THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à MINES ParisTech

Iterative Migration Velocity Analysis: extension to surface-related multiple reflections

Analyse de vitesse par migration itérative : vers une meilleure prise en compte des réflexions multiples

École doctorale n°398

GÉOSCIENCES, RESSOURCES NATURELLES ET ENVIRONNEMENT

Spécialité GÉOSCIENCES ET GÉOINGÉNIERIE

COMPOSITION DU JURY :

M. Jean-Pierre Vilotte Professeur, IPGP	Président du jury
M. Gilles Lambaré Directeur de Recherche, CGG	Rapporteur
M. William Symes Professeur, Rice University	Rapporteur
M. Ludovic Métivier Chargé de Recherche CNRS	Examinateur
M. René-Édouard Plessix Chercheur Sénior, Shell	Examinateur
M. Hervé Chauris Professeur, MINES ParisTech	Examinateur

Soutenue par **Emmanuel COCHER**
le 3 mars 2017

Dirigée par **Hervé CHAURIS**

MINES ParisTech | PSL RESEARCH UNIVERSITY PARIS

Abstract

Seismic waves are widely used for the characterisation of the subsurface as they can travel over long distances and image structures of interest for the oil and gas industry. Signals recorded at the surface during active seismic experiments are analysed to recover parameters governing the wave propagation, such as the pressure wave velocity field. Several techniques aiming at resolving this inverse problem rely on the assumption of a scale separation between rapidly varying components of the velocity model, the reflectivity model, and a smooth macro-model controlling the kinematics of wave propagation. *Migration* algorithms have been developed for the recovery of the reflectivity model but they require a good estimate of the macro-model to deliver accurate images. Analysis of reflectivity images may in turn give indications about the accuracy of the estimated macro-model: this is the principle of Migration Velocity Analysis (MVA) techniques, based on the analysis of primary reflection data. The formulation of MVA considered here is based on *extended* reflectivity images depending on an additional parameter called *subsurface-offset*, a spatial delay introduced during the construction of the reflectivity image. The velocity analysis is carried on panels called Common Image Gathers (CIGs), function of depth and of the subsurface-offset, an accurate macro-model corresponding to energy focused at zero subsurface-offset in CIGs. Using Differential Semblance Optimisation (DSO), the inversion strategy consists of finding the macro-model best focusing events in CIGs. However the method is known to have difficulties in the presence of coherent noise such as multiple reflections, whose removal from recorded data remains a challenging problem. Multiples falsely interpreted as primary reflections result in spurious events in CIGs which do not necessarily focus for the correct macro-model, misleading subsequent velocity analysis. Even with only primary reflections, CIGs are altered with *migration artefacts*, also not focused for the correct macro-model, due to the limited extension of any seismic survey. Recent studies suggest replacing standard migration, which consists of applying the adjoint of the modelling operator to observed data, by inversion. These techniques have been shown to attenuate migration artefacts, but inversion formulas do not necessarily exist for multiples.

In this thesis, I propose to use iterative migration to build CIGs free of migration and multiple artefacts. This makes MVA a nested optimisation procedure: in the inner loop, the reflectivity should minimise data misfit for a fixed value of the macro-model, while the defocused energy in CIGs is minimised in the outer loop. The improvements but also instabilities brought by this strategy are shown and analysed in the case of primaries only and then in the case of first-order surface related multiples modelled with a second-order Born approximation. To limit the additional computational cost of iterative migration, a preconditioner is defined to reduce the number of inner iterations. The preconditioner is a newly developed pseudo-inverse of the modelling operator, derived for primary reflections only. An issue with iterative migration is the apparition of low energy events at large values of h in CIGs. These events have little impact on data misfit but large influence on the subsequent velocity analysis and its stability.

This issue may be mitigated by regularising the inner problem. We also propose a more robust solution applicable in the nested optimisation process, using a slightly modified version of the usual MVA objective function.

This methodology is developed and analysed on synthetic data sets obtained under the constant density acoustic approximation, first with primary reflections only, then with primary reflections and first-order surface multiples. Finally we consider a series of synthetic data sets obtained with modifications of the modelling engine such as the introduction of variable density or the replacement of second-order Born modelling by a free-surface condition, to test the robustness of the approach.

The main contribution of this work is (1) the analysis of existing limitations related to image-domain methods and (2) the propositions for a more robust scheme. Both traditional (pure primary reflections) and extended (incorporation of multiples) approaches benefit from these developments.

