). To gain efficiency, instead of solving the wave equation for each source, it is possible to solve the wave equation for a linear combination of the sources. This is known as source encoding, and has been widely used when combined with steepest descent or conjugate gradient algorithms in the optimization process. With the purpose of reducing even more the computational cost, we combine Newton and quasi-Newton optimization methods with source encoding techniques. We implement this in a 2D frequency domain acoustic modeling engine based on frequency, but the results are easily extendible to the 3D frequency scenario. Our synthetic numerical tests were carried out in the BP-2004 salt model which is a realistic configuration inspired by the Gulf of Mexico, and the real data we used is an ocean bottom cable dataset recorded from the Valhall field in the North Sea. We find that the lowest computational cost needed to attain a predefined misfit value is provided by l -BFGS, with periodic restarts. However, with noisy data, the most accurate and robust direction of descent is provided by Newton methods, with and without source encoding.

The optimization process requires regularization constraints because the model is not entirely constrained by the data, and more than one model may fit the observed data equally well. We see that the total variation of the model as a regularization term provides an adequate description of earth models. To improve the quality of the images of the earth parameters, we propose a local total variation denoising algorithm based on the incorporation of the information provided by a migrated image.

Key words : full waveform inversion, FWI, source encoding, optimization, stochastic optimization, non-linear optimization, regularization, total variation, TV denoising, Newton methods

REMERCIEMENTS

Je ne pourrais pas résumer toute la gratitude que j'ai pour toutes les personnes qui m'ont accueilli, appris, guidé, et partagé des moments avec moi pendant ces années de thèse.

Je tiens tout d'abord à remercier profondément Stéphane Operto, qui m'as guidé dans mes travaux de recherche. J'apprécie énormément les temps qu'il a dédié pour me transmettre toutes ses connaissances, sa disponibilité pour discuter, et son implication dans mon travail. Je le remercie aussi d'avoir rendu ce manuscrit beaucoup plus lisible et de m'avoir aidé pour la rédaction en français. Mes remerciements vont également à Stéphane Gaffet avec qui j'ai commencé à travailler en géophysique pour la première fois lors du master, et qui m'a encouragé à faire une thèse. Je voudrais lui exprimer tout ma reconnaissance pour sa sympathie et ses conseils.

Je voudrais remercier Hervé Chauris, professeur de MINES ParisTech, et René-Eduard Plessix, chercheur principal à Shell, pour avoir accepté d'être les rapporteurs de ce manuscrit. Mes remerciements vont également à Josselin Garneier, Professeur à Paris VII, Ludovic Metivier, chargé de recherche CNRS, Guust Nolet, Professeur à Géoazur, et Jean Virieux, professeur à l'Université Joseph Fourier, pour m'avoir accordé l'honneur de faire partie du jury de ma thèse. Je voudrais exprimer toute ma gratitude à Guust et Jean, avec qui cela a toujours été un vrai bonheur pour moi de partager une discussion.

J'ai eu la chance de faire cette thèse dans un cadre très agréable grâce à l'ensemble du groupe SEISCOPE. Je remercie Ludovic Metivier avec qui j'ai eu le plaisir de travailler pendant ma thèse, Romain Brossier qui m'as toujours aidé à sortir des moments de confusion et Alessandra Ribodetti pour ses explications et sa sympathie en tout moment. A tous les membres de Seiscope actuellement à Geoazur ou déjà partis et avec qui j'ai partagé beaucoup des moments dans les congrès, les conférences et réunions : Yaser Gholami, Vincent Etienne, Vincent Prieux, Stephen Beller, Laure Combe et Vadim Montellier. Un remerciement aussi aux thésards de l'équipe Seiscope à Grenoble avec qui j'ai pu partager des moments : Amir Asnaashari, Isabella Massoni, François Lavoue.

J'ai eu aussi l'occasion de partager des discussion et des réunions avec l'équipe Globalseis. Un grand merci à Laurent Stehly, Dylan Mikesell et Jean Charlety qui ont aussi contribué à mes travaux de recherche.

Je tiens à remercier aussi les professeurs du master Mathmods pour leur sympathie et investissement dans notre formation. Je pense surtout à Bruno Rubino, Pierre- Emmanuel Jabin, Victorita Dolean, Chiara Simeoni et Etienne Tanré.

Chaque jour de ces années de thèse ont été spécialement très agréables grâce à tous les thésards avec qui on partage les choses qui marchent et qui ne marchent pas, les débats à midi, les hauts et les bas. Merci à Clément, Yaser, Alain, Dlung, Nestor, Quentin, Juan Carlos, Maëlle, Victorien, Stephen, Benoit, Flore, Marianne, Eduard, Imane et Alice et Maurin (même si vous n'êtes pas des thésards). Je tiens à remercier spécialement mes collègues avec qui j'ai partagé le bureau : Clément (mon prof de culture française), Alain (c'est trop cool de partager les astuces sur Matlab, linux, bash, etc) et Dlung (ta patience est contagieuse). J'ai appris beaucoup des choses grâce à vous : du stress drop, des failles, des satellites et orbites, jusqu'à la vie française. Keep calm and ...

Un grand merci à toutes les personnes qui font de la vie au laboratoire très agréable et qui sont toujours prêt à donner un coup de main. Je pense surtout à Jenny Trévisan, Jelena Giannetti, Caroline Ramel, Julien Olivier, Lionel Maurino, Valerie Mercier, Arielle Willm (OCA), Laure Miniussi (OCA) et Alain Miniussi (OCA).

Il y a une personne avec qui j'ai commencé cette aventure de la thèse et finalement on est arrivé à la fin. Camilo, merci d'être ma constante.

Mommy, Daddy, Juli, Pao : Gracias infinitas por mantenerme cerca a ustedes a pesar que estamos tan lejos.

Merci du fond du cœur à toutes les personnes à Géoazur avec qui j'ai partagé d'inoubliables moments ces dernières années. C'était un vrai bonheur d'apprendre une langue, une culture et un métier parmi vous.

Clara

"On se dit que le bon temps passe finalement comme une étoile filante..."

RESUMÉ ETENDU

Connaître la composition de l'intérieur de la Terre est d'un intérêt particulier pour de nombreuses applications académiques et industrielles. Cette thèse porte sur l'étude de la méthode d'inversion des formes d'ondes complètes (**FWI**)

$$\min_m \phi_0 = \min_m \|Pu(m) - d\|_2^2, \quad (1)$$

où u est le champ d'onde modélisé, d représente les données enregistrées et m sont les paramètres décrivant les propriétés du sous-sol que l'on cherche à imager. L'opérateur de projection P qui extrait la valeur du champ d'onde modélisé aux positions de l'espace où les données ont été enregistrées. Le champ d'onde modélisé est solution de l'équation d'onde. Par exemple, dans le cas de l'approximation acoustique, u est solution de

$$\left(\nabla^2 - \frac{1}{v^2} \frac{\partial^2}{\partial t^2} \right) u(x, t) = s(x, t) \quad (x, t) \in \Omega \times [0, \infty), \quad (2)$$

où $v(x) = 1/\sqrt{\rho(x)\kappa(x)}$, κ est le module d'incompressibilité, ρ est la densité, v est la vitesse de propagation des ondes P. Ici, les paramètres m représentent aux différentes positions de l'espace une seule propriété physique, la lenteur au carré: $m = \{1/v^2(x)\}$. La résolution de l'équation (2) est dénommée problème direct, et la résolution de l'équation (1) est le problème inverse. La FWI nécessite la résolution de ces deux problèmes de manière alternée. Le problème direct fournit u pour un modèle initial m donné, et ce champ d'onde u est utilisé dans un deuxième temps par le problème inverse pour mettre à jour le modèle du sous-sol. En d'autres termes, le problème direct consiste à résoudre une équation aux dérivées partielles (EDP), et le problème inverse estime les paramètres contenus dans les coefficients de cette EDP à partir de ses solutions. Ce processus est répété de manière itérative jusqu'à ce que les valeurs du champ d'onde u propagé dans le modèle m aux positions des capteurs coïncident avec les données enregistrées d .

Lorsque la totalité du champ d'onde est injectée dans le processus d'optimisation, la FWI a potentiellement la capacité de fournir des modèles du sous-sol plus réalistes en termes de résolution et de caractérisation physique que les autres méthodes d'imagerie. Ses fondements théoriques ont été établis en géophysique dans les années quatre vingt (Lailly, 1983; Tarantola, 1984a), et depuis la compréhension de ses potentialités et de ses limites s'est affinée au gré des applications réalisées grâce aux moyens modernes d'acquisition et de calcul haute performance (Virieux and Operto, 2009). Les challenges méthodologiques à surmonter pour une mise en œuvre pertinente de la FWI ont néanmoins été clairement posés dès les premières études publiées sur le sujet. Par exemple, les premières analyses sur des cas synthétiques indiquent que le coût calcul de la FWI constitue un frein important à son

application à des cas d'étude réels (Gauthier et al., 1986; Crase et al., 1990; Luo and Schuster, 1991; Crase et al., 1992). Plus fondamentalement, contrairement à la plupart des autres méthodes d'imagerie, la FWI est un problème inverse non linéaire et les premiers tests numériques montrent que le processus d'optimisation non linéaire échoue souvent car la fonctionnelle à minimiser ϕ_0 est fortement non convexe faisant converger l'inversion vers un minimum local inacceptable quand le modèle initial est trop éloigné du modèle réel (Gauthier et al., 1986; Mora, 1989; Luo and Schuster, 1991). Bien que la FWI soit supposée reconstruire un large spectre de longueurs d'onde du modèle à partir des champs d'ondes réfléchis et transmis, les premiers essais (Devaney, 1984; Wu and Toksöz, 1987; Mora, 1988, 1989) ont uniquement fourni une image de la réflectivité c'est-à-dire des courtes longueurs d'onde tel que l'aurait fourni une méthode de migration. Cela motiva plusieurs analyses de résolution de la FWI pour comprendre quelle partie du spectre des nombres d'onde dans l'espace des modèles pouvait être reconstruit pour une géométrie d'acquisition et pour une bande passante de source (Jannane et al., 1989; Mora, 1989). Ces difficultés furent identifiées lors des investigations pionnières de Gauthier et al. (1986); Wu and Toksöz (1987); Mora (1988, 1989); Luo and Schuster (1991), et restent aujourd'hui des sujets de recherche d'actualité en FWI. Dans cette thèse, j'aborderai certains aspects de la FWI portant sur la réduction de son coût calculatoire et de sa non linéarité.

Bien que la puissance de calcul soit devenue suffisante dans les années 80 pour résoudre l'équation d'onde pour des modèles du sous-sol d'hétérogénéité arbitraire, par des méthodes de différences finies dans le domaine temps-espace notamment, les applications de la FWI sur des cas d'étude réels restaient hors de portée. Un premier pas ayant permis de réduire considérablement le coût de la FWI a résidé dans l'utilisation de la méthode de l'état adjoint pour calculer le gradient de la fonctionnelle (Lailly, 1983; Tarantola, 1984a). Une deuxième étape vers la réduction du coût de la FWI a été franchie en formulant la FWI dans le domaine fréquentiel. Dans les années 90s, le formalisme de la FWI, tel qu'élaboré par Tarantola (1984a), a été transformé dans le domaine fréquentiel pour les équations d'ondes acoustique et élastique (Pratt and Worthington, 1990; Pratt, 1990; Pratt and Shipp, 1999). L'équation d'onde dans le domaine fréquentiel est une forme généralisée de l'équation d'Helmholtz pouvant s'écrire sous forme matricielle comme

$$A(m, \omega)u(s, m, \omega) = s(\omega), \quad (3)$$

où les coefficients de $A(m, \omega)$, matrice résultant de la discrétisation de l'opérateur de l'équation d'onde en domaine fréquentiel. Ce système linéaire peut être résolu par des méthodes directes ou itératives. Les méthodes directes reposent sur des techniques d'élimination de Gauss telles que les approches fondées sur des décompositions triangulaires supérieure/inférieure de la matrice ($A = LU$). La factorisation LU de la matrice dans l'équation (3) dépend uniquement de m et de la fréquence ω . Par conséquent, pour chaque itération non-linéaire du problème inverse, une seule factorisation LU est effectuée par fréquence pour calculer le gradient. Les champs d'onde calculés pour chacune des sources sont efficacement calculés à partir des facteurs LU par substitutions directe et inverse. Les limites des approches fréquentielles fondées sur des solveurs directs résultent du calcul et du stockage des facteurs LU. Pour des applications élastiques 3D, l'approche par solveur direct semble aujourd'hui inaccessible en raison de la nature vectorielle de l'équation d'onde élastique nécessitant le calcul d'au moins trois composantes de vitesse particulières et des faibles longueurs d'ondes propagées en relation avec la vitesse de propagation des ondes de cisaillement. Dans cette configuration, seules les modélisations en domaine fréquentiel avec des solveurs itératifs ou des modélisations en domaine temporel semblent possibles.

Une réduction du coût calcul est envisageable si le nombre de problèmes directs est diminué. Comme pour des approches fréquentielles fondées sur des solveurs itératifs ou des modélisations temporelles, le coût des problèmes directs est proportionnel au nombre de sources (équations (2) et (3)), la manière la plus naturelle de réduire le coût est de diminuer le nombre de sources. Une méthode courant d'accélération, appelée encodage des sources, consiste à effectuer une acquisition sismique conventionnelle (en émettant une source à la fois). Néanmoins, au lieu de résoudre le problème direct, (2) ou (3), pour chaque source, des super-sources sont formées par combinaison linéaire des sources pondérées par

des facteurs aléatoires (Romero et al., 2000; Neelamani et al., 2008; Krebs et al., 2009; Ben Hadj Ali et al., 2011; Schuster et al., 2011; van Leeuwen et al., 2011; Huang and Schuster, 2012). Pour encoder les données de la sorte, il est nécessaire que les sources soient enregistrées par le même ensemble de récepteurs, désignant ainsi les acquisitions d’extension fixe (fixed-spread en Anglais) comme les acquisitions de fond de mer. Le volume de données généré par N_s sources et enregistré par N_r récepteurs est compressé en une seule collection de traces à tir commun.

Contributions de cette thèse

Généralement, le gradient de la fonction coût de la FWI est implémenté avec la méthode de l’état adjoint à partir de l’équation d’onde du second ordre auto-adjointe. En revanche, l’équation d’onde formulée sous forme d’un système hyperbolique d’ordre 1 en vitesse-contrainte ne conduit pas à une forme auto-adjointe ($A \neq A^\dagger$), nécessitant une implémentation dissociée de l’opérateur direct (équation d’état) et de l’opérateur adjoint. Ma contribution a consisté à développer un formalisme permettant d’exprimer les équations de l’élastodynamiques du premier ordre sous forme d’un opérateur auto-adjoint facilitant ainsi grandement l’implémentation du gradient. Ceci est implémenté via un changement de variable appliqué aux contraintes normales, fourni par la décomposition en vecteurs propre de la matrice de raideur. Ce changement de variable permet de reformuler l’équation d’onde sous une forme pseudo-conservative au sein de laquelle les paramètres du milieu sont regroupés dans une matrice diagonale factorisée aux dérivés temporelles du système hyperbolique (Castellanos et al., 2011).

Bien que les méthodes d’encodage de sources aient été beaucoup utilisées en FWI ou en migration par renversement temporel (Neelamani et al., 2008; Krebs et al., 2009; Ben Hadj Ali et al., 2011; Schuster et al., 2011; van Leeuwen et al., 2011; Huang and Schuster, 2012), elles ont principalement été combinées avec des algorithmes d’optimisation de gradient, reproduisant ainsi les principes des méthodes de gradient stochastiques (Robbins and Monro, 1951; Spall, 2003). Dans le but de réduire encore le coût des calculs et améliorer le facteur d’accélération fourni par les méthodes d’encodage j’ai combiné des algorithmes d’optimisation de quasi-Newton et de Newton avec de l’encodage de source. Dans le cas des données réelles du champ de Valhall, l’inversion avec différentes méthodes d’optimisation ne convergent pas vers le même minimum local. Ceci résulte probablement du fait que, au voisinage du modèle initial, la fonction coût contient plusieurs minimum locaux. L’estimation de l’accélération fournie par l’encodage de sources et les méthodes d’optimisation du second d’ordre est donc biaisée car la qualité des modèles finaux n’est pas identique. Je montre que, avec et sans encodage de sources, les méthodes de Newton suivent la direction de descente qui converge vers les modèles du sous sol pour lesquels la fonction coût a la valeur la plus faible. Pour une meilleure compréhension des artefacts d’interférence générés par l’encodage des sources, une estimation est donnée de la variance du gradient avec encodage de sources pour des données sans bruit à partir de trois types d’assemblages de sources. L’analyse de ces variances peut être utilisée pour concevoir des stratégies optimisées d’encodage de sources. Une partie des résultats présentés ont été soumis pour publication dans Castellanos et al. (2013).

Beaucoup de travaux ont été consacrés à l’exploration de nouvelles normes dans l’espace des données (Djikpéssé and Tarantola, 1999; Guitton and Symes, 2003; Ha et al., 2009; Pyun et al., 2009; Brossier et al., 2010), mais seulement récemment l’influence de la norme utilisée dans l’espace des modèles (la régularisation) a été analysée (Burstedde and Ghattas, 2009; Ramírez and Lewis, 2010; Anagaw and Sacchi, 2012; Guitton, 2012). Je compare dans cette thèse deux normes dans la régularisation, la norme l_2 et la variation totale (TV) à partir d’un test synthétique réaliste (le modèle BP-2004 salt) et les données d’OBC enregistrées sur le champ de Valhall. En l’absence de bruit, le modèle final obtenu avec la norme TV est considérablement meilleur. Pour le cas d’étude réel, les modèles de vitesse finaux obtenus avec les normes l_2 et la TV sont comparables. Néanmoins, la stratification induite par les couches géologiques est mieux restituée dans les modèles construits avec la régularisation par variation totale.

L'algorithme de débruitage de Rudin-Osher-Fatemi (ROF) (Rudin et al., 1992) élimine le bruit d'une image en minimisant la variation totale de l'image, tout en maintenant l'image débruitée aussi similaire que possible à l'image initiale. Néanmoins, les limites du débruitage par TV résident dans le fait qu'il peut éliminer la texture (les structures de petites dimensions) dans l'image. Cela a motivé un certain nombre de travaux dédiés à la conception d'algorithmes de débruitage local par TV qui préserve la texture de l'image (Bertalmio et al., 2003; Vese and Osher, 2003). J'applique un débruitage local par TV où j'incorpore une information a priori sur la réflectivité fournie par une image migrée. L'algorithme de débruitage local par TV n'a aucune action aux positions de l'espace où la migration a positionné des réflecteurs et débruite les autres parties de l'image. L'algorithme de débruitage local par TV remplit alors sa mission tout en préservant les principaux réflecteurs.

CONTENTS

INTRODUCTION

1	INTRODUCTION	17
1.1	Full waveform inversion : the challenges of non-linear seismic imaging.	22
1.1.a	<i>Resolution</i>	23
1.1.b	<i>Computational Cost</i>	24
1.1.c	<i>Non-linear optimization</i>	26
1.2	Introduction to this thesis and description of main results	32
2	INTRODUCTION (FR)	39
2.1	L'inversion des formes d'ondes complètes : le défi de l'imagerie sismique non linéaire.	44
2.1.a	<i>Analyse de résolution</i>	46
2.1.b	<i>Coût de calcul</i>	47
2.1.c	<i>Optimisation non linéaire</i>	49
2.2	Présentation de la thèse et description des principaux résultats	55

CHAPTER 1 FUNDAMENTALS OF INVERSE PROBLEM THEORY

0.3	Linear Inverse Problems.....	64
0.4	Non linear Inverse Problems.....	65
1	ILL POSED INVERSE PROBLEMS.....	66
2	OPTIMIZATION.....	70
2.1	Line Search	71
2.2	Search Directions.....	72
2.2.a	<i>Steepest Descent</i>	72
2.2.b	<i>Linear conjugate gradient</i>	72
2.2.c	<i>Newton methods</i>	73
2.2.d	<i>BFGS</i>	74
2.2.e	<i>Limited memory BFGS</i>	76
2.2.f	<i>Errors with the approximate Hessian, and restart procedures</i>	77

2.3	Preconditioner	78
-----	----------------------	----

CHAPTER 2 FULL WAVEFORM INVERSION

1	SEISMIC IMAGING WITH FULL WAVEFORM INVERSION	81
1.1	Full waveform inversion	81
1.2	The gradient	83
1.2.a	<i>Preconditioner</i>	85
1.2.b	<i>Physical interpretation of the gradient</i>	86
1.3	The Hessian	88
1.3.a	<i>Physical interpretation</i>	92
1.4	Solving FWI numerically in the frequency or time domain	96
1.5	Image resolution analysis	104

CHAPTER 3 SPEED-UP OF FWI WITH SOURCE ENCODING

1	INTRODUCTION	114
2	SPEED-UP FWI : STATE OF THE ART	119
3	METHOD	121
3.1	Full waveform inversion problem	121
3.2	Second-order optimization methods	122
3.3	Preconditioner	124
4	SOURCE ENCODING	125
4.1	Optimization algorithms with source encoding	126
5	NUMERICAL EXAMPLES	127
5.1	Synthetic example	129
5.1.a	<i>Illustration of the problem : Convergence rates of stochastic versus deterministic algorithms</i>	129
5.1.b	<i>Synthetic data without noise</i>	131
5.1.c	<i>Synthetic data with noise</i>	132
5.1.d	<i>Stochastic gradient</i>	134
5.1.e	<i>Summary of the BP-2004 salt experiment</i>	135
5.2	Real data example	136
6	ESTIMATION OF THE VARIANCE WITH SOURCE ENCODING, WITHOUT NOISE.	139
7	CONCLUSION	143
8	TABLES	145
9	FIGURES	147

CHAPTER 4 REGULARIZATION TECHNIQUES FOR FWI

1	REGULARIZATION	167
1.1	Tikhonov regularization	168
1.1.a	Choice of λ for the known noise level case	169
1.1.b	Choice of λ for the unknown noise level case	170
1.1.c	Example 1: No regularization and overfitting the data	170
1.1.d	Example 2: Regularization and use of the L-curve	171
1.1.e	Norm for the regularization term	171
1.1.f	Norm for the data	174
1.2	Total Variation Regularization	174
1.3	TV denoising algorithms for seismic imaging : A modified ROF method.	175
1.4	Numerical Implementation of regularization and denoising (l_2 and TV norms)	177
1.4.a	TV Regularization : numerical examples	178
1.4.b	Denoising : numerical examples	190
1.4.c	Conclusions	191
2	THE MODEL NULL SPACE AND THE SPECTRAL CONTENT OF THE DATA	198
2.1	Frequency response	198
2.2	Spectral content of the data in the BP-2004 Salt Model	206

CONCLUSIONS AND PERSPECTIVES

1	CONCLUSIONS AND PERSPECTIVES	214
2	CONCLUSIONS ET PERSPECTIVES (FR)	216

APPENDICES

1	IMAGING CONDITIONS FOR OTHER SEISMIC IMAGING METHODS	219
1.1	Time Reversal Mirror Imaging Condition	219
1.2	Imaging with noise cross correlations	224
1.3	Imaging condition with adjoint methods	225
1.3.a	Reverse time migration imaging condition	228
1.4	Time reversal imaging conditions : conclusions	228
2	TRAVEL TIME TOMOGRAPHY	230
3	THE GRADIENT OF THE MISFIT FUNCTION : COMPUTATION WITH THE ADJOINT STATE METHOD	233
3.1	The Adjoint State Method	233
3.1.a	Adjoint operators	233
3.1.b	Directional derivatives	234
3.1.c	The Lagrangian and the adjoint state equations	234
4	THE GRADIENT COMPUTATION FOR THE VELOCITY-STRESS ELASTODYNAMIC WAVE EQUATIONS WITHOUT ATTENUATION IN CONSERVATIVE FORM, WITH THE ADJOINT-STATE METHOD	236
4.1	Introduction	236
4.2	Forward Equations	237
4.2.a	Lagrangian and adjoint variables in the time domain	238

4.3	The Adjoint State Equations	240
4.4	The symmetric Pseudo-Conservative Form	241
4.4.a	<i>Adjoint State Equations for the Conservative Formulation</i>	243
4.4.b	<i>State equations</i>	244
4.4.c	<i>Adjoint state equations</i>	245
4.4.d	<i>Gradient of the misfit function</i>	246
4.4.e	<i>Algorithm</i>	247
5	THE HESSIAN : COMPUTATION WITH THE ADJOINT STATE METHOD	247
6	SOURCE ENCODING WITH DIRAC NOTATION	250
6.1	Misfit function	251
6.2	Gradient	253
	BIBLIOGRAPHY	255

INTRODUCTION

1 INTRODUCTION

In several scientific and industrial applications we are interested in knowing the composition of the earth's subsurface. For cost and feasibility reasons, it is not possible to directly observe the subsurface composition (as one would do, for example by drilling). This has driven the development of much cheaper and easily implementable approaches, generically known as seismic imaging techniques. They consist of different algorithms allowing to reconstruct the subsurface structure indirectly from physical measurements at the surface. The data can correspond to the recordings of seismic waves, electromagnetic waves, gravity anomalies, amongst others. The quality of the reconstruction of the earth's composition depends on the *imaging technique* used and ultimately on the *measured data* that are available. The purpose is therefore to reconstruct the subsurface parameters based on the measured data as illustrated in Figure 1. This thesis will focus on imaging techniques based on seismic recordings.

Data acquisition : transmission and reflection energy

On a global scale, earthquakes generate waves that travel through the earth's interior and the particle velocity of the earth's surface is measured on arrays of receivers around the earth as shown schematically in Figure 2a. Most of data used on global scale imaging is known as transmission data because the earthquakes (sources) generate waves that travel through the portion of the earth that needs to be imaged, and the *transmitted energy* is recorded on the other side¹. In reservoir exploration, used to find natural resources (hydrocarbons, gas, water), the perturbations in a delimited zone are created by controlled explosions either on land or marine environments. An example of a streamer acquisition is shown in Figure 2b. The sources are controlled explosions generated by moving boats. The sources and receivers move together and allow for the exploration of a wider area. The interest is to image the subsurface of the earth below the ocean bottom. When the sources and receivers are placed the surface of the earth (even if the boats are on the surface of the water), this is referred to as a surface acquisition. Other configurations, such as cross-hole acquisitions, schematically consist of at least two wells drilled

¹This does not mean that there are no reflected waves in the recorded data. Reflected waves are created between the surface and the shallow structures.

where $v(x) = 1/\sqrt{\rho(x)\kappa(x)}$, κ is the compressibility, ρ is the density, v is the wave velocity. Here, the model m consists of one physical parameters : $m = \{1/v^2(x)\}$. Solving (5) is known as the forward problem, and (4) is known as the inverse problem. Using FWI as an imaging technique is therefore an alternating process. The forward problem is solved to find u for a given model m , and then this wavefield u is used in the inverse problem to determine the model update. In other words, the direct problem consists in solving a partial differential equation (PDE), and the inverse problem estimates the parameters of the PDE. This process is repeated iteratively until the wavefield u propagated in the model m coincides with the observed data d . This work flow is schematically represented in Figure 4.

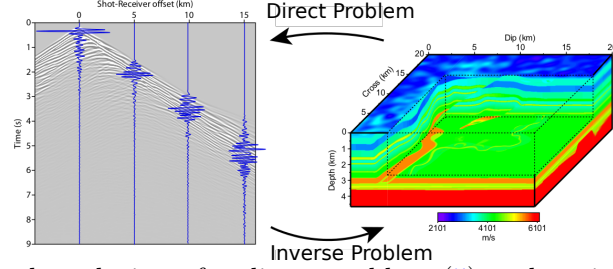


Figure 4: FWI iterates the solution of a direct problem (5) and an inverse problem (4). The direct problem allows to pass from the model to the data space, by using the current model m to solve the equation and find the wavefield u . The inverse problem allows to go from the data space to the model space by using the wavefield u to determine the model that minimizes a misfit ϕ_0 .

As the whole waveform information is used, FWI has the potential to be very accurate compared to other imaging techniques. Its theoretical fundamentals in the time domain were introduced in geophysics in the 1980's (Lailly, 1983; Tarantola, 1984a), and since then the comprehension has evolved and been used with success (Virieux and Operto, 2009). At the same time, FWI faces some important challenges. For example, early synthetic experimental results indicated that the **computational cost** of FWI represented a true limitation if one wished to apply it to real data (Gauthier et al., 1986; Crase et al., 1990; Luo and Schuster, 1991; Crase et al., 1992). Furthermore, unlike previous imaging techniques, FWI is a non-linear inverse problem and early numerical tests revealed that the **non-linear optimization** process often failed because ϕ_0 is not convex, by converging to unsatisfactory local minima when the initial model was far from the true model (Gauthier et al., 1986; Mora, 1989; Luo and Schuster, 1991). Although FWI is expected to update the transmission and the reflectivity information of the model, first attempts (Devaney, 1984; Wu and Toksöz, 1987; Mora, 1988, 1989) only provided the reflectivity and only resolved the high vertical wavenumbers of the model (similar to migration techniques). This prompted several authors to analyze the **resolution** power of FWI, and explore which part of the image wavenumber spectrum can be obtained in an inversion for a given acquisition geometry and source bandwidth (Jannane et al., 1989; Mora, 1989). Other setbacks appeared in multi-parameter inversion. Results showed that the resolution of the reconstructed parameters varies from one parameter to the next depending on the local aperture illumination resulting from the acquisition geometry, the cross-talk between parameters (interaction between parameters) and relative weight of the sensibility of different parameters in the data (Tarantola et al., 1984; Mora, 1987)². These disadvantages that were pointed out early in time (Gauthier et al., 1986; Wu and Toksöz, 1987; Mora, 1988, 1989; Luo and Schuster, 1991), currently remain as challenges and subjects of research in FWI.

²A parametrization is understood as a set of independent parameters that fully describe the model space.

In this PhD thesis we will address some aspects related to the computation costs and non-linearity questions. But, before presenting in more detail the scope of this work, let us briefly go over some details over each of the mentioned points.

1.1 Full waveform inversion : the challenges of non-linear seismic imaging.

Imaging condition

The *imaging condition* is the operator acting from the data space to the model space, that indicates the model parameters that have a sensitivity on the computed wavefield. Therefore, the imaging condition will reveal all the model parameters that, when subject to small variations, generate variations in the data. Consider a reflection wave arrival in the measured data d at time $t = T$, that is not present in the computed wavefield u . The purpose of the imaging condition is to determine where in the image space the model should be modified so that projected wavefield u at the receiver position will also contain the reflection arrival. For one source receiver pair (S-R), the only restriction is that the location x of the diffracting point in the subsurface should satisfy

$$t_{S,x} + t_{x,R} = T. \quad (6)$$

For an homogeneous background model of velocity v , all the points x that satisfy (6) form an *ellipse with focal points at the source and receiver positions*, like that shown in Figure 5a. This can be obtained by explicitly writing the time for the wave to travel from S to x , $t_{S,x}$, and time to travel from R to x , $t_{x,R}$, where D is the distance between the source and the receiver,

$$\begin{aligned} t_{S,x} &= \sqrt{(x/v)^2 + (z/v)^2} \\ t_{x,R} &= \sqrt{(D-x)/v)^2 + (z/v)^2}. \end{aligned}$$

Replacing this back into (6), leads to the following equation of an ellipse,

$$\sqrt{x^2 + z^2} + \sqrt{(D-x)^2 + z^2} = vT,$$

with focal points $f = 0, D$, major axis $a = vT/2 + D/2$ and minor axis $b = \sqrt{(vT)^2 + D^2}/2$.

Summarizing, for one source-receiver pair and a reflection arrival at time T , any point in the ellipse could correspond to the true position of the diffracting point because they all satisfy (6). In the example in Figure 5a, $v = 2m/s$ the reflection arrival time is $T = 5s$, $x_S = 0$ and $x_R = 8m$, resulting in an offset of $D = 8m$. To eliminate the ambiguity and determine where on the ellipse the reflector is located, the sum of the ellipses of all source-receiver pairs in the model space will create constructive interference due to energy arriving in phase, and reveal the true reflector position as shown schematically in Figure 5b for a horizontal reflector.

The imaging condition (6) in the model space can be written in an equivalent way, for one source and all its corresponding receivers as,

$$g(x) = \frac{\partial \phi_0(u(x, t, m))}{\partial m} = \left(P \frac{\partial u}{\partial m} \right)^\dagger (Pu(x, t, m) - d(x, t)) \quad (7)$$

$$= - \int_0^T \frac{\partial^2 u(x, t, m)}{\partial t^2} \lambda(x, T - t, m) dt, \quad (8)$$

where g stands for gradient, u is the solution of the direct problem, $\lambda(T-t)$ is a back-propagated wavefield that is computed using the residual data at the receiver positions as sources. In essence, the imaging condition is proportional to the correlation of the direct and back-propagated

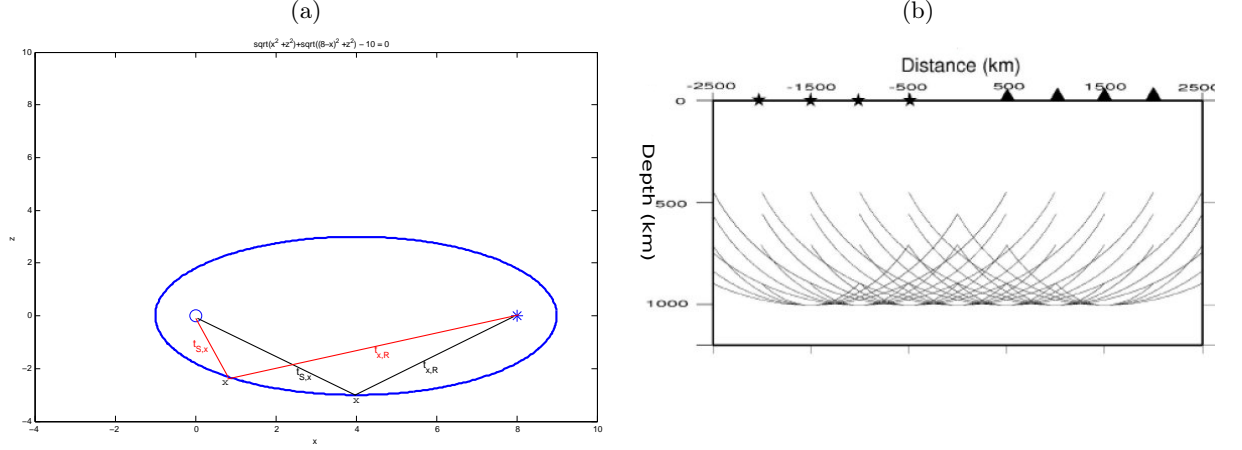


Figure 5: a) Example of the points satisfying the imaging condition $t_{S,x} + t_{x,R} = T$, for a reflection residual at time $T = 5s$, in a model with $v = 2m/s$, with a source at $x_S = 0$ and a receiver at $x_R = 8m$, resulting in an offset of $D = 8m$. The points form an ellipse with focal points at the source and receiver positions. b) Figure from Agudelo (2005). Imaging a horizontal reflector. Intersection of ellipses create coherent interference, and an image is created at the depth of the reflector.

wavefield on all points of the domain, integrated over time. By applying a Fourier transform, the gradient can also be expressed in the frequency domain,

$$g(x) = \int_{-\infty}^{\infty} \omega^2 u(x, \omega, m) \lambda^*(x, \omega, m) d\omega. \quad (9)$$

More details about the interpretation of the imaging condition can be found in Chapter 2, section 1.2.b. The imaging condition of FWI is related to the imaging condition of other seismic imaging techniques, as is explained in more Appendix 1, via the time reversal procedure.

1.1.a Resolution

On a theoretical level, the quality of the images of the earth parameters provided by FWI will supersede those provided by previous imaging techniques because the full waveform content is used in the inversion process and because the model used to solve the forward problem (5) is updated in every iteration. The breakthrough with respect to previous techniques was due to the way the wave equation is solved. Without sufficient computational power, the wave equation was solved analytically and asymptotically using *ray theory*, which required smooth models. It was therefore necessary to use a *linear approximation of the wave equation*, and separate the model m into a smooth background model m_0 and a small unknown model perturbation δm : $m = m_0 + \delta m$. Once this is done, the solution of the wave equation is found using first order perturbation theory, known as the *Born approximation*. The total wavefield u is approximated as $u(m_0 + \delta m) \approx u_0 + (\partial u / \partial m) \delta m$, where u_0 is the solution of the wave equation with a model m_0 and $(\partial u / \partial m) \delta m$ is the scattered wavefield by a model perturbation δm . Note that the Born approximation ignores higher order diffraction terms (i.e waves diffracted two times are not modelled). This distinction between m_0 and δm imposes a *scale separation between a smooth background model and the perturbation model*, as shown in Figure 6a,b. When the computational power became sufficient, the wave equation could be solved numerically using *finite differences* for an arbitrary complex model (Virieux, 1984). Without the need to linearize the wave equation, there was no longer need to make a scale separation between a smooth background model and

the perturbation model, and the waves were propagated in complex models containing both the smooth and reflectivity properties. If the acquisition system allows to measure both transmitted and reflected data and if the source has a sufficiently broad bandwidth, FWI is theoretically amenable to the imaging of a broad and continuous spectrum of wavenumbers. In practice, however, it has been difficult to update simultaneously both the transmission and reflection characteristics of the model. The first inversions only partially reconstructed the image (Devaney, 1984; Wu and Toksöz, 1987). The question asked at the time, and that still remains is, can the inversion algorithm retrieve the low and high wavenumbers simultaneously (Mora, 1989)? A horizontal and vertical resolution capacity of the data should be performed to estimate which wavenumbers in the image domain may be reconstructed. A more detailed explanation of the resolution analysis and limitations are explained in Chapter 2, section 1.5.

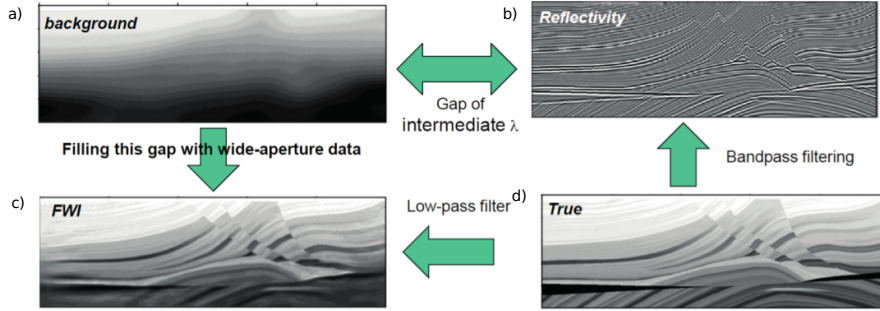


Figure 6: Figure d) shows the true velocity model m . In a linear approximation, the model m is separated in a smooth background model m_0 (Figure a) and a perturbation model δm (Figure b). FWI does not impose a scale separation if transmission and reflection data are available, a low pass version of the wavenumber image spectrum can be reconstructed (Figure c).

1.1.b Computational Cost

Although the computational power became sufficient in the 1980's to solve the wave equation for an arbitrary model with finite differences in the time domain, FWI was still out of reach for real data applications.

Adjoint method

A first step that dramatically decreased the computational cost was the use the *adjoint state method* to compute the gradient (Lailly, 1983; Tarantola, 1984a). Using equation (7), it is necessary to evaluate $\partial u / \partial m_i$ which are the partial derivatives of the computed wavefield with respect to *all* the model parameters, $i = 1, \dots, N \times N_p$. Here N represents the number of grid points in the discretization of the domain, and N_p the number of physical parameters to be reconstructed (ρ, v_p, v_s). It is possible to compute the partial derivatives using finite difference schemes, $\partial u / \partial m_i = (u(m + \Delta m_i) - u(m)) / \Delta m_i$, which represents a computational cost of $C = 2 \times N \times N_p \times N_s$, where N_s is the number of sources. Using the adjoint state method (Lailly, 1983; Tarantola, 1984a), the gradient is computed via equation (8). The computational cost is $C = 2 \times N_s$ because, for each source, one wave equation is solved to find u and another to find the back-propagated wavefield λ . The computation of the gradient with the adjoint state method is explained in Appendices 3,4.