Résumé

L'analyse de la propagation des ondes sismiques dans le sous-sol est une technique courante pour la caractérisation de propriétés de la subsurface car ces ondes peuvent parcourir de grandes distances et imager des structures d'intérêt pour l'industrie pétrolière. Dans la configuration classique, sources et récepteurs sont placés à la surface. Les signaux sismiques enregistrés par les récepteurs sont analysés pour déterminer la valeur de paramètres qui gouvernent la propagation des ondes dans le sous-sol. Le paramètre considéré dans cette étude est le champ de vitesse de propagation des ondes P. De nombreuses stratégies de résolution de ce problème inverse supposent une décomposition du modèle de vitesse en deux composantes : un « macro-modèle » lisse qui explique la cinématique de propagation des ondes d'une part, et un modèle de réflectivité représentant les variations rapides du modèle de vitesse responsables des réflexions d'autre part. Le modèle de réflectivité est classiquement obtenu par des algorithmes de *migration*. Ceux-ci nécessitent une bonne connaissance du macro-modèle pour obtenir une image fidèle du sous-sol. En revanche, l'analyse d'un modèle de réflectivité peut donner des indications sur la qualité du macro-modèle de vitesse utilisé pendant l'étape de migration : c'est le principe des méthodes dites d'*analyse de vitesse par migration* (MVA) qui exploitent l'information contenue dans les ondes réfléchies primaires. Dans la variante de MVA considérée dans cette étude, le modèle de réflectivité est paramétré par les coordonnées spatiales et un paramètre additionnel appelé « offset en profondeur », introduit dans la construction de l'image migrée pour capter toute l'information contenue dans les données. L'analyse de vitesse est réalisée sur des panneaux de cohérence appelés *Common Image Gathers* (CIGs), fonction de la profondeur et de l'offset en profondeur : le critère de qualité du macro-modèle de vitesse est la focalisation de l'énergie à zéro offset dans les CIGs. Suivant la méthode dite d'*optimisation par semblance différentielle* (DSO), l'inversion consiste à trouver le modèle de vitesse qui focalise au mieux les évènements dans les CIGs. Cette technique est sensible à la présence dans les données de bruit cohérent comme les réflexions multiples, dont l'atténuation dans les données demeure une étape délicate de prétraitement des données. L'interprétation de réflexions multiples en tant que primaires se traduit par la présence d'artefacts dans les CIGs qui ne focalisent pas nécessairement pour le bon macro-modèle de vitesse. Même dans le cas de primaires seuls, des artefacts dits « de migration » apparaissent sur les CIGs en raison de l'extension finie de sources et récepteurs à la surface rencontrées lors d'acquisitions de données réelles. Ces artefacts contredisent le critère de cohérence utilisé pour remettre à jour le macro-modèle et compromettent alors le succès de l'inversion. Récemment plusieurs études ont suggéré de construire le modèle de réflectivité par inversion des données observées en remplacement de l'étape de migration, qui consiste mathématiquement à appliquer l'adjoint de l'opérateur de modélisation aux données. Cette stratégie permet d'atténuer les artefacts de migration ; mais si de nombreuses formules d'inversion existent pour les primaires, leur transposition aux réflexions multiples n'est pas nécessairement possible.

Dans cette étude, je propose d'atténuer les artefacts de migrations et ceux dus aux multiples en construisant les CIGs par *migration itérative* des données. La MVA peut alors être vue comme un problème d'optimisation imbriqué : dans la boucle interne, la réflectivité est obtenue en minimisant la différence entre données observées et données calculées pour une valeur fixe du macro-modèle, tandis que le macro-modèle est remis à jour dans la boucle externe pour minimiser l'énergie défocalisée dans les CIGs. Les avantages mais aussi les instabilités de cette stratégie sont illustrés et analysés d'abord dans le cas de primaires seuls, puis avec ajout des multiples de surface du premier ordre, modélisés par l'approximation de Born du second-ordre. Pour réduire le coût de calcul supplémentaire apporté par la migration itérative, un préconditionneur est introduit pour limiter le nombre d'itérations nécessaires dans la boucle interne. Le préconditionneur utilisé ici est une nouvelle proposition de pseudo-inverse de l'opérateur de modélisation des réflexions primaires. Il permet également d'accélérer la convergence de la migration itérative dans le cas de multiples.

Une difficulté liée à la migration itérative est l'apparition d'évènements peu énergétiques pour des grandes valeurs de l'offset en profondeur. Ceux-ci ont une faible incidence sur la boucle d'optimisation interne mais ont une influence néfaste sur l'analyse de vitesse et compromettent sa stabilité. Cette difficulté est atténuée en ajoutant une régularisation sur le modèle de réflectivité. Je propose aussi une alternative plus robuste appliquée à la méthode d'optimisation imbriquée, elle consiste en une modification de la fonction coût habituelle de la MVA.

Cette stratégie est testée sur des jeux de données synthétiques obtenus sous l'approximation acoustique à densité constante de l'équation des ondes, d'abord avec des réflexions primaires seules, puis avec des multiples de surface du premier ordre. Enfin sa robustesse est évaluée en l'appliquant à des données obtenues en modifiant le code de modélisation des données observées, en introduisant par exemple un modèle de densité variable ou en modélisant les multiples avec une condition de surface libre.

Les principales contributions de cette étude sont (1) l'analyse des limitations actuelles des méthodes d'inversion définies dans le domaine image ; (2) des propositions de stratégies plus robustes vis-à-vis de ces difficultés, qui bénéficient à la fois au cas classique (réflexions primaires seules) et au cas étendu aux multiples de surface.

Remerciements

La réalisation de ce travail de thèse doit beaucoup aux nombreuses personnes que j'ai côtoyées au cours de ces trois années et demie passées au laboratoire de géophysique de Fontainebleau. Je pense en premier lieu à mon directeur de thèse Hervé Chauris. Je le remercie pour son accompagnement et son soutien continu. Sa patience, sa très grande disponibilité, sa réactivité et son enthousiasme m'ont été indispensables pour continuer à avancer et progresser malgré les difficultés rencontrées. Je souhaite également remercier René-Édouard Plessix pour nos rencontres régulières au cours de la thèse qui ont apporté un éclairage différent aux problèmes rencontrés et qui les ont replacés dans un cadre plus général. Je remercie aussi Shell d'avoir financé la thèse. Merci également aux autres membres du jury, William Symes, Gilles Lambaré, Ludovic Métivier et Jean-Pierre Vilotte pour avoir accepté d'examiner ce manuscrit et pour leurs remarques pertinentes dans les rapports et lors de la soutenance.

Je pense également à Véronique Lachasse dont l'aide au quotidien a été très précieuse, aux permanents de l'équipe géophysique, Daniela Donno, Alexandrine Gesret, Mark Noble et Pierre Dublanchet, ainsi qu'aux « visiteurs scientifiques » Nidhal Belayouni et Pierre-François Roux, qui participent tous à la vie et à la dynamique du labo. Enfin, je souhaite aussi remercier les doctorants, post-doctorants et stagiaires que j'ai croisés au cours de ces quelques années : Carlos Pérez, Élise Vi Nhu Ba, Fang Wang, Charles-Antoine Lameloise, Sven Schilke, Yves-Marie Batany, Tiago Barros, Jihane Belhadj, Julien Cotton, Yubing Li, Keurfon Luu, Hao Jiang, Alexandre Kazantsev, Michelle Almakari, Tianyou Zhou, Abed Benaichouche, Maxime Godano, Shanker Krishna, Fares Mehouachi, Zufri, Bénédicte Loubamono et Joshua Lartey. J'ai une pensée particulière pour Charles-Antoine qui m'a accueilli au labo, introduit au monde de la géophysique et dont les résultats de thèse ont été indispensables pour mon travail, ainsi que pour Yves-Marie avec qui j'ai parcouru la fameuse « dernière ligne droite ».

Finalement, je ne peux pas ne pas mentionner tous mes amis, que je ne listerai pas de peur d'en oublier, et bien sûr mes proches, en particulier mes parents, pour leur soutien sans faille.

Contents

Chapter 1.

Introduction

Contents

Résumé du chapitre 1

Les méthodes d'imagerie sismique sont couramment employées par l'industrie pétrolière pour caractériser les matériaux qui constituent le sous-sol. Elles mettent en œuvre une source sismique active et un ensemble de récepteurs. Plusieurs géométries d'acquisition sont possibles. Dans ce manuscrit, la source et les récepteurs sont supposés placés en surface et l'enregistrement de données est répété pour plusieurs positions de la source en surface. Après prétraitement, les données sont analysées pour déterminer la valeur des propriétés physiques du sous-sol qui gouvernent la propagation des ondes ; il s'agit alors de résoudre un problème inverse.