Frequency domain

A next step in reducing the computational cost of FWI was taken by working in the *frequency domain*. In 1990 the equations for FWI (Tarantola, 1984a) were transformed to the frequency domain, for the acoustic and elastic forward problem (Pratt and Worthington, 1990; Pratt, 1990; Pratt and Shipp, 1999). Applying a Fourier transform to the forward problem (5), the discretization with finite differences (FD) is performed directly (the forward modeling operator in the frequency domain is sometimes referred to as Helmholtz operator). The resulting forward problem is

$$A(m, \omega)u(s, m, \omega) = s(\omega), \quad (10)$$

where $A(m, \omega)$ is the discretization of the wave equation operator in the frequency domain, and depends on the frequency $\omega = 2\pi f$. Equation (10) is a linear system where A is a matrix of dimensions $N \times N$, and u and s are vectors of dimensions $N \times 1$. This linear system can be solved with direct or iterative methods. With iterative methods, the number of iterations needed to find the solution depends on the conditioning of matrix, and therefore adequate preconditioners are needed (Erlangga, 2005; Plessix, 2007; Erlangga and Nabben, 2008). Each source function requires an independent iterative solution of the system. Direct methods to solve linear systems are based on Gauss elimination and include methods such as the $A = LU$ factorization of the matrix. The LU factorization of A in (10) only depends on m and the frequency ω . Therefore, for each iteration of the inverse problem, one LU factorization is performed per frequency. For all the source functions, the wavefield is easily found by forward and backward substitutions. If the number of frequencies is less than the number of sources, using the LU factorization to solve (10) allows to solve the direct problems efficiently because the LU factorization, which represents the highest computational effort in the solution of the forward problem, is only done once per source. In FWI this represents a considerable gain (Pratt and Worthington, 1990; Pratt, 1990) because in a typical 2D seismic experiment the number of sources is of the order of $10^2 - 10^3$ and in a 3D seismic experiment it is around $10^3 - 10^4$. Nonetheless, for a certain period it was thought that working in the frequency domain would not represent a computational gain in real applications, specially with short offset data, because all the frequencies had to be taken into account in the inversion, with $\Delta f \geq 1/2T_{max}$ (Freudenreich and Singh, 2000). Based on a resolution analysis of wave number coverage, it was shown that the number of frequencies needed were much less than those dictated by the Nyquist criteria and indeed FWI could be successful with a few frequencies (Sirgue and Pratt, 2004).

Besides reducing the number of forward problems to be solved, working in the frequency domain also has computational advantages in the inverse problem in terms of the memory required to store the wavefields needed to compute the gradient. Because of disk memory limitations, normally it is not possible to store the wavefields $u(x, t_i)$, $\lambda(x, t_i)$ on all the imaging domain ($x \in \Omega$) for all time steps ($t_i \in [1, \dots, N_t]$) required to evaluate the gradient (8), because the number of time steps N_t is large. Several solutions have been proposed, such as storing the wavefield $u(x, t_i)$ in memory for some specific values of t_i . The wavefield $u(x, t_i)$ which are stored in memory, are used as initial conditions to recompute the wavefields $u(x, t)$ for other times. On the other hand, if the inverse problem is performed in the frequency domain, the gradient is proportional to the multiplication of the direct and back-propagated field, for each frequency $u(x, \omega_i)$ and $\lambda(x, \omega_i)$, $\omega_i \in [1, N_f]$. Since $N_f \ll N_t$, generally the wavefields can be directly stored in memory to compute the gradient. More details can be found in Chapter 2, section 1.4.

Source encoding

The limitations of working in the frequency domain with direct solvers is that the LU factorization has to be stored in memory. With a finite difference grid, the matrix A is a sparse banded matrix. For 2D applications there are 3 bands and the distance between bands is N . For 3D applications there are five bands and the distance between bands is N^2 . Unlike A , the L

and U matrices are full and thus are memory demanding. For *3D acoustic* applications it may still be feasible to apply direct solvers (Operto et al., 2007). The first real data applications of 3D acoustic FWI the forward problem was solved in the frequency domain with iterative solvers (Plessix, 2009; Plessix and Perkins, 2010) or time marching algorithms (Sirgue et al., 2010). For *3D elastic* applications it seems almost certain that direct solvers may face difficulties. The reason is that finite difference schemes will require a minimum number of discretization points per wavelength to avoid numerical dispersion. Typically, in the earth $3000 \text{ m/s} < v_p < 5000 \text{ m/s}$, and $1300 \text{ m/s} < v_s < 3000 \text{ m/s}$. Therefore, for a fixed frequency $\lambda_s < \lambda_p$ and modeling shear waves will require a much finer grid, greatly increasing N in each dimension. The size of L and U are therefore too large to store in memory, and iterative solvers must be employed.

Computational gains have been attempted to reduce the number of the direct problems, which is often the most computationally expensive part of FWI. Note that with iterative or direct solvers the number of direct problems is proportional to the number of sources (equations (5) and (10)). However, when working with direct solvers, the computational burden is greatly relieved because the same LU factorization may be used for all sources. Therefore, to reduce the number of direct problems, the number of right hand sides of the direct problem have to be reduced. In *source blending* the acquisition of the data is modified. Several sources are shot simultaneously, or with some time delays between them, and the recorded data by the sensors therefore has the information of each source, and the interference effects between them (Beasley, 2008; Berkhout, 2008). Source blending reduces the time of collection data in the field and also the number of sources that are processed when solving the direct problem. Another popular speed-up technique, called *source encoding*, consists in performing a traditional seismic acquisition experiment using only one source at time, for N_s sources. However, instead of solving the direct problem (5) or (10) for each source, encoded super-sources \tilde{s} are formed by creating random linear combinations of the sources (Romero et al., 2000; Neelamani et al., 2008; Krebs et al., 2009; Ben Hadj Ali et al., 2011; Schuster et al., 2011; van Leeuwen et al., 2011; Huang and Schuster, 2012),

$$\tilde{s} = \sum_{i=1}^{N_s} \alpha_i s_i, \quad \tilde{d}_o = \sum_{i=1}^{N_s} \alpha_i d_{o_i}, \quad (11)$$

where α are random coefficients. The observed data is also encoded into \tilde{d}_o . To be able to encode the data in this way, it is necessary for all the sources to share the same set of receivers (referred as fixed-spread acquisitions). The volume of data produced by N_s sources and measured at N_r receivers is encoded into only one recorded data set at the receiver positions.

Even though source encoding with gradient descent algorithms reduces the computational cost per iteration (2 direct problems, instead of $2 \times N_s$), more iterations need to be performed because each model update is less accurate due to interference effects between sources (referred to as cross-talk), as illustrated in Figure 7a. In the end, there will be a computational gain (speed-up) if the number of direct problems solved to attain a predefined value of misfit function is less with source encoding than using the sources individually in the standard way, as illustrated in Figure 7b. Source encoding techniques have been applied successfully to real data sets (Baumstein et al., 2011; Routh et al., 2011; Bansal et al., 2013; Schiemenz and Igel, 2013).

1.1.c Non-linear optimization

Early applications of FWI showed that, although the accuracy of the forward problem was greatly improved and FWI seemed feasible, there were several additional difficulties with the optimization problem (Gauthier et al., 1986; Luo and Schuster, 1991). The misfit function (4) is non-linear with respect to m . This means that small changes in the model parameters can

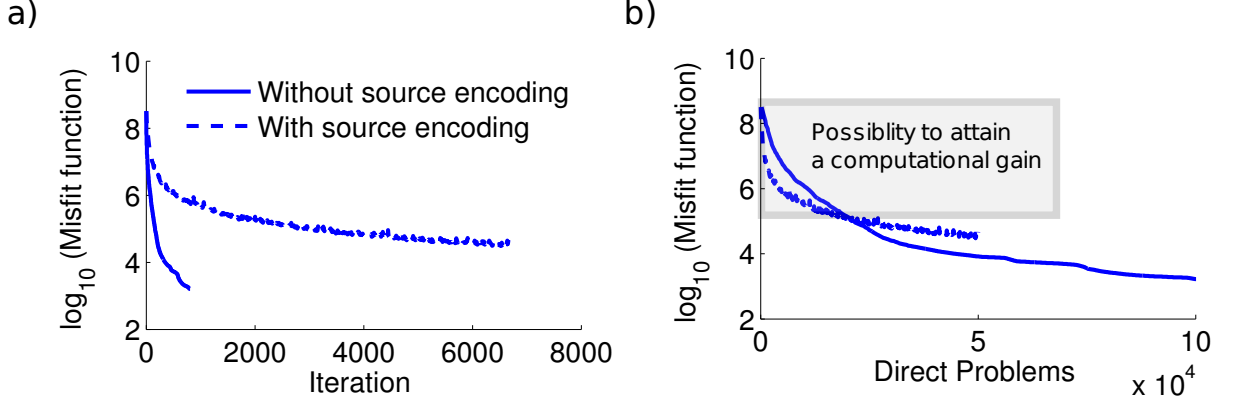


Figure 7: FWI inversion convergence curves with source encoding (dashed lines) and without source encoding (solid lines). (a) Comparison between the convergence rates. Less iterations are required without source encoding to attain the same misfit value. (b) Comparison of the computational efficiency (direct problems). If the reduction of the misfit function remains in the gray box, a computational gain can be attained because less direct problems are solved with source encoding than without. This example is presented in more detail in Chapter 3.

generate large changes in the misfit function ϕ_0 , and the variations occur in a non-linear fashion. In an acoustic 2D cross-well (transmission) data inversion [Luo and Schuster \(1991\)](#) compared travel time inversion and FWI. The numerical inversions showed that the resolution power of FWI is higher. However, travel time inversion was more successful because FWI converged to local minima.

Non-convex optimization

A numerical illustration of the difficulties in minimizing the l_2 norm of the waveform misfit (4) which is non-convex compared to a travel time misfit which is convex, is illustrated in Figure 8. Travel time tomography defines the misfit ϕ_0 as the difference between the observed and calculated arrivals times of the maximum of the wavefield

$$\min_m \phi_0 = \min_m \|T_c(m) - T_o\|_2^2, \quad (12)$$

where T_o represents the observed arrival time of the maximum and T_c represents the calculated arrival time (see Appendix 2 for a more precise description.). Consider a wavefield whose displacement can be described by $d(t) = e^{(t-2.5)^2} \sin(\omega t)$. For a low frequency (long period) $T = 2$ s, the wavefield is plotted in dark blue in Figure 8a. Imagine the computed wavefield $u(t)$ is equal to the true wavefield but arriving τ seconds in advance, $u(t) = d(t - \tau)$. The computed wavefield $u_c(t)$ for $\tau = 0.5$ s is plotted in dashed red. The maximum of the computed wavefield T_c and observed wavefield T_o are marked with light blue stars. Figure 8d shows the misfit function as a function of the delay time error τ in the computed wavefield. For the range of values plotted here, the travel-time misfit function (light blue) increases as the delay increases. As can be appreciated graphically, the optimization algorithm will be able to locate the minima. The waveform misfit (dark blue) shows that only if $\tau < 1$ s, a local optimization algorithm will provide the right solution. In particular, for $\tau = 0.5$ s which corresponds to the red wavefield 8a, the minimization will converge because the red star in Figure 8d shows that this point is in the valley of attraction. The optimization of the full waveform difference becomes more difficult as the frequency increases (period decreases). For Figure 8e where $T = 1$ s, the initial delay must satisfy $\tau < 0.5$ s to converge, and in Figure 8f where $T = 0.5$ s, the initial delay must satisfy $\tau < 0.25$ s to converge. As can be deduced from the figures, the condition

to avoid converging to secondary minima is that $\tau < T/2$. This means that the model should provide delays within the first maximum of the waveform. When this condition is not satisfied, *cycle-skipping* occurs. For example, for a frequency of $f = 2$ Hz and $\tau = 0.5$ s, Figure 8f reveals that using the l_2 norm of the waveform misfit, the inversion will remain at a local minimum, where cycle skipping occurred. The non-linearity therefore requires that the starting model be sufficiently accurate to be in the valley of the global minimum and avoid cycle skipping.

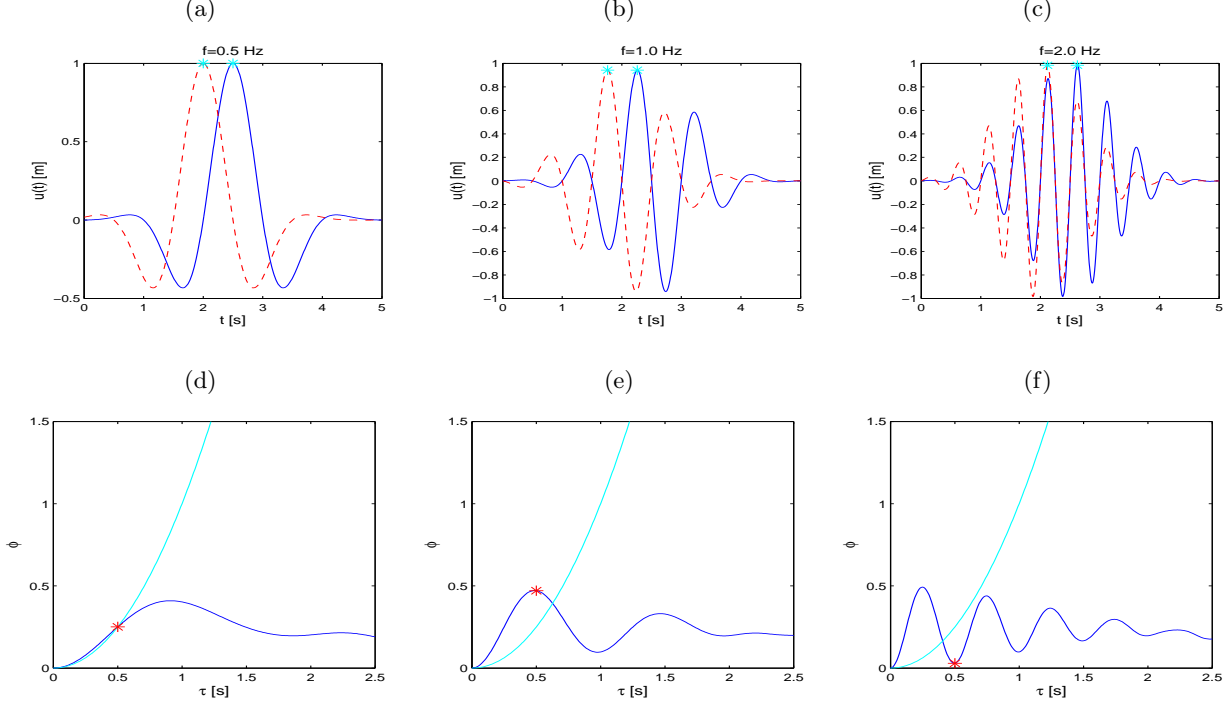


Figure 8: The first row plots the observed wavefield $d(t) = e^{(t-2.5)^2} \sin(2\pi ft)$ in blue, and the computed wavefield in red, $u(t) = d(t - \tau)$ for $\tau = 0.5$ s. The wavefields are plotted for a) $f = 0.5$ Hz b) $f = 1.0$ Hz, d) $f = 2$ Hz. The second row plots the travel time misfit function (12) in light blue as a function of the delay τ , and the full waveform misfit function (4) in dark blue. The misfit functions are plotted for d) $f = 0.5$ Hz e) $f = 1.0$ Hz, f) $f = 2$ Hz. For the values plotted here, the travel time misfit function is convex (light blue). The misfit function measuring the difference in the full waveform (dark blue) is non-convex. For a delay of $\tau = 0.5$ Hz, the optimization will be successful for $f = 0.5$ Hz but will converge to a secondary minima for $f = 2$ Hz. This is called cycle skipping.

Alternate definitions of the misfit function have been proposed to overcome this difficulty both in the time and in the frequency domain (Shin et al., 2002; Sheng et al., 2006; Shin and Min, 2006; Pyun et al., 2007; Shin et al., 2007; Shin and Ha, 2008; van Leeuwen et al., 2010; Hale, 2013). The purpose is to create misfit functions that are less-sensitive to cycle skipping, which is attained by creating misfit functions with a less oscillatory behaviour (as in Figure 8a), at the expense of losing information in the waveform (for example, use the envelope of the misfit function, time damping or cross correlating the wavefields). Alternate misfit functions may be particularly useful in the absence of low frequencies in the data, to avoid premature cycle-skipping and to obtain an initial model that will be then fed to FWI with an l_2 norm misfit function.

Hierarchical inversion

process challenging because the gradient of with respect to one parameter, contains the influence of other parameters. That is, a model perturbation for one parameter is in fact a linear contribution of the gradients with respect to all the parameters. In addition, the sensitivity of the data to each parameter $\partial u/\partial m$ may have different orders of magnitude. As a result, the inversion will only update the parameters with the highest partial derivatives. A re-parametrization of the physical parameters in the inversion, will help to rescale and decouple the physical parameters. A re-parametrization of m will modify $\partial A/\partial m$, known as the radiation pattern, present in the gradient expression. To reduce the cross talk between parameters, re-parametrizations can be searched such that the radiation patterns of different parameters have the smallest intersection possible. The Hessian will also help to decouple the effects and to rescale the magnitude of the parameters. For example, if ρ and v_p are the parameters to be reconstructed, the Gauss-Newton approximation of the Hessian will contain four main blocks with the cross-correlations $(\partial u/\partial v_p)^\dagger (\partial u/\partial v_p)$, $((\partial u/\partial v_p)^\dagger (\partial u/\partial \rho))$, $(\partial u/\partial \rho)^\dagger (\partial u/\partial v_p)$, $(\partial u/\partial \rho)^\dagger (\partial u/\partial \rho)$. Each block will have a different magnitude that will, theoretically, balance the magnitudes of different entries of the gradient so as to update both v_p and ρ . The Hessian will depend on the parametrization. Different parametrizations will provide different weights to the partial derivatives. An adequate parametrization will lead to a well conditioned Hessian, and each of the blocks will become similar in magnitude. In addition, in the same way that the Hessian corrected the smearing ambiguity artefacts and focused the energy on the reflector (Pratt et al., 1998), with multi-parameter inversion this will be useful to decrease the interaction between parameters, and refocus the energy on one parameter. An asymptotic (infinite frequency and Born approximation) study of the eigen-vectors of the Hessian, can help to determine if a parametrization is suitable and allows to decouple the physical parameters (Forgues and Lambaré, 1997; Plessix and Cao, 2011; Operto et al., 2013). Real data elastic anisotropic applications studying the effect of the parametrization have been performed (Gholami et al., 2013b,a; Prioux et al., 2013a,b).

Regularization of the inverse problem

The difficulties in the optimization process and the under-determination of the inverse problem require *regularization* terms to constrain the space of solutions and guide the inversion towards a final model that a geophysicist considers acceptable. For example, noisy models are discarded and the smoothest models are preferred. The regularization term can be included in the misfit function,

$$\phi = \phi_0 + \lambda \|R(m)\|_2^2 \quad (13)$$

$$= \|Pu_c(m) - d_o\|_2^2 + \lambda \|R(m)\|_2^2, \quad (14)$$

where $R(m)$ is the regularization term and $\lambda \|R(m)\|_2^2$ is a penalty term via a Lagrange multiplier. The optimization will seek to find the model that best explains the data *and* the regularization term simultaneously. Traditionally, smooth models are favored and thus $R(m) = \nabla m$ so as to minimize the gradient. If there is a-priori information about the values of the model at certain positions, this can also be included in $R(m)$ (Asnaashari et al., 2013). There have been very few works using other regularization terms or norms. For example, it is possible to change the norm and minimize the total variation (TV) of models (Ramírez and Lewis, 2010; Anagaw and Sacchi, 2012; Guitton, 2012),

$$\phi = \|Pu_c(m) - d_o\|_2^2 + \lambda \|\nabla(m)\|_1, \quad (15)$$

or minimize the number of coefficients of the model under a change of basis W (Candes and Romberg, 2005; Loris et al., 2010; Herrmann and Li, 2012),

$$\phi = \|Pu(m) - d\|_2^2 + \lambda \|W(m)\|_1. \quad (16)$$

Other regularization schemes have been proposed, where the regularization term is not added to the misfit function, but instead it is multiplied (Abubakar et al., 2002, 2004). Theoretically, multiplicative regularization has the advantage that the relative weights between the data gradient and the model gradient are naturally determined.