Dans le cas le plus courant, on cherche à déterminer uniquement le champ de vitesse de propagation des ondes P. Cela correspond à l'approximation acoustique à densité constante de l'équation des ondes. La formulation la plus naturelle pour ce problème inverse est l'inversion des formes d'ondes (Full Waveform Inversion, FWI). Elle consiste en la minimisation itérative d'une fonction coût qui mesure l'écart entre les données observées et des données simulées en résolvant l'équation des ondes avec l'estimation courante du modèle de vitesse. En raison du coût numérique élevé de la simulation de données calculées, la minimisation de cette fonction coût est réalisée par des méthodes d'optimisation locale utilisant le gradient de la fonction coût. Une première difficulté est que la fonction coût de la FWI présente de nombreux minima locaux. Un bon modèle de départ est alors nécessaire pour assurer que le processus d'optimisation converge vers le minimum global. À défaut, la présence de basses fréquences dans les données observées ou d'enregistrements à grands offsets (distance source-récepteur) permet de définir des stratégies hiérarchiques limitant le risque de converger vers un minimum secondaire. Une deuxième difficulté est la reconstruction des grandes longueurs d'onde du modèle de vitesse. Dans la partie peu profonde du modèle, elles sont retrouvées grâce aux ondes plongeantes. En revanche, seules les ondes réfléchies illuminent la partie plus profonde du modèle ; dans le gradient de la fonction coût, ces ondes apportent des informations sur les variations rapides du modèle de vitesse comme la position des interfaces mais ne remettent pas à jour les grandes longueurs d'ondes du modèle.

Des stratégies alternatives ont été définies pour contourner ces difficultés. Elles supposent que l'on peut décomposer le modèle de vitesse en une partie lisse, appelée « macro-modèle » et responsable de la cinématique de propagation des ondes, et une partie hautes fréquences, représentant les variations rapides du modèle de vitesse et appelée « réflectivité ». Ces deux modèles sont déterminés séparément. Le modèle de réflectivité peut être retrouvé par des techniques dites de « migration », définies comme des algorithmes qui repositionnent les interfaces dans le sous-sol à partir de l'information contenue dans les ondes réfléchies enregistrées dans les données. Classiquement, une estimation du macro-modèle est utilisée pour simuler d'une part la propagation dans le sous-sol de la source sismique émise en surface et d'autre part la rétropropagtion des données enregistrées par les récepteurs. Ces deux champs d'onde coïncident à la position des réflecteurs et l'image « migrée » est obtenue par corrélation croisée de ces deux champs d'onde. Le modèle de réflectivité obtenu dépend alors fortement du macro-modèle utilisé pour calculer la propagation des champs d'onde.

L'estimation du macro-modèle fait couramment appel aux méthodes de « tomographie ». Elles sont définies à l'aide de la théorie des rais qui fait l'hypothèse d'une approximation hautes fréquences de l'équation des ondes : les évènements enregistrés dans les données sont alors

caractérisés uniquement par le temps de trajet source-récepteur en laissant de côté l'amplitude et la signature des signaux sismiques. Définies dans le domaine temps, les méthodes de tomographie minimisent l'écart entre des temps de trajet calculés avec une estimation du modèle de vitesse et des temps de trajet pointés dans les données enregistrées. Elles sont couramment utilisées dans l'industrie et de nombreuses évolutions ont été proposées telles les tomographies de pente et la stéréotomographie.

L'autre famille de méthodes visant à estimer le macro-modèle est connue sous le nom d'« analyse de vitesse par migration » (Migration Velocity Analysis, MVA). Ces méthodes exploitent la redondance d'information contenue dans les données sismiques. Ces techniques sont définies dans le domaine image et peuvent utiliser la théorie des rais ou considérer le champ d'onde dans son ensemble. L'analyse se fait sur les ondes réfléchies primaires et repose sur l'approximation de Born, qui définit une relation linéaire entre données réfléchies et modèle de réflectivité. Les différentes variantes de MVA peuvent être classées en deux catégories. Dans l'*orientation surface*, le jeu de données est divisé en sous-ensembles indexés par un paramètre de redondance tel que la position de la source en surface ou la distance entre source et récepteur. Une image de réflectivité est obtenue par migration pour chaque sous-jeu de données : on obtient alors une collection d'images représentant une même région de la subsurface. Si le bon modèle de vitesse est utilisé au cours de l'étape de migration, ces images doivent être semblables. Si le modèle de vitesse est erroné, les réflecteurs peuvent ne pas être positionnés à la même profondeur pour chaque valeur du paramètre de redondance. Ce critère de cohérence permet la définition d'une fonction coût qui mesure la similarité entre les images. Le macro-modèle est déterminé par minimisation de cette fonction coût. Dans l'*orientation profondeur*, étudiée dans cette thèse, un délai spatial et/ou temporel est introduit pendant l'étape de migration lors de la corrélation des champs source et récepteur. Dans cette étude je considérerai uniquement un délai spatial appelé « offset en profondeur » et supposé horizontal. Pour chaque valeur de l'offset en profondeur, une image de réflectivité est construite avec l'ensemble des données. On parle alors de modèle (de réflectivité) « étendu ». L'introduction de ce degré de liberté supplémentaire permet de compenser les erreurs du macro-modèle et conduit à un nouveau critère pour l'évaluation du macro-modèle. Celui-ci est défini sur des panneaux 2D, appelés Common Image Gathers (CIGs), représentant le modèle de réflectivité fonction de la profondeur et de l'offset en profondeur pour une position latérale fixe. Avec le bon modèle de vitesse, l'énergie doit se focaliser autour de la valeur nulle de l'offset en profondeur, qui correspond au modèle de réflectivité qui a un sens physique. La présence d'énergie à des valeurs non nulles de l'offset en profondeur, correspondant à des modèles de réflectivité non physiques, traduit la présence d'erreurs dans le modèle de vitesse. J'utiliserai dans cette étude la stratégie d'« optimisation par semblance différentielle » (DSO). Elle construit le macro-modèle de vitesse par minimisation d'une fonction coût qui pénalise l'énergie défocalisée dans les CIGs. Il a été montré, au moins dans des cas simples, que cette fonction est lisse et ne présente pas de minima secondaires dans une large région autour du minimum global. Le macro-modèle peut alors être obtenu par minimisation de cette fonction coût avec des méthodes d'optimisation locale. Le calcul du gradient associé est une étape essentielle de l'inversion et fera l'objet d'une grande attention dans cette étude.

La formulation profondeur de la DSO est théoriquement attrayante mais peu d'applications sur données réelles ont été publiées jusqu'ici. En plus du coût de calcul numérique de la méthode, plusieurs difficultés l'empêchent d'être plus largement adoptée. Tout d'abord, on constate en

pratique l'apparition dans les CIGs d'artefacts liés à l'extension finie de sources et de récepteurs en surface. Ces artefacts ne focalisent pas pour le bon modèle de vitesse, qui ne constitue alors plus un minimum de la fonction coût de la DSO. Une deuxième difficulté, connue sous le nom d'« artefacts du gradient », est la présence de fortes oscillations dans le gradient de la fonction coût aux troncatures du modèle de réflectivité et autour de la position des réflecteurs. Ces oscillations ne sont pas nécessairement atténuées par un lissage du gradient. Pour s'en affranchir, la méthode de « contraction horizontale » (Fei et Williamson, 2010 ; Shen et Symes, 2015) a été proposée, mais elle présente le désavantage de ne plus définir la remise à jour du modèle de vitesse en tant que gradient d'une fonction objective. Finalement, les méthodes de MVA restent limitées à l'analyse de réflexions primaires seules. Si un formalisme similaire a récemment été défini pour le cas des ondes transmises et des ondes plongeantes, les réflexions multiples sont toujours considérées comme du bruit cohérent. En effet, interprétées comme des primaires elles conduisent à de nouveaux artefacts dans les CIGs qui, de même que les artefacts de migration, ne focalisent pas nécessairement pour le bon modèle de vitesse.

L'atténuation des réflexions multiples dans les données observées a fait l'objet de nombreuses recherches et demeure un problème ouvert. Plusieurs méthodes de *prédiction* des multiples contenus dans un jeu de données ont été proposées, certaines supposant la connaissance d'un modèle de vitesse. La méthode SRME (Surface Related Multiple Elimination, Verschuur et Berkhout, 1997) est aujourd'hui parmi les plus populaires et ne requiert aucune connaissance sur le sous-sol. En revanche, elle suppose une couverture dense de sources et récepteurs en surface. Quelle que soit la méthode employée, la prédiction des multiples n'est en pratique jamais parfaite et une étape de « soustraction adaptative » est nécessaire pour faire correspondre au mieux la prédiction aux multiples présents dans les données. Le danger est alors d'atténuer également l'information contenue dans des réflexions primaires chevauchant des multiples. Depuis quelques années, les multiples sont cependant abordés sous un autre angle : ils peuvent être considérés non pas comme du bruit mais comme une source d'information supplémentaire, notamment en raison de l'illumination différente du sous-sol qu'ils procurent. Des algorithmes de migration prenant en compte les réflexions multiples ont récemment été proposés. Ils utilisent des procédures itératives telles qu'une migration au sens des moindres carrés. L'utilisation de multiples pour la reconstruction du macro-modèle de vitesse a jusqu'à présent été moins explorée.