1.2 Introduction to this thesis and description of main results

The brief discussion of the main questions and challenges in FWI shows why this field is an active area of research since the 1980s. This thesis aims to add to the growing understanding of FWI by addressing some points concerning the computational cost and the non-linear optimization process.

We start by presenting in Chapter 1 the inverse problem theory and optimization algorithms for a general setting, without restraining ourselves to the specific case of seismic imaging. We introduce the concepts of linear and non-linear inverse problems, and ill-posed problems. As inverse problems are generally formulated as minimization problems, we introduce the optimization algorithms that we make reference to throughout this thesis. For the specific application of FWI, this chapter is not essential.

Chapter 2 complements the introduction to FWI. A physical interpretation is given to the two main objects that appear in the minimization algorithms: the gradient and the Hessian. For a small size problem we compute the full Hessian and we illustrate the action of one line on the inverse Hessian on the gradient to get an understanding on how the artefacts are corrected. The physical interpretation of the imaging conditions of other seismic methods that also make use of time reversal are briefly reviewed in Appendix 1. The equations for the gradient and the Hessian found in the literature with the adjoint-state method (Lions, 1972; Chavent, 2009; Plessix, 2006; Métivier et al., 2013a) are used here, and re-derived in Appendix 3 and 5.

Self-adjoint formulation of the elastic isotropic wave equation

Generally, in FWI, the gradient computation with the adjoint state method is implemented with a second order self-adjoint expression of the wave equation. For the elastic wave equation, the first-order velocity-stress formulation does not provide a self-adjoint formulation ($A \neq A^\dagger$), implying that the forward and adjoint modeling operator have to be implemented independently. Our contribution, presented in Appendix 4, lies in developing a formalism to recast the isotropic first order elastic velocity-stress equations in a self-adjoint fashion. This is done through linear transformations based on a change of basis into the eigen-vector space of the operators. This allows to compute the gradient with the only one implementation of modeling operator. This work can be found in Castellanos et al. (2011).

Source encoding with second order optimization methods

With iterative and direct solvers, the number of forward problems are proportional to the number of sources. With the LU factorization, the forward problem is solved more efficiently because the same factorization can be used for all sources, and the wavefield solution to each source is simply found by substitutions. However, to perform 3D elastic FWI in time or frequency domain with the current computational resources, time marching and iterative solvers appear to be the most feasible solution, and source encoding can greatly reduce the computational burden. Although source encoding techniques have been widely used (Neelamani et al., 2008; Krebs et al., 2009; Ben Hadj Ali et al., 2011; Schuster et al., 2011; van Leeuwen et al., 2011; Huang and Schuster, 2012), they have been mostly combined with gradient optimization algorithms,

resembling stochastic gradient algorithms (Robbins and Monro, 1951; Spall, 2003). With the purpose of reducing even more the computational cost and improving the gain (increasing the gap in the curves with and without source encoding in Figure 7), we combine quasi-Newton and Newton optimization methods with source encoding techniques, which had not been done in the context of FWI. In other application areas such as machine learning, this exploration has also recently begun (Schraudolph et al., 2007). The lack of convergence proofs of second-order stochastic methods (Bottou and Le Cun, 2005), and the lack of efficient formulations to compute the Newton descent directions (Métivier et al., 2013b, 2014) are some of the reasons. The equations for the efficient Hessian computation are shown in Appendix 5. However, the computational cost per iteration of the truncated Newton methods (Gauss-Newton and Full Newton) is higher per iteration, because the Newton descent direction requires the solution of an additional linear system. In the case of l -BFGS, the Hessian approximation needs to be periodically restarted. Therefore, a priori, it is not clear whether the computational savings can be further improved with second order optimization methods.

In Chapter 3, we predefine stopping criteria based on the reduction of the misfit function, and compare the convergence (iterations) and computational efficiency (forward problems) of four optimization methods (non-linear conjugate gradient, l -BFGS, Gauss Newton and Full Newton), when they are implemented in an efficient frequency-domain FWI with random source encoding. We compare the convergence, costs and final velocity models with the results obtained with the individual sources. This work-flow is first applied on a realistic synthetic experiment inspired by the geology of the Gulf of Mexico both for noise-free data and noisy data. Then, we assess the benefit provided by random source encoding and second-order optimization methods when applied on a 2D real ocean-bottom-cable (OBC) dataset recorded from the Valhall oil field. Even though the maximum advantage of source encoding can be seen with time marching or iterative solvers, we work with direct solvers. However, the conclusions we draw are based on the number of direct problems solved and therefore are directly applicable to iterative solvers in the frequency domain. Part of these results were submitted for publication (Castellanos et al., 2013).

Conclusions on source encoding with second order optimization methods with synthetic data

Our FWI results on the synthetic case when the initial model is sufficiently accurate and the data does not contain noise, indicate that the computational gain is boosted with source encoding combined with second order methods. The highest computational efficiency is provided by l -BFGS and Gauss-Newton, and the highest convergence rate is attained by Gauss-Newton. The savings in computational cost for our numerical test are shown in Figure 11. When very noisy data is considered (the power of the noise is 25% of the power of the data), the convergence rates of all optimization algorithms without source encoding is similar. This suggests that the action of the inverse Hessian on the gradient does not considerably improve the descent direction, due to the noise in the data. With source encoding, there is a similar situation and the convergence rates and computational cost of all methods are comparable. However, Newton methods show the lowest statistical variance in the final model, when many FWI inversions are performed. The computational savings for our numerical test with noise are lower compared to the noiseless scenario.

Conclusions on source encoding with second order optimization methods with real data

When we use the real data set of the Valhall oil field, the inversion with different optimization methods does not converge to the same local minima. This is probably due to the fact that

réel (Gauthier et al., 1986; Mora, 1989; Luo and Schuster, 1991). Bien que la FWI soit supposée reconstruire un large spectre de longueurs d'onde du modèle à partir des champs d'ondes réfléchis et transmis, les premiers essais (Devaney, 1984; Wu and Toksöz, 1987; Mora, 1988, 1989) ont uniquement fourni une image de la réflectivité c'est-à-dire des courtes longueurs d'onde tel que l'aurait fourni une méthode de migration. Cela motiva plusieurs analyses de résolution de la FWI pour comprendre quelle partie du spectre des nombres d'onde dans l'espace des modèles pouvait être reconstruit pour une géométrie d'acquisition et pour une bande passante de source (Jannane et al., 1989; Mora, 1989). D'autres difficultés inhérentes aux inversions multi-paramètres sont apparues. Les résultats montrent que la résolution avec laquelle chaque classe de paramètre est reconstruite varie d'un paramètre à l'autre en fonction de l'illumination angulaire locale fournie par la géométrie du dispositif d'acquisition, les couplages entre paramètres et l'influence variable de chaque paramètres dans les résidus des données (Tarantola et al., 1984; Mora, 1987)³. Ces difficultés furent identifiées lors des investigations pionnières de Gauthier et al. (1986); Wu and Toksöz (1987); Mora (1988, 1989); Luo and Schuster (1991), et restent aujourd'hui des sujets de recherche d'actualité en FWI.

Dans cette thèse, j'aborderai certains aspects de la FWI portant sur la réduction de son coût calculatoire et de sa non linéarité. Mais avant d'entre plus en détail dans mon travail de thèse, je présente un panorama général sur les principales caractéristiques de la FWI.

2.1 L'inversion des formes d'ondes complètes : le défi de l'imagerie sismique non linéaire.

Condition d'imagerie

La condition d'imagerie peut être défini comme l'opérateur agissant de l'espace des données vers l'espace des modèles et fournissant l'information sur les paramètres ayant une influence sur le champ d'onde calculé. Par conséquent, la condition d'imagerie révèle l'ensemble des paramètres qui, soumis à de faibles perturbations, génèrent des perturbations dans les données. A titre d'illustration, considérons une arrivée réfléchie dans les données d enregistrée au temps $t = T$, qui n'est pas prédite par les données modélisées u . La contribution de la condition d'imagerie est de déterminer les positions de l'espace où le modèle du sous-sol doit être modifié tel que le champ d'onde modélisé projeté aux positions des récepteurs contienne cette arrivée réfléchie. Pour un couple source-récepteur (S-R), la seule condition est que la position x du paramètre permette de vérifier l'équation

$$t_{S,x} + t_{x,R} = T. \quad (19)$$

Pour un modèle de référence homogène de vitesse v , toutes les positions x qui permettent de vérifier (19) sont situés sur une ellipse dont les points focaux sont situés aux positions de la source et du capteur, comme illustré sur la Figure 19a. Cela se montre aisément en écrivant explicitement le temps de trajet d'une onde se propageant de S à x , $t_{S,x}$, et de R à x , $t_{x,R}$,

$$\begin{aligned} t_{S,x} &= \sqrt{(x/v)^2 + (z/v)^2} \\ t_{x,R} &= \sqrt{(D-x)/v)^2 + (z/v)^2}, \end{aligned}$$

où D est la distance entre la source et le capteur. Substituer ces expressions dans (19) fournit l'équation suivante de l'ellipse,

$$\sqrt{x^2 + z^2} + \sqrt{(D-x)^2 + z^2} = vT,$$

³La paramétrisation du sous-sol est ici définie comme un jeu de paramètres physiques indépendants caractérisant les propriétés du sous-sol.

dont les points focaux sont aux positions $f = 0, D$, le grand axe est de longueur $a = vT/2 + D/2$ et le petit axe de largeur $b = \sqrt{(vT)^2 + D^2}/2$.

En résumé, pour un couple source-récepteur et une arrivée réfléchie enregistrée au temps T , tout point de l'ellipse représente une position potentielle où serait situé le diffractant réel ayant généré l'onde réfléchie car tous ces points permettent de vérifier (19). Sur l'exemple de la Figure 19a, $v = 2m/s$ le temps d'arrivée de la réflexion $T = 5s$, $x_S = 0$ et $x_R = 8m$, donnant un déport source-récepteur de $D = 8m$. Pour éliminer cette ambiguïté sur la position du réflecteur, la sommation des ellipses associées à tous les couples source-récepteur la position du réflecteur par interférence constructive de toutes les ellipses à cette position. Ce processus de construction par interférences constructives est illustré sur la Figure 19b pour un réflecteur horizontal.

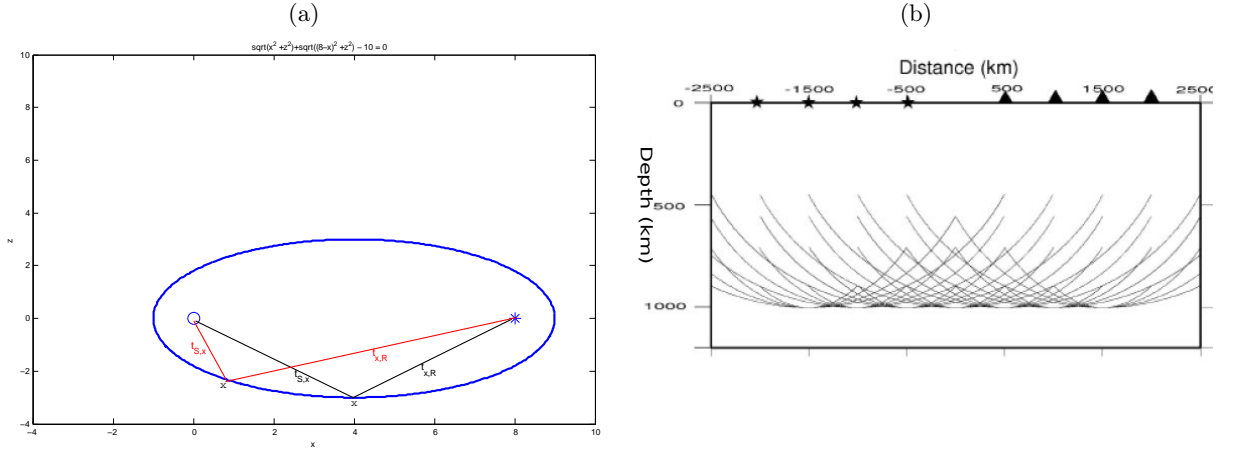


Figure 19: a) Exemple de points diffractants vérifiant la condition d'imagerie $t_{S,x} + t_{x,R} = T$, pour une onde réfléchie enregistrée au temps $T = 5s$, dans un modèle homogène de vitesse $v = 2m/s$, pour une source située à $x_S = 0$ et un récepteur situé à $x_R = 8m$. Les points forment une ellipse dont les points focaux coïncident avec la position du tir et du capteur. b) Figure tirée de Agudelo (2005). Imagerie d'un réflecteur horizontal. L'interférence constructive des ellipses à la position du réflecteur forme l'image de celui-ci.

La condition d'imagerie (19) dans l'espace des modèles peut être écrite de manière équivalente, pour une source et un ensemble de capteurs comme,

$$g(x) = \frac{\partial \phi_0(u(x, t, m))}{\partial m} = \left(P \frac{\partial u}{\partial m} \right)^\dagger (Pu(x, t, m) - d(x, t)) \quad (20)$$

$$= - \int_0^T \frac{\partial^2 u(x, t, m)}{\partial t^2} \lambda(x, T - t, m) dt, \quad (21)$$

où g est le gradient de la fonctionnelle, u est la solution du problème direct, $\lambda(T - t)$ est le champ d'onde rétro-propagé qui est calculé en utilisant les données résiduelles aux positions des récepteurs comme une source. Par construction, la condition d'imagerie correspond à la corrélation pondérée à décalage nul du champ direct avec le champ rétro-propagé en tout point du milieu, intégré sur le temps. La sommation sur le temps peut être représentée de manière équivalente par une sommation sur les fréquences,

$$g(x) = \int_{-\infty}^{\infty} \omega^2 u(x, \omega, m) \lambda^*(x, \omega, m) d\omega. \quad (22)$$

Plus de détails sur l'interprétation de la condition d'imagerie sont fournis dans le chapitre 2, Section 1.2.b. La condition d'imagerie de la FWI est analogue à celle d'autres méthodes

d'imagerie reposant sur une procédure de renversement temporel, comme cela est expliqué dans l'annexe 1.

2.1.a Analyse de résolution

D'un point de vue théorique, la qualité des images fournies par la FWI sera supérieure à celle fournie par d'autres méthodes d'imagerie car la FWI utilise toute l'information contenue dans les formes d'onde et le modèle du sous-sol utilisé pour résoudre le problème direct (18) est mis à jour à chaque itération. La percée fournie par la FWI réside dans la manière dont l'équation d'onde est résolue. Sans puissance de calcul suffisante, l'équation d'onde était résolue analytiquement ou asymptotiquement avec la théorie des rais qui nécessite des modèles lisses. Il était alors nécessaire d'utiliser une approximation linéarisée de l'équation d'onde et de séparer le modèle du sous-sol m en un modèle de référence lisse m_0 et un modèle de petites perturbations δm : $m = m_0 + \delta m$. Une fois cette séparation d'échelle effectuée, la solution de l'équation d'onde est trouvée par la théorie des perturbations à l'ordre 1, connue sous le nom de l'approximation de Born du premier ordre. Une approximation du champ d'onde total u est défini comme $u(m_0 + \delta m) \approx u_0 + (\partial u / \partial m) \delta m$, où u_0 est la solution de l'équation d'onde dans le modèle m_0 et $(\partial u / \partial m) \delta m$ est le champ d'onde diffracté par la perturbation du modèle δm .

Notons que l'approximation de Born d'ordre 1 néglige les termes de diffraction multiple. La distinction explicite entre m_0 et δm induit une séparation d'échelles entre le modèle de référence lisse et le modèle de perturbations, comme illustré sur la Figure 20a,b. Quand les ressources de calcul devinrent suffisantes, l'équation d'onde pu être résolue de manière "exacte" avec des méthodes numériques volumétriques telles que la méthode des différences finies pour des modèles complexes du sous-sol d'hétérogénéité arbitraire (Virieux, 1984). En s'affranchissant du besoin de linéariser l'équation d'onde, la séparation d'échelle entre le modèle de référence et le modèle de perturbations devient caduque car la propagation des ondes peut être calculée dans des modèles de complexité arbitraire. Si la géométrie d'acquisition permet l'enregistrement d'ondes transmises et réfléchies et si la source a une bande passante suffisamment large, FWI a le pouvoir de résolution théorique pour imager un spectre large et continu de nombres d'ondes. En pratique, il s'est néanmoins révélé difficile de mettre à jour simultanément toutes les composantes spectrales du modèle du sous-sol (Devaney, 1984; Wu and Toksöz, 1987; Mora, 1989). Une description plus détaillée de cette analyse de résolution est fournie dans le chapitre 2, Section 1.5.

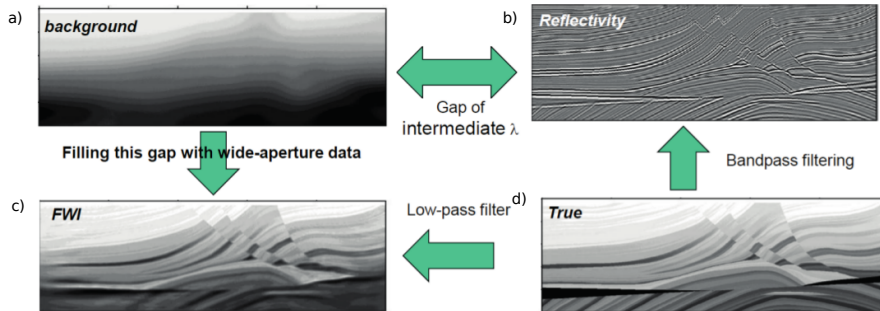


Figure 20: Illustration schématique des différentes échelles possibles impliquées en exploration sismique. (d) Le modèle exact du sous-sol m . Dans des approches linéarisables, le modèle m est décomposé en un modèle de référence lisse m_0 (a) et un modèle de perturbation δm (b). La FWI ne requiert pas cette séparation d'échelle si les données contiennent à la fois des arrivées transmises et réfléchies. Dans ce cas, le modèle du sous-sol reconstruit (c) est une version passe bas du vrai modèle (d).

2.1.b Coût de calcul

Bien que la puissance de calcul soit devenue suffisante dans les années 80 pour résoudre l'équation d'onde pour des modèles du sous-sol d'hétérogénéité arbitraire, par des méthodes de différences finies dans le domaine temps-espace notamment, les applications de la FWI sur des cas d'étude réels restaient hors de portée.

Méthode de l'état adjoint

Un premier pas ayant permis de réduire considérablement le coût de la FWI a résidé dans l'utilisation de la méthode de l'état adjoint pour calculer le gradient de la fonctionnelle (Lailly, 1983; Tarantola, 1984a).

L'équation (20) nécessite le calcul des dérivées partielles du champ d'onde modélisé par rapport à la totalité des paramètres du modèle, $i = 1, \dots, N \times N_p$. Ici, N représente le nombre de points de grille dans le domaine de calcul et N_p le nombre de classes de paramètres à reconstruire, par exemple, $(\rho, v_p, v_s) \rightarrow N_p = 3$. La méthode des différences finies fournit une approche brutale pour calculer les dérivées partielles, $\partial u / \partial m_i = (u(m + \Delta m_i) - u(m)) / \Delta m_i$, impliquant un coût calcul proportionnel à $C = 2 \times \mathbf{N} \times \mathbf{N}_p \times N_s$, où N_s est le nombre de sources sismiques. Avec la méthode de l'état adjoint (Lailly, 1983; Tarantola, 1984a), le gradient de la fonctionnelle est calculé avec l'équation (21) pour un coût proportionnel à $C = 2 \times N_s$. En effet, l'implémentation de l'équation (20) nécessite deux résolutions de l'équation d'onde par source, une pour calculer le champ d'onde incident u et une autre pour calculer le champ adjoint rétro-propagé λ . Le calcul du gradient avec la méthode de l'état adjoint est détaillé dans les annexes 3 et 4.

La FWI dans le domaine fréquentiel

Une deuxième étape vers la réduction du coût de la FWI a été franchie en formulant la FWI dans le domaine fréquentiel. Dans les années 90s, le formalisme de la FWI, tel qu'élaboré par Tarantola (1984a), a été transformé dans le domaine fréquentiel pour les équations d'ondes acoustique et élastique (Pratt and Worthington, 1990; Pratt, 1990; Pratt and Shipp, 1999). Appliquer une transformée de Fourier par rapport au temps à l'équation d'onde transforme un problème d'évolution à un problème stationnaire de conditions aux limites pour chaque composante fréquentielle à modéliser. L'équation d'onde dans le domaine fréquentiel est une forme généralisée de l'équation d'Helmholtz pouvant s'écrire sous forme matricielle comme

$$A(m, \omega)u(s, m, \omega) = s(\omega), \quad (23)$$

où les coefficients de $A(m, \omega)$, matrice résultant de la discrétisation de l'opérateur de l'équation d'onde en domaine fréquentiel, dépendent de la fréquence $\omega = 2\pi f$. L'équation (23) est un système d'équations linéaires où A est une matrice carrée de dimensions $N \times N$ tandis que u et s sont des vecteurs de dimension N . Ce système linéaire peut être résolu par des méthodes directes ou itératives. En ce qui concerne les méthodes itératives, le nombre d'itérations nécessaires à l'estimation de la solution dépend du conditionnement de la matrice. Par conséquent, la conception de pré-conditionneurs performants sont nécessaires et constitue la difficulté de ces approches (Erlangga, 2005; Plessix, 2007; Erlangga and Nabben, 2008). Chaque source sismique (terme de droite) nécessite une nouvelle résolution itérative du système, indépendante de celles effectuées pour les sources précédentes. A contrario, les méthodes directes reposent sur des techniques d'élimination de Gauss telles que les approches fondées sur des décompositions triangulaires supérieure/inférieure de la matrice ($A = LU$). La factorisation LU de la matrice dans l'équation (23) dépend uniquement de m et de la fréquence ω . Par conséquent, pour chaque itération non-linéaire du problème inverse, une seule factorisation LU est effectuée par fréquence pour calculer le gradient. Les champs d'onde calculés pour chacune des sources sont efficacement

calculés à partir des facteurs LU par substitutions directe et inverse. Si l'inversion peut être limitée à quelques fréquences discrètes, des analyses de complexité montrent que les approches par solveur direct sont les plus efficaces (au moins en deux dimensions) en raison de l'efficacité des phases de substitutions et du nombre limité de factorisation à effectuer. Pour des applications de la FWI le gain peut être considérable (Pratt and Worthington, 1990; Pratt, 1990) car le nombre de sources pour des expériences 2D et 3D est respectivement de l'ordre de $10^2 - 10^3$ et de $10^3 - 10^4$. Les domaines d'application privilégiés de l'inversion en domaine fréquentiel concernent les acquisitions dites grand-angle pour lesquelles une analyse de résolution permet de montrer que seules quelques fréquences discrètes peuvent être injectées dans l'inversion grâce à la redondance d'illumination fournie par les fréquences temporelles et les angles d'ouverture (Sirgue and Pratt, 2004). A contrario, les géométries d'acquisition en réflexion fournissent un éclairage angulaire du milieu beaucoup plus étroit qui doit être compensé par un échantillonnage plus fin des fréquences dans la FWI rendant l'approche fréquentielle moins pertinente pour ces configurations (Freudenreich and Singh, 2000).

Au delà du bénéfice fourni par le coût réduit du problèmes direct, les approches fréquentiels permettent la manipulation de volumes de données plus faibles lorsque quelques fréquences discrètes sont inversées. Dans le domaine temporel, le calcul du gradient nécessite la corrélation des champs directs et adjoints à chaque pas de temps t . Comme le champ adjoint rétro-propagé λ ne peut être calculé qu'après que le champ incident ait été complètement calculé (car le terme de source du champ adjoint est formé par les résidus entre les données enregistrées et les valeurs aux positions des capteurs des champs directs modélisés), deux approches sont possibles pour calculer ces corrélations. Une première approche consiste à stocker en mémoire le champ direct à tous les temps $u(x, t), x \in \Omega$, où N_t est le nombre de pas de temps. La deuxième consiste à recalculer le champ incident durant le calcul du champ adjoint soit par de manière pro-grade avec des techniques de point de reprise soit de manière rétrograde à partir des valeurs du champ incident stocké aux frontières du domaine de calcul. La première méthode est rarement utilisé, en particulier en 3D, en raison de son coût de stockage prohibitif. La deuxième approche nécessite de recalculer au moins une fois le champ incident. En domaine fréquentiel, le gradient se résume à un produit pondéré de deux champs d'onde monochromatiques dont le stockage en mémoire ne pose pas de problème. Plus de détails sont fournis dans le chapitre 2, Section 1.4.

Assemblage et encodage de sources

Les limites des approches fréquentielles fondées sur des solveurs directs résultent du calcul et du stockage des facteurs LU. Pour des méthodes aux différences finies, la matrice A est bande diagonale. En 2D, il y a trois bandes et la distance entre bandes est proportionnelle à N . Pour des applications 3D, il y a cinq bandes et la distance entre bandes est proportionnelle à N^2 . Contrairement à A , les matrices L et U sont pleines et leur stockage est par conséquent coûteux. Pour des applications acoustiques 3D de dimension réduite, la faisabilité des approches directes a été montrée (Operto et al., 2007). Les premières applications de FWI 3D acoustique en domaine fréquentiel ont été réalisées en domaine fréquentiel à partir de méthodes itératives (Plessix, 2009; Plessix and Perkins, 2010) ou de modélisation temporelles pour le problème direct (Sirgue et al., 2010). Dans ce dernier cas, les champs monochromatiques sont extraits à la volée dans la boucle sur les pas de temps par transformée de Fourier discrète. Pour des applications élastiques 3D, l'approche par solveur direct semble aujourd'hui inaccessible en raison de la nature vectorielle de l'équation d'onde élastique nécessitant le calcul d'au moins trois composantes de vitesse particulières et des faibles longueurs d'ondes propagées en relation avec la vitesse de propagation des ondes de cisaillement. Dans cette configuration, seules les modélisations en domaine fréquentiel avec des solveurs itératifs ou des modélisations en domaine temporel semblent possibles.

Une réduction du coût calcul est envisageable si le nombre de problèmes directs est diminué.

Comme pour des approches fréquentielles fondées sur des solveurs itératifs ou des modélisations temporelles, le coût des problèmes directs est proportionnel au nombre de sources (équations (18) et (23)), la manière la plus naturelle de réduire le coût est de diminuer le nombre de sources. Lors des acquisitions sismiques, cette approche est implémentée en mélangeant les sources: plusieurs sources sont émises simultanément ou avec un décalage entre elles, et le champ d'onde enregistré contient la contribution mélangée de chaque source en vertu de la linéarité de l'équation d'onde par rapport au terme de source (Beasley, 2008; Berkhout, 2008). Cette technique d'assemblage des sources sismiques réduit considérablement les temps d'acquisition mais aussi le nombre de sources à considérer lors de la résolution du problème direct en FWI. Une autre méthode courant d'accélération, appelée encodage des sources, consiste à effectuer une acquisition sismique conventionnelle (en émettant une source à la fois). Néanmoins, au lieu de résoudre le problème direct, (18) ou (23), pour chaque source, des super-sources sont formées par combinaison linéaire des sources pondérées par des facteurs aléatoires (Romero et al., 2000; Neelamani et al., 2008; Krebs et al., 2009; Ben Hadj Ali et al., 2011; Schuster et al., 2011; van Leeuwen et al., 2011; Huang and Schuster, 2012),

$$\tilde{s} = \sum_{i=1}^{N_s} \alpha_i s_i, \quad \tilde{d}_o = \sum_{i=1}^{N_s} \alpha_i d_{o_i}. \quad (24)$$

où α désigne des coefficients aléatoires. Les données observées sont encodées de manière analogue, \tilde{d}_o . Pour encoder les données de la sorte, il est nécessaire que les sources soient enregistrées par le même ensemble de récepteurs, désignant ainsi les acquisitions d'extension fixe (fixed-spread en Anglais) comme les acquisitions de fond de mer. Le volume de données généré par N_s sources et enregistré par N_r récepteurs est compressé en une seule collection de traces à tir commun.

Même si la méthode d'encodage des sources couplée à des algorithmes optimisation de gradient permet de réduire le coût calcul par itération (2 problème directs au lieu de $2 \times N_s$ problèmes directs), plus d'itérations de l'optimisation doivent être effectuées car chaque mise à jour du modèle du sous-sol est moins précise en raison des interférences entre les différences sources élémentaires formant la super-source (dénommées cross-talk en Anglais), comme cela est illustré sur la Figure 21a. In fine, un gain en termes de coût ou accélération est obtenu si le nombre total de problèmes directs nécessaires pour atteindre une valeur prédéfinie de la fonction coût est plus faible lorsque l'encodage des sources est utilisé, comme cela est illustré sur la Figure 21b. La méthode d'encodage de sources a été appliquée avec succès sur des données réelles par Baumstein et al. (2011); Routh et al. (2011); Bansal et al. (2013); Schiemenz and Igel (2013).

2.1.c Optimisation non linéaire

Les applications pionnières de la FWI ont révélé des difficultés supplémentaires associées au problème d'optimisation malgré la précision accrue du problème direct et la faisabilité théorique de la FWI (Gauthier et al., 1986; Luo and Schuster, 1991). La relation entre les données et les paramètres est non linéaire. Cela signifie que de petites modifications des paramètres du modèle peuvent générer de fortes perturbations des données et de la fonction coût ϕ_0 , ces variations se comportant de façon non linéaire. Lors d'une inversion acoustique de données entre-puits, Luo and Schuster (1991) ont comparé l'inversion des temps de trajet de la première arrivée et la FWI. Les résultats ont montré le pouvoir de résolution supérieur de la FWI. Néanmoins, l'inversion des temps de trajet a fourni des résultats plus robustes car la FWI convergerait vers des minimums secondaires.

Optimisation non convexe

thington, 1990). Néanmoins, en raison du coût calcul, ce n'est que lors de ces dernières années qu'une démonstration claire des améliorations fournies par les méthodes de Newton fut présentée. Lorsque l'on considère une approximation du second ordre de la fonction coût (fonction coût localement quadratique), le Hessien est constitué de deux termes, un formé par la corrélation entre les champs d'onde diffractés $\partial u / \partial m$ (les dérivés de Fréchet) et l'autre formé par des termes de diffraction double $\partial^2 u / \partial m^2$. Il a été montré que les artefacts d'étalement qui apparaissent en raison de ambiguïté de la condition d'imagerie peuvent être diminués par l'action déconvolante du Hessien dans l'approximation de Gauss-Newton (qui utilise uniquement les termes de diffraction d'ordre 1) (Pratt et al., 1998). Le Hessien complet fournit une amélioration supplémentaire en corrigeant le gradient des contributions parasites des diffractions d'ordre 2. Plus précisément, la solution complète de l'équation d'onde modélise des ondes diffractées multiples ce qui génère une incompatibilité avec la condition d'imagerie (20) qui considère uniquement les termes en diffraction simple. Cela implique que la partie des résidus dus à de la diffraction double seront interprétés à tort comme des événements simplement diffractés conduisant à une contribution erronée dans l'image. Le terme du second ordre dans le Hessien contribue à supprimer ces artefacts (Pratt et al., 1998). En théorie, quand le Hessien n'est pas utilisé, ces artefacts tendent à disparaître au cours des itérations. Néanmoins, en pratique, en raison de la non linéarité de l'inversion, l'inversion implémentée avec des méthodes de gradient peut rester bloquée dans un minimum local en raison de ces artefacts. Par exemple, reconstruite le modèle de la Figure 24a avec une fréquence de 7 Hz est difficile car la fréquence est élevée relativement à la taille caractéristique des structures présentes dans le modèle initial. Cela peut générer des sauts de phase et la convergence vers un minimum local. De plus, il y a des événements doublement diffractés entre les deux inclusions circulaires. L'inversion fondée sur l'algorithme de Newton fournit clairement la meilleure solution (Figure 24e). L'autre avantage des méthodes de Newton est de réduire le nombre d'itérations nécessaires pour atteindre une valeur donnée de la fonction coût, avec l'inconvénient que le coût d'une itération est plus élevé comparativement à celle d'une méthode de plus grande pente. Des applications sur données réelles avec des algorithmes de Newton et de quasi-Newton (l-BFGS) sont présentées dans Brossier et al. (2009a,b); Plessix et al. (2012); Métivier et al. (2014).

Inversion multi-paramètres

Les bénéfices que l'on peut tirer de la prise en compte du Hessien s'appliquent aussi aux inversions multi-paramètres. Les applications pionnières de la FWI multi-paramètres ont montré que la résolution avec laquelle les paramètres de nature différente sont reconstruits dépend du choix de la paramétrisation du sous-sol (Tarantola et al., 1984; Mora, 1987; Pratt and Worthington, 1990). Il y a des phénomènes de couplages entre les paramètres de nature différente (la vitesse de propagation des ondes de compression v_p et de cisaillement v_s , la densité, l'atténuation, les paramètres décrivant l'anisotropie ϵ et δ , ce qui rend complexe le processus d'optimisation car le gradient de la fonctionnelle par rapport à un paramètre est pollué par l'influence des autres paramètres. Plus précisément, le modèle de perturbation pour un paramètre est en fait une combinaison linéaire des gradients par rapport aux différents paramètres. De plus, la sensibilité des données à chaque paramètre, que l'on peut mesurer par les dérivés de Fréchet $\partial u / \partial m$, peut avoir des ordres de grandeur fortement variable d'un paramètre à l'autre générant un problème d'optimisation mal conditionné. La conséquence est que l'optimisation tendra à ne mettre à jour que les paramètres dont les dérivés de Fréchet sont les plus élevées. Une re-paramétrisation des paramètres physiques dans l'inversion contribuera d'une part à adimensionnaliser les dérivés de Fréchet et d'autre part à minimiser les couplages entre paramètres. Une re-paramétrisation du modèle du sous-sol modifiera les dérivés de Fréchet, dont les amplitudes sont contrôlées par le diagramme de rayonnement $\partial A / \partial m$ de la source virtuelle secondaire positionnée à la position du paramètre du modèle (Pratt et al., 1998). Seul ce diagramme de rayonnement différencie

La non convexité de la fonctionnelle à minimiser (non linéarité) et le caractère sous-déterminé du problème inverse (non unicité de la solution) nécessitent d'introduire des termes de régularisation pour restreindre l'espace des solutions et guider l'inversion vers un modèle du sous-sol que le géophysicien considérera plausible. Par exemple, des modèles du sous-sol seront rejetés et des modèles lisses seront favorisés dans la recherche. La régularisation peut être introduite dans la fonction coût,

$$\phi = \phi_0 + \lambda \|R(m)\|_2^2 \quad (26)$$

$$= \|Pu_c(m) - d_o\|_2^2 + \lambda \|R(m)\|_2^2, \quad (27)$$

où $R(m)$ est le terme de régularisation et $\lambda \|R(m)\|_2^2$ est un terme de paramètre avec λ jouant le rôle de multiplicateur de Lagrange. L'optimisation vise à trouver le modèle qui explique au mieux de manière conjointe les données et le terme de régularisation. Généralement, des modèles lisses sont favorisés ce qui conduit à $R(m) = \nabla m$. Si des informations a priori sur les valeurs du modèle à certaines positions de l'espace sont disponibles, celles-ci peuvent être aussi incorporées dans le terme régularisant (Asnaashari et al., 2013). Relativement peu de travaux ont été publiés sur le choix optimal de régularisation et de normes en FWI. A titre d'exemple, citons néanmoins l'utilisation de régularisations fondées sur la variation totale (TV) des modèles du sous-sol dont l'objet est de favoriser la reconstruction des contrastes dans les modèles du sous-sol (Ramírez and Lewis, 2010; Anagaw and Sacchi, 2012; Guitton, 2012),

$$\phi = \|Pu_c(m) - d_o\|_2^2 + \lambda \|\nabla(m)\|_1, \quad (28)$$

ou des régularisation visant à favoriser le caractère creux de la solution sur des bases adaptées W (généralement d'ondelettes) (Candes and Romberg, 2005; Loris et al., 2010; Herrmann and Li, 2012),

$$\phi = \|Pu(m) - d\|_2^2 + \lambda \|W(m)\|_1. \quad (29)$$

D'autres régularisations ont été proposées, où la régularisation n'est pas additionnée à la fonctionnelle mais est plutôt multipliée (Abubakar et al., 2002, 2004). Théoriquement, les régularisations multiplicatives ont l'avantage que le poids relatif entre le gradient généré par les données et celui généré par la régularisation est déterminé de manière automatique.

2.2 Présentation de la thèse et description des principaux résultats

Cette brève discussion sur les principales questions et les défis soulevés par la FWI montre pourquoi cette technologie est toujours un champ actif de recherche depuis les années 80. Cette thèse a pour objectif d'ajouter une contribution à la compréhension de la FWI en explorant certains aspects spécifiques en rapport avec son coût numérique et sa non linéarité.

Je commence par présenter de manière générale la théorie des problèmes inverses et des algorithmes d'optimisation dans le chapitre 1, sans me restreindre au cas spécifique de l'imagerie sismique. J'introduis les concepts de problèmes inverses linéaires versus non linéaires et des problèmes inverses mal posés. Comme les problèmes inverses sont généralement formulés sous forme d'un problème de minimisation, j'introduis les algorithmes d'optimisation que je serai amenée à manipuler tout au long de cette thèse. Le lecteur familier de ces concepts pourra directement se reporter au chapitre sur la FWI.

Le chapitre 2 complète l'introduction fournie ci-dessus sur la FWI. Une interprétation physique des deux principaux ingrédients intervenant dans les algorithmes d'optimisation est fournie: le gradient de la fonctionnelle et le Hessien.

Pour un problème de petite dimension, je calcule le Hessien complet et illustre l'action d'une ligne de son inverse sur le gradient de la fonctionnelle afin de fournir des éléments d'information sur la mécanique avec laquelle il corrige les artéfacts. L'interprétation physique des conditions d'imagerie associées à d'autres méthodes d'imagerie fondées sur le concept de renversement temporel est passée en revue dans l'annexe 1. Les expressions du gradient et du Hessien qui sont développées dans la littérature avec la méthode de l'état adjoint (Lions, 1972; Chavent, 2009; Plessix, 2006; Métivier et al., 2013a) sont utilisées dans cette thèse et sont redéveloppées dans les annexes 3 et 5 pour le cas spécifique des équations de l'élastodynamique formulées sous forme d'un système hyperbolique du premier ordre.

Formulation auto-adjointe de l'équation d'onde élastique isotrope en vitesse-contrainte

Généralement, le gradient de la fonction coût de la FWI est implémenté avec la méthode de l'état adjoint à partir de l'équation d'onde du second ordre auto-adjointe. En revanche, l'équation d'onde formulée sous forme d'un système hyperbolique d'ordre 1 en vitesse-contrainte ne conduit pas à une forme auto-adjointe ($A \neq A^\dagger$), nécessitant une implémentation dissociée de l'opérateur direct (équation d'état) et de l'opérateur adjoint. Ma contribution, présentée dans l'annexe 4, a consisté à développer un formalisme permettant d'exprimer les équations de l'élastodynamiques du premier ordre sous forme d'un opérateur auto-adjoint facilitant ainsi grandement l'implémentation du gradient. Ceci est implémenté via un changement de variable appliqué aux contraintes normales, fourni par la décomposition en vecteurs propre de la matrice de raideur. Ce changement de variable permet de reformuler l'équation d'onde sous une forme pseudo-conservative au sein de laquelle les paramètres du milieu sont regroupés dans une matrice diagonale factorisée aux dérivés temporelles du système hyperbolique. Ce travail a été publié dans Castellanos et al. (2011).

Encodage de sources combiné avec des méthodes d'optimisation du second ordre

En imagerie sismique, le nombre de simulations à effectuer est proportionnel au nombre de source. Dans les approches fréquentielles fondées sur des solveurs directs, le problème direct est résolu de manière plus efficace car la factorisation LU de la matrice est commune à toutes les sources et les solutions sont calculées efficacement par substitution. Néanmoins, pour des applications en trois dimensions de la FWI élastique, les méthodes de modélisations fondées sur des algorithmes explicites d'évolution temporelle ou sur des solveurs itératifs en domaine fréquentiel semblent les plus réalistes. Ces méthodes doivent tirer un bénéfice important des techniques d'encodages de sources car le coût des modélisations est directement proportionnel au nombre de sources contrairement aux approches fondées sur des solveurs directs. Bien que les méthodes d'encodage de sources aient été beaucoup utilisées en FWI ou en migration par renversement temporel (Neelamani et al., 2008; Krebs et al., 2009; Ben Hadj Ali et al., 2011; Schuster et al., 2011; van Leeuwen et al., 2011; Huang and Schuster, 2012), elles ont principalement été combinées avec des algorithmes d'optimisation de gradient, reproduisant ainsi les principes des méthodes de gradient stochastiques (Robbins and Monro, 1951; Spall, 2003). Dans le but de réduire encore le coût des calculs et améliorer le facteur d'accélération fourni par les méthodes d'encodage (i.e., augmenter l'écart entre les courbes obtenues avec et sans encodage sur la Figure 21), j'ai combiné des algorithmes d'optimisation de quasi-Newton et de Newton avec de l'encodage de source, ce qui n'avait jamais été proposé dans le cadre applicatif de la FWI.