Dans cette étude, je propose d'itérer l'étape de migration avant de remettre à jour le macro-modèle. La migration itérative (chapitre 2) vise à déterminer un modèle de réflectivité étendu minimisant une fonction coût qui mesure l'écart entre données observées et données calculées. L'objectif est d'atténuer les artefacts de migration et ceux dus aux multiples qui apparaissent à la première itération, équivalente à la migration classique. L'étude est restreinte au cas de multiples de surface du premier ordre modélisés par une approximation de Born du second ordre. L'*analyse de vitesse par migration itérative* ainsi définie est un problème d'optimisation imbriqué à deux niveaux et une méthode de calcul du gradient de la fonction coût externe doit être définie. Deux stratégies sont étudiées dans le chapitre 3. Dans un premier temps la méthode est testée dans le cas de primaires seuls, puis étendue aux multiples. Dans le chapitre 4, elle est comparée, dans le cas de primaires seuls, à une stratégie d'inversion « directe » où l'image migrée est obtenue par application aux données inverse d'un pseudo-inverse de l'opérateur de modélisation dans le domaine étendu. Une nouvelle formule d'inversion est proposée à cette occasion. De même que

celles proposées par ten Kroode (2012) et Hou et Symes (2015, 2017), et contrairement au cas étudié dans Lameloise *et al.* (2014), elle ne fait pas intervenir de quantités liées au rais, même si les calculs intermédiaires reposent sur l'approximation hautes fréquences. Cette stratégie atténue efficacement les artefacts de migration et améliore la qualité de la remise à jour du macro-modèle (section 4.2). Une légère modification de la fonction coût de la DSO permet également de fortement atténuer les oscillations résiduelles localisées au niveau des réflecteurs. Par ailleurs l'introduction du pseudo-inverse comme préconditionneur de la migration itérative permet d'en diminuer le coût numérique. Comparé à l'inversion directe, l'approche itérative produit des résultats de réflectivité similaires. En revanche, le calcul du gradient de la boucle externe se révèle instable : celui-ci est très sensible à de légères modifications du modèle de réflectivité apparaissant entre deux itérations successives et ayant un faible impact sur la fonction objective de la migration. Cette difficulté est illustrée dans le chapitre 3 et des régularisations stabilisant le calcul du gradient sont étudiées dans les chapitres 3 et 5. Finalement la robustesse de cette stratégie est testée dans le chapitre 6 sur des jeux de données synthétiques obtenus avec un code de modélisation différent de celui utilisé pour l'inversion, par exemple avec une ondelette de source différente ou un modèle de densité variable.

In exploration geophysics, seismic imaging consists of characterising the subsurface model parameters such as pressure and shear wave velocity or rock density with the analysis of non-destructive measurements performed at the surface. Emitted seismic waves propagate in all the directions in the subsurface and are distorted during propagation. In simple models, they are reflected and diffracted at interfaces characterised by rapid variations of physical properties of the rocks. Sensors located at the surface record the waves reaching the surface after having interacted with the subsurface and provide the data used by geophysicists. These time recordings usually cannot be used directly for geological interpretation. Surface recordings have to be converted into maps of the subsurface. This crucial step is formulated as the resolution of an inverse problem aiming at finding the value of physical parameters governing the propagation of waves in the subsurface. The resolution of this inverse problem is a difficult step and many methods addressing it have been proposed. Among others issues, data may depend in a non-linear way on the model parameters. This is the case for example when the model corresponds to the large wavelengths of the velocity model or when seismic waves reflect several times in the subsurface. This thesis especially focuses on these two cases.

In this introduction, we first briefly summarize the main steps of seismic imaging used in exploration geophysics and then pay attention to the resolution of the inverse problem. Finally we motivate our work and explain the current limitations of image-domain methods, in particular how multiples can bias traditional imaging techniques. For that, we recap the physics of multiple bouncing waves and explain why multiples may bring valuable information on the subsurface instead of being removed in a pre-processing step.