Dans d'autres domaines comme en Machine learning, cette combinaison commence à peine à être explorée (Schraudolph et al., 2007). L'absence de preuve de convergence des méthodes stochastiques du second ordre (Bottou and Le Cun, 2005) et le manque d'algorithmes efficaces pour calculer la direction de descente de Newton (Métivier et al., 2013b, 2014) figurent parmi les raisons ayant freiné ces investigations. Les équations permettant le calcul efficace du Hessien

complet à partir de la méthode de l'état adjoint du second ordre sont présentées dans l'annexe 5. Néanmoins, le coût de calcul par itération de la FWI des méthodes de Newton tronqué (Gauss-Newton et Full Newton) est plus élevé que celui des méthodes de plus grande pente, car le calcul de la direction de descente de Newton fondé sur des approches non matricielles nécessite de calculer plus de problèmes directs. Par ailleurs, dans le cas des méthodes de quasi-Newton fondées sur l -BFGS, l'approximation du Hessien a besoin d'être périodiquement ré-initialisé contribuant à ralentir la convergence. Dès lors, il n'est pas clair quel gain réel peut être tiré des méthodes d'optimisation du second ordre en FWI que des techniques d'encodage de sources soient utilisées ou pas.

Dans le chapitre 3, j'introduis un critère d'arrêt des itérations fondé sur le taux de réduction de la fonction coût et je compare la convergence (mesuré par le nombre d'itérations effectuées) et le coût de calcul (mesuré par le nombre de problèmes directs effectué) pour quatre méthodes d'optimisation (gradient conjugué non linéaire, l -BFGS, Gauss-Newton et Full Newton), lorsque ceux-ci sont implémentés dans la FWI en domaine fréquentiel couplée avec de l'encodage de sources. Je compare la convergence, les coûts et la qualité des modèles du sous-sol obtenus avec source encoding avec ceux obtenus lorsque les tirs sont traités indépendamment. L'analyse est tout d'abord appliquée à un cas d'étude synthétique inspiré de la géologie du Golf du Mexique pour des données sans et avec bruit. Ensuite, je quantifie l'apport de la méthode d'encodage de sources et des méthodes d'optimisation du second-ordre avec des données réelles enregistrées par un dispositif de câbles de fond de mer sur le champ pétrolier de Valhall (mer du Nord). Bien que la méthode d'encodage de sources soit plus adaptée à des méthodes de modélisation fondées sur des solveurs itératifs ou des méthodes explicites d'évolution, j'utilise un code fondé sur un solveur direct pour des raisons d'efficacité calculatoire (les cas d'étude présentés sont 2D et acoustiques). Néanmoins, les conclusions que je tire reposent sur le nombre de problèmes directs effectués (le nombre de phase de substitutions) et s'appliquent donc à l'identique à d'autres méthodes de modélisation. Une partie des résultats présentés ont été soumis pour publication dans [Castellanos et al. \(2013\)](#).

Conclusions sur l'encodage des sources couplée avec des méthodes d'optimisation du second-ordre: cas synthétique

Mes résultats de FWI sur le cas synthétique, lorsque le modèle initial est de précision suffisante et les données ne contiennent pas de bruit, indiquent que le gain calcul fourni par l'encodage des sources couplée avec des méthodes d'optimisation du second-ordre est significatif. La plus grande accélération est obtenue avec l'algorithme de l -BFGS et de Gauss-Newton et le meilleur taux de convergence est atteint par Gauss-Newton. Les gains en temps calcul pour ce cas d'étude sont illustrés sur la Figure 25. Quand du bruit est introduit dans les données (la puissance du bruit représente 25% de celle des données), le taux de convergence de toutes les méthodes d'optimisation sans encodage de sources tendent à se niveler. Cela suggère que l'action du Hessien sur le gradient n'améliore plus de manière significative la direction de descente, en raison du bruit dans les données. Avec l'encodage des sources, cette tendance à niveler les performances de chaque méthode d'optimisation est reproduite. Néanmoins, les méthodes de Newton fournissent une variance plus faible des modèles de vitesse obtenus, lorsque plusieurs réalisations de la FWI sont effectuées. L'accélération fournie par l'encodage des sources est plus faible lorsque du bruit est introduit dans les données.

Conclusions sur l'encodage des sources couplée avec des méthodes d'optimisation du second-ordre: cas réel

Dans le cas des données réelles du champ de Valhall, l'inversion avec différentes méthodes

SPEED-UP OF FWI WITH SOURCE ENCODING

To perform 3D elastic full waveform inversion in time or frequency domain with the current computational resources, time marching or iterative solvers appear to be the most feasible solution. The number of direct problems are proportional to the number of sources, and with time marching or iterative solvers, each direct problem is computationally expensive. Instead of working with each source individually, source encoding techniques allow to work with linear combinations of the sources. The advantage is that the computational cost per iteration is greatly reduced. The counterpart that many more iterations need to be performed because the model update in each iteration is less precise, due to undesired interference effects between sources. In the end, there will be a computational gain (speed-up) if the number of direct problems solved to attain a predefined value of misfit function is less with source encoding than using the sources individually in the standard way.

Although source encoding techniques have been widely used (Neelamani et al., 2008; Krebs et al., 2009; Ben Hadj Ali et al., 2011; Schuster et al., 2011; van Leeuwen et al., 2011; Huang and Schuster, 2012), they have been mostly combined with gradient optimization algorithms, resembling stochastic gradient algorithms (Robbins and Monro, 1951; Spall, 2003). With the purpose of reducing even more the computational cost and improving the computational efficiency, we combine quasi-Newton and Newton optimization methods with source encoding techniques. However, the computational cost per iteration of the truncated Newton methods (Gauss-Newton and Full Newton) is higher per iteration, because the Newton descent direction requires the solution of an additional linear system. In the case of l -BFGS, we periodically restart the Hessian approximation. Therefore, a priori, it is not clear whether the computational savings can be further improved with second order optimization methods.

We compare the convergence and the computational efficiency of four optimization methods (non-linear conjugate gradient, l -BFGS, Gauss Newton (GN) and Full Newton (FN)) when they are implemented in efficient frequency-domain FWI with and without random source encoding. A stopping criterion of iterations is carefully designed allowing for a fair comparison of the opti-

mization methods and a fair assessment of the speed-up provided by the random source encoding method. These work flows are first applied on a realistic synthetic experiment inspired by the geology of the Gulf of Mexico both for noise-free data and noisy data. Then, we assess the benefit provided by random source encoding and second-order optimization methods when applied on a 2D real ocean-bottom-cable (OBC) dataset recorded from the Valhall oil field. Even though the maximum advantage of source encoding can be seen with time marching or iterative solvers, we work with direct solvers. However, the conclusions we draw are based on the number of direct problems solved and therefore are directly applicable to iterative solvers in the frequency domain.

We found that, in an ideal noise-free data scenario with frequency groups that determine an approximately convex misfit function, truncated Newton methods remain more computationally expensive than the quasi-Newton method *l*-BFGS. As noise is added to the synthetic data and more aggressive regularization is used, the action of the Hessian becomes less effective and the convergence rate of the Newton-based methods is thus degraded. This contributes to level down the convergence rate of Newton-based methods relative to steepest-descent method. While all of the optimization methods generate subsurface models of similar accuracy for the synthetic example, application on real data from the Valhall field shows that the truncated Newton methods attain a lower misfit function value than the other optimization methods, hence suggesting a more robust behavior to noise and other source of errors such as incomplete wave physics. A speed-up of nearly one order of magnitude was reached for the selected stopping criterion of iterations. The accuracy of the subsurface models that was achieved for this stopping criterion of iteration was validated against published previous works, a sonic log and reverse time migration. Finally, we derive some formulas for the estimation of the variance of the encoded gradient, that can aid towards source encoding strategies. Parts of the contents of this chapter can be found in (Castellanos et al., 2013).

1 INTRODUCTION

FWI has been shown to be quite successful in 2D applications under the acoustic (Plessix et al., 2012) and elastic (Brossier et al., 2009a) approximation. More recently, 3D FWI has also been made possible thanks to the increasing development of high performance computing (Plessix, 2009; Plessix and Perkins, 2010; Sirgue et al., 2010; Vigh et al., 2013). However, FWI remains computationally intensive, particularly in three dimensions, when high frequencies are injected in the inversion, or complex wave propagation (visco-elastic) is accounted for. Hence, there is a natural interest to reduce this computational burden.

The full wave form inversion algorithm alternates the solution of a direct and an inverse problem. The direct problem consists in solving the wave equation, and the inverse problem consists in solving an optimization problem. Solving the discrete direct problem (in the time or frequency domain) amounts to solving the linear system

$$A(x, m)u(x, m) = s(x), \quad (3.1)$$

where $A(x, m)$ is the wave equation operator, $s(x)$ the source term and $u(x, m)$ the wave field solution. In FWI, the direct problem is often a computational bottleneck because the number of times the wave equation must be solved is proportional to the number of sources. For example, at each iteration, the computation of the gradient requires the solution of $2 N_s$ direct problems, and this may increase if a more precise descent direction is taken and due to the line search evaluations.

One possible solution to reduce the number of right hand sides in equation (3.1), known as source blending, is to fire several sources simultaneously or with some time delays at the acquisition level when the data is being collected, (Beasley, 2008; Berkhout, 2008). The measured data therefore has the wavefield response of the individual sources and the interference effects with each other. Source blending reduces the time of collection data in the field and also the number of sources that are processed when solving the direct problem. Another popular speed-up technique, called *source encoding*, consists in performing a traditional seismic acquisition experiment using only one source at time, for N_s sources. When the information is being processed, a limited number of super sources \tilde{s} are created by performing a linear combination of individual sources that are weighted by random coefficients ($\tilde{s} = \sum_{i=1}^{N_s} \alpha_i s_i$). This technique does not save time in the acquisition of the data, but is more flexible at the processing level. It was first proposed for migration by Romero et al. (2000) using random phase encoding. After the pioneering work of Romero et al. (2000), random source encoding has been widely used in time-domain and frequency-domain FWI (Neelamani et al., 2008; Krebs et al., 2009; Ben Hadj Ali et al., 2011; Schuster et al., 2011; van Leeuwen et al., 2011; Huang and Schuster, 2012). As a result of the linearity in the wave equation between s and u , the encoded wavefield \tilde{u} is a superposition of the individual wavefield solutions ($\tilde{u} = \sum_{i=1}^{N_s} \alpha_i u_i$). The same assembling operation is applied on the observed data, and on the adjoint wavefield in the gradient computation. The random coefficients must satisfy $\mathbb{E}[\alpha_i^* \alpha_j] = \delta_{i,j}$.

Figure 3.1 shows schematically how the encoded data are generated. Let N_s denote the number of sources, N_r the number of receivers and we will assume that all the sources share the same number of receivers. Under this assumption, it is possible to represent the recorded data in a matrix form, like the gray matrix in Figure 3.1. Each column represents the recorded data generated by one source, at all the receiver positions. The encoded recorded data \tilde{d} is created by performing a linear combination of the data generated by all the sources, with weights given by a random encoding vector. For example, by taking the recorded data at the first receiver position generated by all sources (first row in blue of the data matrix), and multiplying it by a random encoding vector (vector in yellow), we obtain the entry of the encoded data at the first receiver. Doing the multiplication of the whole data matrix with the encoding vector, results in the encoded data at all receiver positions. Note that this approach is only valid when all sources share the same receivers, so that we can assemble the data matrix. In general, K different super sources can be assembled by using K different realizations of the random encoding vectors.

Ideally, using source encoding will provide a computational gain because instead of solving the wave equation N_s times to solve the forward problem (3.1), we only need to solve the wave equation K times (where usually $K \ll N_s$). However, the computational gain is in practice much less because the descent direction with source encoding is less accurate than the descent direction provided by the full set of sources because it will be contaminated with artefacts. Consequently, with source encoding more iterations will be required to attain the same level of accuracy in the final model as that obtained without source encoding. For a fixed computational time and using source encoding, many low cost iterations with low accurate updates can be performed. Using the full set of sources for the same computational time, fewer iterations with more accurate updates can be done. The trade off that source encoding provides lies in the idea that each update does not need to be as accurate as possible and, on the other hand, it is more important to perform iterations that will allow the model to quickly evolve.

The model update provided in each iteration using source encoding methods is less accurate

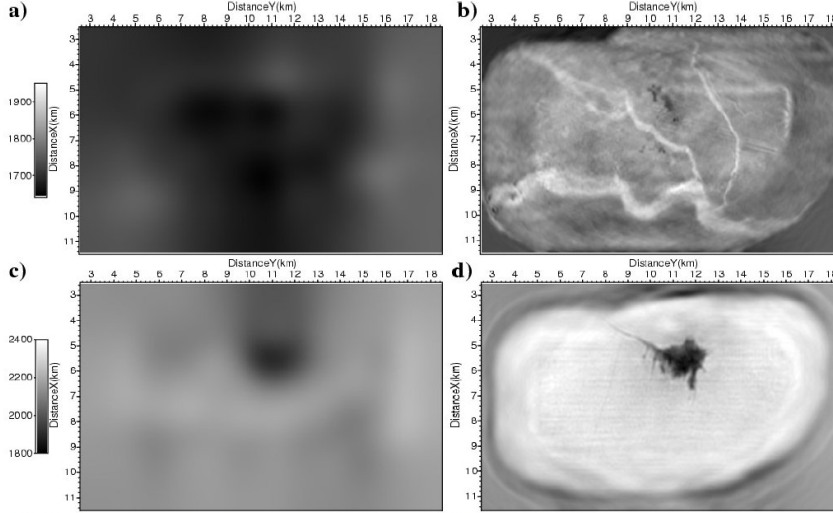


Figure 3.4: Figure from (Sirgue et al., 2009) 2D horizontal velocity profiles. a) At a depth of 150m, using reflection tomography. b) At a depth of 150m, using FWI. c) At a depth of 1050m, using reflection tomography. d) At a depth of 1050m, using FWI.

show that, when the misfit function is strongly non-convex, source encoding can help to guide the inversion toward an improved minimum of the misfit function, thanks to a broader exploration of the model space. The real data case study shows that the truncated Newton methods provide the most robust direction of descent leading to subsurface models of similar accuracy, regardless of the encoding. A speed-up of nearly one order of magnitude was attained thanks to a careful design of the stopping criterion of iteration.

2 SPEED-UP FWI : STATE OF THE ART

FWI has been shown to be quite successful in 2D applications under the acoustic (Plessix et al., 2012) and elastic (Brossier et al., 2009a) approximation. More recently, 3D FWI has also been possible thanks to the increasing development of high performance computing (Plessix, 2009; Plessix and Perkins, 2010; Sirgue et al., 2010; Vigh et al., 2013). However, FWI remains computationally intensive, particularly in three dimensions, when high frequencies are injected in the inversion, or complex wave propagation (visco-elastic) is accounted for. Hence, there is a natural interest to reduce this computational burden.

In the simplest scenario, computational savings can be achieved by sub-sampling the data, hence reducing the volume of information injected in the inversion. An illustrative sub-sampling approach is the so-called efficient frequency-domain FWI (Sirgue and Pratt, 2004), where only a few discrete frequency components are processed in a hierarchical manner. Note that the sub-sampling is successful if the redundancy contained in the original data set is carefully decimated. In efficient frequency-domain FWI, the redundancy in the wavenumber coverage is reduced by coarsening the frequency interval, while avoiding wraparound in the space domain (Sirgue and Pratt, 2004). This frequency sub-sampling can be combined with another one performed in the source-receiver space (Gao et al., 2010; Habashy et al., 2011). In the source-receiver space, monochromatic data are stacked with judicious weighting coefficients determined by a singular value decomposition (SVD) of the source-receiver data matrix. This method not only reduces the data volume, but also improves the signal to noise ratio because small eigenvalues, which may correspond to noise, are not taken into account in the inversion. The computational gain

depends on the accuracy of the truncated SVD, which in turn, also depends on the redundancy of the original data set for each frequency. The setback of the sub-sampling methods is that we may not always have a redundant original data set, and not using all information at the same time may be detrimental.

Another possibility for compressing the data volume consists in creating a limited number of super sources by linear combinations of individual shots that are weighted by random coefficients, thanks to the linearity of the solution of the wave equation with respect to the source terms. This approach was first proposed for migration by [Romero et al. \(2000\)](#) using random phase encoding. More recently [Godwin and Sava \(2013\)](#) compared this random encoding method with deterministic methods to choose the coefficients and assemble the sources. In the deterministic methods, the weighting coefficients can be inferred from discrete Fourier transform (DFT) ([Nihei and Li, 2007](#); [Dai et al., 2013](#)), discrete Hartley transform ([Strang and Nguyen, 1996](#)), which resembles a real valued Fourier transform, or appropriate time delays as when plane-wave migration is performed ([Vigh and Starr, 2008](#); [Dai and Schuster, 2013](#)). The motivation behind the projection of the dataset on a Fourier or Hartley basis results because the basis vectors of these transformations are orthonormal. Thus, in the limiting case where the number of super sources is equal to the number of sources, each super source is encoded with a basis vector and the source encoded inverse problem is equivalent to the non encoded one. However, when the number of super sources is less than the number of sources, as should be the case, the encoding matrix has non-zero off diagonal terms. These off diagonal elements create undesirable cross-correlations between different sources, referred to as crosstalk noise. The illustrative and insightful comparison in [Godwin and Sava \(2013\)](#) shows that the DFT and Hartley transform perform better as the number of super sources tends towards the original number of sources, as expected. The plane wave migration produces the poorest results for any number of super sources. The random phase encoding produces the migrated image with the lowest error when a few number of super sources is used. It thus appears that the maximum computational gain can be obtained through the use of random encodings. This is the reason why we focus on this method in this study.

After the pioneering work of [Romero et al. \(2000\)](#), random source encoding has been widely used in time-domain and frequency-domain FWI ([Neelamani et al., 2008](#); [Krebs et al., 2009](#); [Ben Hadj Ali et al., 2011](#); [Schuster et al., 2011](#); [van Leeuwen et al., 2011](#); [Huang and Schuster, 2012](#)). Random codes include phase shifts, time shifts, convolution with random variables, among others. As is the case for migration, the setback of assembling the sources in super sources is the crosstalk generated by the undesired correlations of the incident wavefield emitted by one individual shot with the back-propagated wavefield emitted by the data residuals associated with the other shots involved in the super-shot setting during the gradient building. Regeneration of the random encoding at each non-linear iteration of the FWI is a possible approach to make the sum incoherent as the number of iterations increase. The sum over iterations to make the average term of the crosstalk noise tend to zero is an averaging technique that, as any Monte-Carlo method, has a slow convergence rate of the order $O(1/\sqrt{I})$ where I is the number of iterations ([Nemirovskiĭ and ĩlĕUdin, 1983](#)). In contrast, deterministic methods using all the sources may have better convergence rates as, for example, $O(1/I)$ ([Boyd and Vandenberghe, 2004](#)). Therefore, using a few super sources allows one to dramatically reduce the computational cost of each non-linear iteration of FWI, but the low convergence rate of averaging techniques may impact the overall computational gain. Another setback is that the random encoding strategy is only valid for common receiver acquisitions. Some attempts have been made to handle marine streamer acquisitions by creating, for example, subsets of data that share the same receivers, or using correlation-based misfit function, which are less sensitive to amplitude effects ([Baumstein et al., 2011](#); [Choi and Alkhalifah, 2012](#)).

Alternatively a hybrid technique has been developed to improve the convergence rate of stochastic methods ([Friedlander and Schmidt, 2012](#)). This technique does not use super sources but rather it is based on a batching strategy where an increasing number of randomly chosen

independent sources are used as the number of iterations grows. Since there is no assembling of super sources, no crosstalk noise is generated, and fixed-spread acquisition geometries are no longer required. The inversion resembles a stochastic optimization problem at the beginning of the inversion when a few number of sources are employed. Although the overall convergence rate of this hybrid method is theoretically superior (Friedlander and Schmidt, 2012), van Leeuwen and Herrmann (2012) showed numerically that, for FWI applications, the convergence rate of this method coupled with a l -BFGS minimization is similar to the convergence rate of the random encoding technique combined with a steepest descent algorithm. This may result because the hybrid method does not use all the information from all sources at the beginning of the minimization process. We therefore do not follow a hybrid approach to improve the convergence rate in the present study. Instead, we interface random source encoding methods with second-order optimization algorithms, which are expected to have higher convergence rates.

Theoretically, stochastic optimization is only proven to converge using a steepest-descent optimization algorithm (Robbins and Monro, 1951; Spall, 2003), under certain restrictions on the step-length computed at each iteration. However, in standard FWI, l -BFGS has shown to improve the convergence (Brossier et al., 2009a). This quasi-Newton method approximates the inverse of the Hessian by performing successive rank-2 updates of an initial estimation from the gradients and the models of the previous l iterations (Byrd et al., 1995). The recursive update of the Hessian over iterations in the l -BFGS method can be affected by the source encoding method when the random codes are regenerated at each FWI iteration because a regeneration of the code changes the misfit function and hence its gradient. Therefore, we first study how the l -BFGS method can be coupled with source encoding strategy based on random encodings, and if a computational gain can be expected from this optimization method.

Truncated Newton optimization methods such as Gauss-Newton (GN) or the full Newton (FN) (Métivier et al., 2013b, 2014) are second-order optimization methods that can be considered in the random source encoding framework. From a source encoding point of view, the advantage of these methods relative to l -BFGS is to account for the action of the Hessian only from quantities available at the current iteration. Therefore, the regeneration of the random codes at each non-linear iteration of the inversion is no longer an issue, as it is for the l -BFGS method. The drawback is that the (Gauss-)Newton approaches require additional seismic modelings per iteration. Therefore, we need to assess whether the higher computational cost of one non-linear iteration of the truncated Newton methods can be balanced by an improved convergence rate provided by a more accurate estimation of the Hessian.

3 METHOD

3.1 Full waveform inversion problem

Let us define a space $\Omega \subset \mathbb{R}^2$ as a subsurface medium with spatially varying model parameters $m(x)$ which may be, for example, the density $\rho(x)$ and the P-wave velocity $v_p(x)$ in the acoustic approximation. In the frequency domain, the wavefield $u(x, m, \omega)$ satisfies the wave equation,

$$A(x, m, \omega)u(x, m, \omega) = s(x, \omega), \quad (3.4)$$

where $A(x, m, \omega)$ is the forward modeling operator, which in the acoustic approximation is

$$A(x, m, \omega) = -\frac{1}{\rho(x)v_p(x)^2}\omega^2 - \nabla \cdot \left(\frac{1}{\rho(x)}\nabla u(x, m) \right), \text{ on } \Omega, \quad (3.5)$$

and the source function is denoted by $s(x, \omega)$. We impose a free surface boundary condition at the surface and absorbing boundary conditions on the other boundaries to simulate an infinite

BIBLIOGRAPHY

- Abubakar, A., van den Berg, P. M., and Habashy, T. M. (2004). A Multiplicative Regularization approach for Deblurring Problems. *IEEE Transactions on Image Processing*, 13:1524–1532.
- Abubakar, A., van den Berg, P. M., and Mallorqui, J. J. (2002). Imaging of biomedical data using a multiplicative regularized contrast source inversion method. *IEEE Transactions on Microwave and Techniques*, 50:1761–1770.
- Agudelo, W. (2005). Imagerie sismique quantitative de la marge convergent d’équator-colombie. Université Paris 6.
- Almomin, A. and Biondi, B. (2012). Tomographic full waveform inversion: Practical and computationally feasible approach. *SEG Technical Program Expanded Abstracts 2012*, pages 1–5.
- Amestoy, P., Duff, I. S., and L’Excellent, J. Y. (2000). Multifrontal parallel distributed symmetric and unsymmetric solvers. *Computer Methods in Applied Mechanics and Engineering*, 184(2-4):501–520.
- Ammari, H., Bretin, E., Garnier, J., and Wahab, A. (2011). Time reversal in attenuating acoustic media. *Contemporary Mathematics*, 548:151–163.
- Anagaw, A. Y. and Sacchi, M. D. (2012). Edge-preserving seismic imaging using the total variation method. *Journal of Geophysics and Engineering*, 9(2):138.
- Andrei, N. (2007). A scaled BFGS preconditioned conjugate gradient algorithm for unconstrained optimization. *Applied Mathematics Letters*, 20:645–650.
- Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P., and Virieux, J. (2013). Regularized seismic full waveform inversion with prior model information. *Geophysics*, 78(2):R25–R36.
- Aster, R. C., Borchers, B., and Thurber, C. H. (2005). *Parameter Estimation and Inverse Problems*. Elsevier Academic Press.
- Bansal, R., Routh, P., Krebs, J., Lee, S., Baumstein, A., Anderson, J., Downey, N., Lazaratos, S., Lu, R., and Saldarriaga, S. (2013). Full wavefield inversion of ocean bottom node data. In *EAGE Technical Program Expanded Abstracts 2013*, page We1104.

- Baumstein, A., Ross, W., and Lee, S. (2011). Simultaneous source elastic inversion of surface waves. In *Expanded Abstracts*, page C040. European Association of Geoscientists and Engineers.
- Beale, E. (1972). A derivation of conjugate gradients. *Numerical Methods for non-linear optimization*, pages 39–43.
- Beasley, C. J. (2008). A new look at marine simultaneous sources. *The Leading Edge*, 27(7):914–917.
- Ben Hadj Ali, H., Operto, S., and Virieux, J. (2011). An efficient frequency-domain full waveform inversion method using simultaneous encoded sources. *Geophysics*, 76(4):R109.
- Bérenger, J.-P. (1994). A perfectly matched layer for absorption of electromagnetic waves. *Journal of Computational Physics*, 114:185–200.
- Berkhout, A. J. (2008). Changing the mindest in seismic data acquisition. *The Leading Edge*, 27(7):924–938.
- Bertalmio, M., Vese, L., Sapiro, G., and Osher, S. (2003). Simultaneous structure and texture image inpainting. *Image Processing, IEEE Transactions on*, 12(8):882–889.
- Beylkin, G. (1987). *Mathematical theory for seismic migration and spatial resolution*, pages 291–304. Blackwell scientific publications (Oxford).
- Biondi, B. and Almomin, A. (2013). Tomographic full waveform inversion (TFWI) by combining FWI and wave-equation migration velocity analysis. *The Leading Edge*, September, special section: full waveform inversion:1074–1080.
- Bottou, L. (1991). Stochastic gradient learning in neural networks. In *Proceedings of Neuro-Nîmes*, volume 91.
- Bottou, L. and Bousquet, O. (2011). The tradeoffs of large-scale learning. *Optimization for Machine Learning*, page 351.
- Bottou, L. and Le Cun, Y. (2005). On-line learning for very large data sets. *Applied Stochastic Models in Business and Industry*, 21(2):137–151.
- Boyd, S. P. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.
- Brandwood, D. (1983). A complex gradient operator and its application in adaptive array theory. *Microwaves, Optics and Antennas, IEE Proceedings H*, 130(1):11–16.
- Brossier, R. (2011). Two-dimensional frequency-domain visco-elastic full waveform inversion: Parallel algorithms, optimization and performance. *Computers & Geosciences*, 37(4):444 – 455.
- Brossier, R., Métivier, L., Operto, S., Ribodetti, A., and Virieux, J. (2013a). VTI acoustic equations: a first-order symmetrical PDE. Technical Report Technical report *n° 50*, SEISCOPE project.
- Brossier, R., Operto, S., and Virieux, J. (2009a). Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion. *Geophysics*, 74(6):WCC105–WCC118.
- Brossier, R., Operto, S., and Virieux, J. (2009b). Two-dimensional seismic imaging of the Valhall model from synthetic OBC data by frequency-domain elastic full-waveform inversion. *SEG Technical Program Expanded Abstracts*, 28(1):2293–2297.

- Brossier, R., Operto, S., and Virieux, J. (2010). Which data residual norm for robust elastic frequency-domain full waveform inversion? *Geophysics*, 75(3):R37–R46.
- Brossier, R., Operto, S., and Virieux, J. (2013b). Velocity model building from seismic reflection data by full waveform inversion. *Geophysical Prospecting*, submitted.
- Bunks, C., Salek, F. M., Zaleski, S., and Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473.
- Burstedde, C. and Ghattas, O. (2009). Algorithmic strategies for full waveform inversion: 1d experiments. *Geophysics*, 74(6):WCC37–WCC46.
- Byrd, R. H., Lu, P., and Nocedal, J. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific and Statistical Computing*, 16:1190–1208.
- Campillo, M. and Paul, A. (2003). Long Range correlations in the diffuse seismic coda. *Science* 299, 547.
- Candes, E. and Romberg, J. (2005). l_1 -MAGIC: Recovery of sparse signals via convex programming,. *Technical Report*.
- Caselles, V., Chambolle, A., and Novaga, M. (2011). Total variation in imaging. In *Handbook of Mathematical Methods in Imaging*, pages 1016 – 1057. Springer.
- Castellanos, C., Etienne, V., Hu, G., Operto, S., Brossier, R., and Virieux, J. (2011). Algorithmic and methodological developments towards full waveform inversion in 3d elastic media. *SEG Technical Program Expanded Abstracts*, 30:2793–2798.
- Castellanos, C., Métivier, L., Operto, S., Brossier, R., and Virieux, J. (2013). Fast full waveform inversion with source encoding and second-order optimization methods. *Geophysical Journal International*, submitted.
- Chambolle, A. (2004). An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20:89–97.
- Chambolle, A., Levine, S. E., and Lucier, B. J. (2011). An upwind finite-difference method for total variation-based image smoothing. *SIAM Journal on Imaging Sciences*, 4(1):277–299.
- Chambolle, A. and Lions, P.-L. (1997). Image recovery via total variation minimization and related problems. *Numer. Math.*, 76:167–188.
- Chan, T. and Shen, J. (2005). *Image processing and analysis: variational, PDE, wavelet, and stochastic methods*. Siam.
- Chan, T. F., Osher, S., and Shen, J. (2001). The digital tv filter and nonlinear denoising. *Image Processing, IEEE Transactions on*, 10(2):231–241.
- Chavent, G. (2009). *Nonlinear least squares for inverse problems*. Springer Dordrecht Heidelberg London New York.
- Chen, J. and Schuster, G. T. (1999). Resolution limits of migrated images. *Geophysics*, 64(4):1046–1053.
- Choi and Alkhalifah, T. (2012). Multi-source waveform inversion of marine streamer data using the normalized wavefield. In *Expanded Abstracts*. EAGE.
- Claerbout, J. (1985). *Imaging the Earth’s interior*. Blackwell Scientific Publication.

- Crase, E., Pica, A., Noble, M., McDonald, J., and Tarantola, A. (1990). Robust elastic non-linear waveform inversion: application to real data. *Geophysics*, 55:527–538.
- Crase, E., Wideman, C., Noble, M., and Tarantola, A. (1992). Nonlinear elastic inversion of land seismic reflection data. *Journal of Geophysical Research*, 97:4685–4705.
- Cupillard, P., Stehly, L., and Romanowicz, B. (2011). The one-bit noise correlation: A theory based on the concepts of coherent and incoherent noise. *Geophysical Journal International*, 184(3):1397–1414.
- Dai, W., Huang, Y., and Schuster, G. T. (2013). Least-squares reverse time migration of marine data with frequency-selection encoding. *Geophysics*, 78(4):S233–S242.
- Dai, W. and Schuster, G. T. (2013). Plane-wave least-squares reverse time migration. *Geophysics*, 78(4):S165–S177.
- Darbon, J. and Sigelle, M. (2005). A fast and exact algorithm for total variation minimization. In *Pattern recognition and image analysis*, pages 351–359. Springer.
- De Hoop, M., Fedrizzi, E., Garnier, J., Solna, K., et al. (2012). Imaging with noise blending. *Multi-Scale and High-Contrast PDE: From Modelling, to Mathematical Analysis, to Inversion*, pages 105–124.
- Derode, A., Larose, E., Tanter, M., De Rosny, J., Tourin, A., Campillo, M., and Fink, M. (2003). Recovering the green’s function from field-field correlations in an open scattering medium (1). *The Journal of the Acoustical Society of America*, 113:2973.
- Devaney, A. (1984). Geophysical diffraction tomography. *Geoscience and Remote Sensing, IEEE Transactions on*, (1):3–13.
- Devaney, A. J. and Zhang, D. (1991). Geophysical diffraction tomography in a layered background. *Wave motion*, 14:243–265.
- Djiekpéssé, H. A. and Tarantola, A. (1999). Multiparameter l_1 norm waveform fitting: Interpretation of gulf of mexico reflection seismograms. *Geophysics*, 64(4):1023–1035.
- Engl, H. W., Hanke, M., and Neubauer, A. (2000). *Regularization of Inverse Problems*, volume 35. Kluwer Academic Publishers.
- Erlangga, Y. A. (2005). *A robust and efficient iterative method for the numerical solution of the Helmholtz equation*. PhD thesis, Delft University of Technology.
- Erlangga, Y. A. and Nabben, R. (2008). Deflation and balancing preconditioners for Krylov subspace methods applied to nonsymmetrix matrices. *SIAM Journal on Matrix Analysis and Applications*, 30(2):684–699.
- Fink, M. (1993). Time-reversal mirrors. *Journal of Physics D: Applied Physics*, 26(9):1333–1350.
- Fink, M. (2008). Time-reversal waves and super resolution. In *Journal of Physics: Conference Series*, volume 124, page 012004. IOP Publishing.
- Forgues, E. and Lambaré, G. (1997). Parameterization study for acoustic and elastic ray+born inversion. *Journal of Seismic Exploration*, 6:253–278.
- Freudenreich, Y. and Singh, S. (2000). Full waveform inversion for seismic data - frequency versus time domain. In *EAGE Technical Program Expanded Abstracts 2000*, page C54.

- Friedlander, M. P. and Schmidt, M. (2012). Hybrid deterministic-stochastic methods for data fitting. *To appear in SIAM Journal on Scientific Computing*.
- Gao, F., Atle, A., and Williamson, P. (2010). Full waveform inversion using deterministic source encoding. *SEG Technical Program Expanded Abstracts*, 29(1):1013–1017.
- Garnier, J. and Papanicolaou, G. (2009). Passive sensor imaging using cross correlations of noisy signals in a scattering medium. *SIAM Journal on Imaging Sciences*, 2(2):396–437.
- Gauthier, O., Virieux, J., and Tarantola, A. (1986). Two-dimensional nonlinear inversion of seismic waveform : numerical results. *Geophysics*, 51:1387–1403.
- George, A. and Liu, J. W. (1981). *Computer solution of large sparse positive definite systems*. Prentice-Hall, Inc.
- Gholami, Y., Brossier, R., Operto, S., Prioux, V., Ribodetti, A., and Virieux, J. (2013a). Which parametrization is suitable for acoustic VTI full waveform inversion? - Part 2: application to Valhall. *Geophysics*, 78(2):R107–R124.
- Gholami, Y., Brossier, R., Operto, S., Ribodetti, A., and Virieux, J. (2013b). Which parametrization is suitable for acoustic VTI full waveform inversion? - Part 1: sensitivity and trade-off analysis. *Geophysics*, 78(2):R81–R105.
- Gilbert, J. C. (1997). On the regularization of the Wolfe conditions in reduced quasi Newton methods for equality constrained optimization. *SIAM J. Optim*, 7(3):780–813.
- Gill, P. and Leonard, M. W. (2003). Limited memory reduced Hessian methods for large scale unconstrained optimization. *SIAM J. Optim*, 12:380–401.
- Gill, P. E. and Murray, W. (1979). Conjugate-gradient methods for large-scale nonlinear optimization. Technical report, DTIC Document.
- Godwin, J. and Sava, P. (2013). A comparison of shot-encoding schemes for wave-equation migration. *Geophysical Prospecting*.
- Goldstein, T. and Osher, S. (2009). The split bregman method for l1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343.
- Guitten, A. (2012). Blocky regularization schemes for full waveform inversion. *Geophysical Prospecting*, 60(5):870–884.
- Guitten, A. and Symes, W. W. (2003). Robust inversion of seismic data using the Huber norm. *Geophysics*, 68(4):1310–1319.
- Ha, T., Chung, W., and Shin, C. (2009). Waveform inversion using a back-propagation algorithm and a Huber function norm. *Geophysics*, 74(3):R15–R24.
- Habashy, T. M., Abubakar, A., Pan, G., and Belani, A. (2011). Source-receiver compression scheme for full-waveform seismic inversion. *Geophysics*, 76(4):R95–R108.
- Hale, D. (2013). Dynamic warping of seismic images. *Geophysics*, 78(2):S105–S115.
- Hansen, C. (1998). *Rank-deficient and discrete ill-posed problems - Numerical aspects of linear inversion*. Society for Industrial and Applied Mathematics - Mathematical modeling and computation.
- Herrmann, F. J. and Li, X. (2012). Efficient least-squares imaging with sparsity promotion and compressive sensing. *Geophysical prospecting*, 60(4):696–712.

- Hicks, G. J. (2002). Arbitrary source and receiver positioning in finite-difference schemes using kaiser windowed sinc functions. *Geophysics*, 67:156–166.
- Hou, S., Solna, K., and Zhao, H. (2006). A direct imaging algorithm for extended targets. *Inverse Problems*, 22(4):1151.
- Huang, Y. and Schuster, G. T. (2012). Multisource least-squares migration of marine streamer and land data with frequency-division encoding. *Geophysical Prospecting*, 60(4):663–680.
- Hustedt, B., Operto, S., and Virieux, J. (2004). Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling. *Geophysical Journal International*, 157:1269–1296.
- Jannane, M., Beydoun, W., Crase, E., Cao, D., Koren, Z., Landa, E., Mendes, M., Pica, A., Noble, M., Roeth, G., Singh, S., Snieder, R., Tarantola, A., and Trezeguet, D. (1989). Wavelengths of Earth structures that can be resolved from seismic reflection data. *Geophysics*, 54(7):906–910.
- Kaltenbacher, B., Neubauer, A., and Scherzer, O. (2008). *Iterative Regularization Methods for Nonlinear Problems*. de Gruyter, Berlin, New York.
- Kelley, C. (1999). *Iterative Methods for Optimization*. SIAM.
- Krebs, J., Anderson, J., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A., and Lacasse, M. D. (2009). Fast full-wavefield seismic inversion using encoded sources. *Geophysics*, 74(6):WCC105–WCC116.
- Lailly, P. (1983). The seismic inverse problem as a sequence of before stack migrations. In Bednar, R. and Weglein, editors, *Conference on Inverse Scattering, Theory and application, Society for Industrial and Applied Mathematics, Philadelphia*, pages 206–220.
- Lailly, P. (1984). The seismic inverse problem as a sequence of before stack migrations. In Bednar, R. and Weglein, editors, *Conference on Inverse Scattering, SIAM, Philadelphia*, pages 206–220. Soc. Ind. appl. Math.
- Lambaré, G., Operto, S., Podvin, P., Thierry, P., and Noble, M. (2003). 3-D ray+Born migration/inversion - part 1: theory. *Geophysics*, 68:1348–1356.
- Lambaré, G., Virieux, J., Madariaga, R., and Jin, S. (1992). Iterative asymptotic inversion in the acoustic approximation. *Geophysics*, 57:1138–1154.
- Larmat, C., Montagner, J. P., Fink, M., and Capdeville, Y. (2006). Time-reversal imaging of seismic sources and application to the great Sumatra earthquake. *Geophysical Research Letters*, 33:L19312.
- Laruelle, S. and Pagés, G. (2012). Stochastic approximation with averaging innovation applied to finance. *Monte Carlo Methods Appl.*, 18(1):1–51.
- Lions, J. (1972). *Nonhomogeneous boundary value problems and applications*. Springer Verlag, Berlin.
- Lions, J. L. (1968). *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*. Dunod, Paris.
- Liu, D. C. and Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45:503–528.