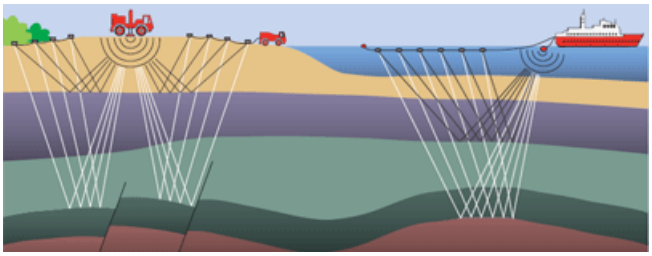
1.1. From seismic acquisition to interpretation

Seismic experiments are in common use in the oil and gas industry for the subsurface imaging and reservoir management of a reservoir. Seismic waves can travel over long distances and are thus well suited to the study of the geologic structures involved in oil production, located at a depth of a few kilometres in the subsurface. Seismic exploration uses active seismic experiments, meaning the source at the origin of wave propagation is artificially triggered, contrary to passive seismic considered in seismology for the study of natural earthquakes. We review the main steps of active seismic experiments in this section. The reader is referred to Sheriff and Geldart (2006) and Yilmaz (2001) for more details.

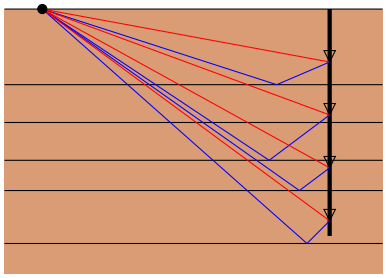
1.1.1. Seismic data acquisition

Seismic surveys can take place in land or marine environments. In the usual land acquisition, the source is a vibrating truck and an array of receivers called geophones measures the motion of particles. They are evenly placed at the surface along one or several lines (figure 1.1a). The experiment is run many times with different source and receiver positions. Under the approximation of single scattered energy, a point of the subsurface is in this way illuminated through many different angles. The presence of a drilled borehole can also provide a different acquisition geometry called Vertical Seismic Profiling (VSP) with sources at the surface and receivers in the well (figure 1.1b). In the case of cross-well acquisition, sources and receivers are located in two different wells. In marine acquisition, the source is an air-gun and receivers

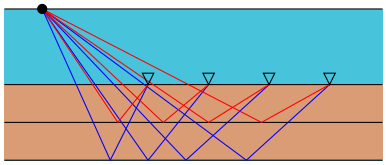
called hydrophones sensitive to the pressure are distributed along several streamers towed by a specific vessel. Alternatively receivers can be located at the sea floor in a display called Ocean Bottom Cable (OBC) (figure 1.1c).



(a) surface acquisition (taken from Danish Energy Agency).



(b) VSP acquisition.



(c) OBC acquisition.

Figure 1.1. – Acquisition geometries for different types of seismic surveys.

The data recorded by a receiver is a function of time and is usually called a *trace*. The collection of traces recorded during an experiment is represented on a panel called *shot gather* with the time on the vertical axis, and the distance between source and receiver (also called offset) on the horizontal axis (figure 1.2).

A shot gather records the Earth's response to the source excitation. Many events corresponding to different kinds of waves can be observed. We can distinguish between pressure waves (P-waves), where particles move parallel to the wave propagation direction, and shear waves (S-waves), where the particle motion is orthogonal to the wave propagation (Aki and Richards, 2002). Besides, energy travels as surface and body waves. Surface waves are more energetic and provide information about the near-surface (Socco and Strobbia, 2004; Pérez Solano, 2013). They are used in seismology but commonly considered as noise in seismic exploration which uses body waves. The latter can be classified according to their path in the subsurface:

- transmitted waves, such as direct and diving waves travel between the source and the receiver without being reflected;
- (primary-)reflected and diffracted waves are generated at discontinuities of the Earth with strong impedance contrast;
- refracted waves travel along these interfaces;