- Liu, F., Zhang, G., Morton, S. A., and Leveille, J. P. (2011). An effective imaging condition for reverse-time migration using wavefield decomposition. *Geophysics*, 76(1):S29–S39.
- Loris, I., Douma, H., Nolet, G., Daubechies, I., and Regone, C. (2010). Nonlinear regularization techniques for seismic tomography. *Journal of Computational Physics*, 229:890–905.
- Luo, Y. and Schuster, G. T. (1991). Wave-equation travelttime inversion. *Geophysics*, 56(5):645–653.
- Ma, Y., Hale, D., Meng, Z. J., and Gong, B. (2010). Full waveform inversion with image-guided gradient. *SEG Technical Program Expanded Abstracts*, 29(1):1003–1007.
- Menke, W. (1984). *Geophysical Data Analysis: Discrete Inverse Theory*. Academic Press, Inc., Orlando, USA.
- Mercerat, E. and Nolet, G. (2012). Comparison of ray-and adjoint-based sensitivity kernels for body-wave seismic tomography. *Geophysical Research Letters*, 39(12).
- Mercerat, E. D. and Nolet, G. (2013). On the linearity of cross-correlation delay times in finite-frequency tomography. *Geophysical Journal International*, 192(2):681–687.
- Métivier, L., Breteau, F., Brossier, R., Operto, S., and Virieux, J. (2014). Full waveform inversion and the truncated newton method: quantitative imaging of complex subsurface structures. *Geophysical Prospecting*, In press.
- Métivier, L., Brossier, R., Virieux, J., and Operto, S. (2013a). Full waveform inversion and the truncated newton method. *SIAM Journal On Scientific Computing*, 35(2):B401–B437.
- Métivier, L., Brossier, R., Virieux, J., and Operto, S. (2013b). Full waveform inversion and the truncated newton method. *SIAM Journal On Scientific Computing*, 35(2):B401–B437.
- Miller, D., Oristaglio, M., and Beylkin, G. (1987). A new slant on seismic imaging: Migration and integral geometry. *Geophysics*, 52(7):943–964.
- Montagner, J.-P., Larmatand, C., Capdeville, Y., Fink, M., Phung, H., and Romanowicz, B. (2012). Time-reversal method and cross-correlation techniques by normal mode theory: a three point problem. *Geophysical Journal International*, (191):637–652.
- Montelli, R., Nolet, G., Dahlen, F. A., Masters, G., Engdahl, E. R., and Hung, S. H. (2004). Finite-frequency tomography reveals a variety of plumes in the mantle. *Science*, 303:338–343.
- Mora, P. R. (1987). Nonlinear two-dimensional elastic inversion of multi-offset seismic data. *Geophysics*, 52:1211–1228.
- Mora, P. R. (1988). Elastic wavefield inversion of reflection and transmission data. *Geophysics*, 53:750–759.
- Mora, P. R. (1989). Inversion = migration + tomography. *Geophysics*, 54(12):1575–1586.
- Neelamani, R., Krohn, C. E., Krebs, J. R., Deffenbaugh, M., Anderson, J. E., and Romberg, J. K. (2008). Efficient seismic forward modeling usin simultaneous random sources and sparsity. In *78th Annual SEG Conference & Exhibition, Las Vegas*. Society of Exploration Geophysicists.
- Nemirovski, A. (1999). *Optimization II: Standard Numerical Methods for Nonlinear Continuous Optimization, Lecture Notes*. Technion - Israel Institute of Technology.
- Nemirovskiĭ, A. S. and ėlĕUdin, D. (1983). *Problem complexity and method efficiency in optimization*. Wiley (Chichester and New York).

- Neumaier, A. (1997). On convergence and restart conditions for non-linear conjugate gradient method. *Manuscript*.
- Ng, M. K., Qi, L., Yang, Y.-F., and Huang, Y.-M. (2007). On semismooth newton’s methods for total variation minimization. *Journal of Mathematical Imaging and Vision*, 27(3):265–276.
- Nihei, K. T. and Li, X. (2007). Frequency response modelling of seismic waves using finite difference time domain with phase sensitive detection (TD-PSD). *Geophysical Journal International*, 169:1069–1078.
- Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization*. New York, US : Springer.
- Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, 2nd edition.
- Nolet, G. (2008). *A Breviary of Seismic Tomography*. Cambridge University Press, Cambridge, UK.
- Operto, S., Brossier, R., Gholami, Y., Métivier, L., Prieux, V., Ribodetti, A., and Virieux, J. (2013). A guided tour of multiparameter full waveform inversion for multicomponent data: from theory to practice. *The Leading Edge*, September, Special section Full Waveform Inversion:1040–1054.
- Operto, S., Virieux, J., Amestoy, P., L’Écellent, J.-Y., Giraud, L., and Ben Hadj Ali, H. (2007). 3D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study. *Geophysics*, 72(5):SM195–SM211.
- Operto, S., Virieux, J., Ribodetti, A., and Anderson, J. E. (2009). Finite-difference frequency-domain modeling of visco-acoustic wave propagation in two-dimensional TTI media. *Geophysics*, 74 (5):T75–T95.
- Osher, S., Burger, M., Goldfarb, D., Xu, J., and Yin, W. (2005). An iterative regularization method for total variation based image restoration. *SIAM Multiscale Model Simulation*, 4(2):460–489.
- Plessix, R. E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503.
- Plessix, R. E. (2007). A Helmholtz iterative solver for 3D seismic-imaging problems. *Geophysics*, 72(5):SM185–SM194.
- Plessix, R. E. (2009). Three-dimensional frequency-domain full-waveform inversion with an iterative solver. *Geophysics*, 74(6):WCC53–WCC61.
- Plessix, R.-E., Baeten, G., de Maag, J. W., and ten Kroode, F. (2012). Full waveform inversion and distance separated simultaneous sweeping: a study with a land seismic data set. *Geophysical Prospecting*, 60:733 – 747.
- Plessix, R. E. and Cao, Q. (2011). A parametrization study for surface seismic full waveform inversion in an acoustic vertical transversely isotropic medium. *Geophysical Journal International*, 185:539–556.
- Plessix, R. E. and Perkins, C. (2010). Full waveform inversion of a deep water ocean bottom seismometer dataset. *First Break*, 28:71–78.
- Powell, M. (1977). Restart Procedures for the Conjugate Gradient Method. *Mathematical Programming*, 12:241–254.

- Pratt, R. G. (1990). Frequency-domain elastic modeling by finite differences: a tool for crosshole seismic imaging. *Geophysics*, 55(5):626–632.
- Pratt, R. G., Shin, C., and 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. and Shipp, R. M. (1999). Seismic waveform inversion in the frequency domain, part II: Fault delineation in sediments using crosshole data. *Geophysics*, 64:902–914.
- Pratt, R. G. and Worthington, M. H. (1990). Inverse theory applied to multi-source cross-hole tomography. Part I: acoustic wave-equation method. *Geophysical Prospecting*, 38:287–310.
- Prieux, V., Brossier, R., Gholami, Y., Operto, S., Virieux, J., Barkved, O., and Kommedal, J. (2011). On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study. *Geophysical Journal International*, 187:1495–1515.
- Prieux, V., Brossier, R., Operto, S., and Virieux, J. (2013a). Multiparameter full waveform inversion of multicomponent OBC data from valhall. Part 1: imaging compressional wavespeed, density and attenuation. *Geophysical Journal International*, 194(3):1640–1664.
- Prieux, V., Brossier, R., Operto, S., and Virieux, J. (2013b). Multiparameter full waveform inversion of multicomponent OBC data from valhall. Part 2: imaging compressional and shear-wave velocities. *Geophysical Journal International*, 194(3):1665–1681.
- Pyun, S., Shin, C., and Bednar, J. B. (2007). Comparison of waveform inversion, part 3: amplitude approach. *Geophysical Prospecting*, 55(4):477–485.
- Pyun, S., Shin, C., and Son, W. (2009). Frequency-domain waveform inversion using an L1-norm objective function. In *Expanded Abstracts*, page P005. EAGE.
- Ramírez, A. C. and Lewis, W. R. (2010). Regularization and full-waveform inversion: A two-step approach. In *2010 SEG Annual Meeting*.
- Ravaut, C., Operto, S., Imbrota, L., Virieux, J., Herrero, A., and dell’Aversana, P. (2004). Multi-scale imaging of complex structures from multi-fold wide-aperture seismic data by frequency-domain full-wavefield inversions: application to a thrust belt. *Geophysical Journal International*, 159:1032–1056.
- Ribodetti, A., Operto, S., Agudelo, W., Collot, J.-Y., and Virieux, J. (2011). Joint ray+born least-squares migration and simulated annealing optimization for high-resolution target-oriented quantitative seismic imaging. *Geophysics*, 76(2):R23.
- Robbins, H. and Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407. Mathematical Reviews number (MathSciNet): MR42668; Zentralblatt MATH identifier: 0054.05901.
- Romero, L. A., Ghiglia, D. C., Ober, C. C., and Morton, S. A. (2000). Phase encoding of shot records in prestack migration. *Geophysics*, 65, (2):426–436.
- Roosta-Khorasani, F. and Ascher, U. (2013). Improved bounds on sample size for implicit matrix trace estimators. *Computing Research Repository*.
- Routh, P., Krebs, J., Lazaratos, S., Baumstein, A., Lee, S., Cha, Y. H., Chikichev, I., Downey, N., Hinkley, D., and Anderson, J. (2011). Encoded simultaneous source full-wavefield inversion for spectrally shaped marine streamer data. In *2011 SEG Annual Meeting*.

- Roux, N. L. and Fitzgibbon, A. W. (2010). A fast natural newton method. In Fürnkranz, J. and Joachims, T., editors, *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 623–630. Omnipress.
- Rudin, L., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268.
- Saad, Y. (2003). *Iterative methods for sparse linear systems*. SIAM, Philadelphia.
- Sava, P. and Fomel, S. (2005). Coordinate-independent angle-gathers for wave equation migration. *75th SEG meeting, Houston, Texas, USA, Expanded Abstracts*, pages 2052–2055.
- Sava, P. and Fomel, S. (2006). Time-shift imaging condition in seismic migration. *Geophysics*, 71(6):S209–S217.
- Schiemenz, A. and Igel, H. (2013). Accelerated 3-d full-waveform inversion using simultaneously encoded sources in the time domain: application to valhall ocean-bottom cable data. *Geophysical Journal International*, 195(3):1970–1988.
- Schittkowski, K. (2011). A robust implementation of a sequential quadratic programming algorithm with successive error restoration. *Optimization Letters*, 5:283–296. 10.1007/s11590-010-0207-9.
- Schraudolph, N. N., Yu, J., and Günter, S. (2007). A stochastic quasi-newton method for online convex optimization. In *In Proceedings of 11th International Conference on Artificial Intelligence and Statistics*.
- Schuster, G. T. (1996). Resolution limits for crosswell migration and travelttime tomography. *Geophysical Journal International*, 127(2):427–440.
- Schuster, G. T. (2007). *Basics of seismic wave theory*. University of Utah.
- Schuster, G. T., Wang, X., Huang, Y., Dai, W., and C., B. (2011). Theory of multisource crosstalk reduction by phase encoded statics. *Geophysical Journal International*, 184:1289–303.
- Shapiro, N. M., Campillo, M., Stehly, L., and Ritzwoller, M. H. (2005). High-Resolution Surface-Wave Tomography from Ambient Seismic Noise. *Science*, 307(5715):1615.
- Shen, P. and Symes, W. W. (2008). Automatic velocity analysis via shot profile migration. *Geophysics*, 73(5):VE49–VE59.
- Sheng, J., Leeds, A., Buddensiek, M., and Schuster, G. T. (2006). Early arrival waveform tomography on near-surface refraction data. *Geophysics*, 71(4):U47–U57.
- Shewchuk, J. R. (1994). An introduction to the conjugate gradient method without the agonizing pain. Technical Report Computer Science Technical Report CMU-CS-94-125, School of computer science, Carnegie Mellon University.
- Shin, C. and Ha, W. (2008). A comparison between the behavior of objective functions for waveform inversion in the frequency and laplace domains. *Geophysics*, 73(5):VE119–VE133.
- Shin, C., Jang, S., and Min, D. J. (2001a). Improved amplitude preservation for prestack depth migration by inverse scattering theory. *Geophysical Prospecting*, 49:592–606.
- Shin, C. and Min, D.-J. (2006). Waveform inversion using a logarithmic wavefield. *Geophysics*, 71(3):R31–R42.

- Shin, C., Min, D.-J., Marfurt, K. J., Lim, H. Y., Yang, D., Cha, Y., Ko, S., Yoon, K., Ha, T., and Hong, S. (2002). Traveltime and amplitude calculations using the damped wave solution. *Geophysics*, 67:1637–1647.
- Shin, C., Pyun, S., and Bednar, J. B. (2007). Comparison of waveform inversion, part 1: conventional wavefield vs logarithmic wavefield. *Geophysical Prospecting*, 55(4):449–464.
- Shin, C., Yoon, K., Marfurt, K. J., Park, K., Yang, D., Lim, H. Y., Chung, S., and Shin, S. (2001b). Efficient calculation of a partial derivative wavefield using reciprocity for seismic imaging and inversion. *Geophysics*, 66(6):1856–1863.
- Sirgue, L. (2003). *Inversion de la forme d’onde dans le domaine fréquentiel de données sismiques grand offset*. PhD thesis, Université Paris 11, France - Queen’s University, Canada.
- Sirgue, L. (2006). The importance of low frequency and large offset in waveform inversion. In *Presented at the 68th EAGE Conference & Exhibition, Vienna, EAGE*, page A037.
- Sirgue, L., Barkved, O. I., Dellinger, J., Etgen, J., Albertin, U., and Kommedal, J. H. (2010). Full waveform inversion: the next leap forward in imaging at Valhall. *First Break*, 28:65–70.
- Sirgue, L., Barkved, O. I., Gestel, J. P. V., Askim, O. J., and Kommedal, J. H. . (2009). 3D waveform inversion on Valhall wide-azimuth OBC. In *Presented at the 71th Annual International Meeting, EAGE, Expanded Abstracts*, page U038.
- Sirgue, L. and Pratt, R. G. (2004). Efficient waveform inversion and imaging : a strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248.
- Snieder, R. (1998). The role of nonlinearity in inverse problems. *Inverse Problems*, 14:387.
- Snieder, R. and Trampert, J. (2000). Linear and nonlinear inverse problems. In *Geomatic Method for the Analysis of Data in the Earth Sciences*, volume 95 of *Lecture Notes in Earth Sciences*, pages 93–164. Springer Berlin Heidelberg.
- Spall, J. C. (2003). *Introduction to Stochastic Search and Oprimization: Estimation, Simulation and Control*. Wiley-Interscience Series in Discrete Mathematics and Optimization. 1st edition.
- Stehly, L., Fry, B., Campillo, M., Shapiro, N., Guilbert, J., Boschi, L., and Giardini, D. (2009). Tomography of the Alpine region from observations of seismic ambient noise. *Geophysical Journal International*, 178:338–350.
- Strang, G. and Nguyen, T. (1996). *Wavelets and filter banks*. Wellesley-Cambridge Press, Wellesley, MA.
- Symes, W. W. (2007). Reverse time migration with optimal checkpointing. *Geophysics*, 72(5):SM213–SM221.
- Symes, W. W. (2008). Migration velocity analysis and waveform inversion. *Geophysical Prospecting*, 56:765–790.
- Tarantola, A. (1984a). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266.
- Tarantola, A. (1984b). Linearized inversion of seismic reflection data. *Geophysical Prospecting*, 32:998–1015.
- Tarantola, A. (1986). A strategy for non linear inversion of seismic reflection data. *Geophysics*, 51(10):1893–1903.

- Tarantola, A. (2005). *Inverse Problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics, Philadelphia.
- Tarantola, A. et al. (1984). The seismic reflection inverse problem. *Inverse problems of acoustic and elastic waves*, pages 104–181.
- Tarantola, A. and Valette, B. (1982). Generalized nonlinear inverse problems solved using the least square criterion. *Reviews of Geophysical and Space Physics*, 20:219–232.
- Thierry, P., Operto, S., and Lambaré, G. (1999). Fast 2D ray-Born inversion/migration in complex media. *Geophysics*, 64(1):162–181.
- Tikhonov, A. and Arsenin, V. (1977). *Solution of ill-posed problems*. Winston, Washington, DC.
- Tromp, J., Tape, C., and Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, 160:195–216.
- van Gestel, J., Kommedal, J., Barkved, O., Mundal, I., Bakke, R., and Best, K. (2008). Continuous seismic surveillance of Valhall field. *The Leading Edge*, pages 1616–1621.
- van Leeuwen, T., Aravkin, A., and Herrmann, F. (2010). Seismic waveform inversion by stochastic optimization. *Tech. Rep. TR-2010-5*.
- van Leeuwen, T., Aravkin, A. Y., and Herrmann, F. J. (2011). Seismic waveform inversion by stochastic optimization. *International Journal of Geophysics*, Volume 2011, ID 689041:18 pages.
- van Leeuwen, T. and Herrmann, F. (2012). Fast waveform inversion without source-encoding. *Geophysical Prospecting*, 61(s1):10–19.
- Vese, L. A. and Osher, S. J. (2003). Modeling textures with total variation minimization and oscillating patterns in image processing. *Journal of Scientific Computing*, 19(1-3):553–572.
- Vigh, D., Moldoveanu, N., Jiao, K., Huang, W., and Kapoor, J. (2013). Ultralong-offset data acquisition can complement full-waveform inversion and lead to improved subsalt imaging. *The Leading Edge*, 32(9):1116–1122.
- Vigh, D. and Starr, E. W. (2008). 3D prestack plane-wave, full waveform inversion. *Geophysics*, 73:VE135–VE144.
- Virieux, J. (1984). SH wave propagation in heterogeneous media, velocity-stress finite difference method. *Geophysics*, 49:1259–1266.
- Virieux, J. (1986). P-SV wave propagation in heterogeneous media, velocity-stress finite difference method. *Geophysics*, 51:889–901.
- Virieux, J. and Operto, S. (2009). An overview of full waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26.
- Virieux, J. and Operto, S. (2010). *An overview of full waveform inversion in exploration geophysics*, page ISBN: 9781560802266 (13). Society of Exploration Geophysics.
- Vogel, C. (2002). *Computational methods for inverse problems*. Society of Industrial and Applied Mathematics, Philadelphia.
- Vogel, C. R. and Oman, M. E. (1996). Iterative methods for total variation denoising. *Society for Industrial and Applied Mathematics Journal on Scientific Computing*, 17(1):227–238.

- Wang, C., Chen, X., Smola, A., and Xing, E. (2013a). Variance reduction for stochastic gradient optimization. In *Advances in Neural Information Processing Systems*, pages 181–189.
- Wang, F., Chauris, H., Donno, D., and Calandra, H. (2013b). Taking advantage of wave field decomposition in full waveform inversion. In *EAGE Technical Program Expanded Abstracts 2013*, page Tu0708.
- Whitney, M. L. (2009). *Theoretical and Numerical Study of Tikhonov’s Regularization and Morozov’s Discrepancy Principle*. PhD thesis, Georgia State University, Department of Mathematics and Statistics.
- Williamson, P. R. (1990). Tomographic inversion in reflection seismology. *Geophysical Journal International*, 100:255–274.
- Wittman, T. (2012). Lecture notes on mathematical imaging processing. In *Medical Imaging*. Fields Institute for Mathematics Research, Toronto.
- Woodward, M. J. (1992). Wave-equation tomography. *Geophysics*, 57:15–26.
- Wu, R. S. and Toksöz, M. N. (1987). Diffraction tomography and multisource holography applied to seismic imaging. *Geophysics*, 52:11–25.
- Xu, S., Chauris, H., Lambaré, G., and Noble, M. (2001). Common-angle migration: a strategy for imaging complex media. *Geophysics*, 66:1877–1894.
- Zhang, Y. and Sun, J. (2009). Practical issues in reverse time migration: true amplitude gathers, noise removal and harmonic source encoding. *First break*, 27:53–59.
- Zhu, C., Byrd, R. H., and Nocedal, J. (1997). L-bfgs-b: Algorithm 778: L-bfgs-b, fortran routines for large scale bound constrained optimization. *ACM Transactions on Mathematical Software*, 23(4):550–560.