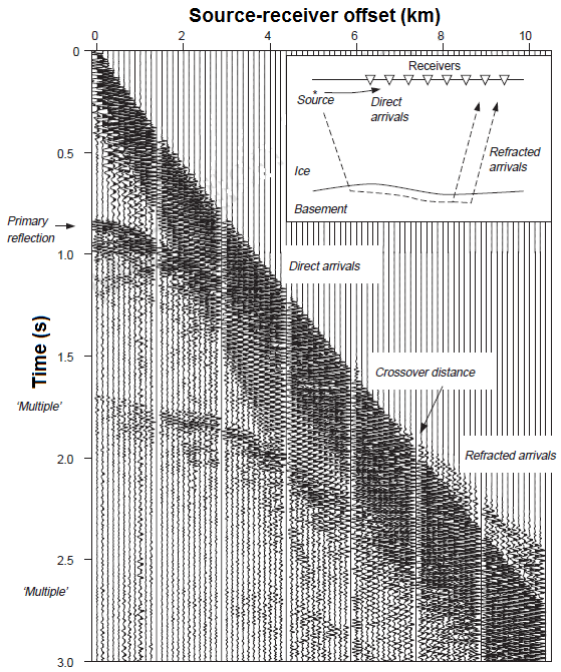


Figure 1.2. – An example of a 2D seismic data recorded in marine acquisition (from Anandakrishnan *et al.*, 1998).

- multiple reflected waves are caused by strong reflectors such as the sea surface or the water bottom. Upgoing energy is sent back to the subsurface and can reflect a few times on the subsurface structures before being recorded.

1.1.2. Physical modelling

For the determination of the subsurface parameters, we first need to define a physical law representative for the wave propagation observed on data recordings. It is mathematically formalised through a partial differential equation. The most general visco-elastic wave equation involves the density, the attenuation of P and S waves and the 21 elastic coefficient of the stiffness tensor relating the stress tensor to the strain tensor. Simplified physics is usually assumed. If the subsurface is considered isotropic, the stiffness tensor reduces to the Lamé parameters λ and μ. The simplest approximation is the isotropic acoustic case. The Earth is assimilated to a fluid parametrised by the P-wave velocity V_P and the density ρ. Then only P-waves are considered and reflected events are caused by rapid variations of the acoustic impedance $I_P = \rho V_P$. Note that the model parameters are function of the spatial coordinates: even for the constant density acoustic case, a large number of parameters should be specified for solving the wave equation.

The simulation of wave propagation is usually performed with a finite-difference resolution of the wave-equation (Virieux, 1986; Levander, 1988; Operto *et al.*, 2007), or with finite element schemes (see Virieux *et al.*, 2011, for a review). Modelling can also rely on a high-frequency

approximation such as ray theory (Červený, 2005). The result of the forward problem is synthetic data calculated for all subsurface points, in particular at the receiver positions.

1.1.3. Preprocessing

Earth's parameters cannot be directly inverted from seismic measurements. Usually, preprocessing is needed before subsequent imaging steps. The first processing stage consists of selecting the wavefield in the data that will be used for the analysis and of attenuating the noise inherent to every practical experiment (Yilmaz, 2001). Many seismic imaging procedures rely on primary reflection data only and consider other events like transmitted waves or multiples as coherent noise. These should be removed from the data before further analysis. In particular, the removal of coherent noise such as multiples has been an intense research topic (Verschuur, 2013) and will be reviewed in section 1.4.3.

However, with the improvement of seismic imaging algorithms and computation capabilities, recent developments attempt to use all the information available in the data, like multiples or transmitted events, and more generally the full wavefield. In particular, the purpose of this study is to investigate the use of multiple reflections as signal rather than noise.

1.1.4. Definition of an objective function

The accuracy of model parameters used during modelling can be assessed by comparing observed data and calculated data. For a quantitative estimation, a scalar *objective* or *cost function* is defined on the space of admissible models. It is designed such that the best model is a global minimiser or maximiser of the function. It can be defined directly in the data-domain, measuring for example the differences between observed and calculated data in the least-squares sense (Tarantola, 1984). Alternative formulations in the image domain detailed in section 1.3 rely on focusing or coherency criteria defined on reconstructed images of the Earth (Al-Yahya, 1989; Symes, 2008). A simple example of image-domain strategy will be given at the end of this section.

The purpose of the inverse problem is to find a set of parameters minimising the objective function. In practice, it is an ill-posed problem, meaning that several models can explain perfectly the data (Tarantola, 2005), due to an imperfect illumination of the subsurface or to coupling between parameters. This issue can be mitigated by adding a regularisation term to the objective function, usually enforcing the smoothness of the recovered model, or its consistency with a priori information about the subsurface (Asnaashari *et al.*, 2013).

1.1.5. Solving the inverse problem

Optimisation strategy

The objective of the inverse problem is to determine a set of model parameters which minimise the objective function through an optimisation procedure. The most general approach involves global optimisation methods (Sen and Stoffa, 2013) which requires only the ability to compute the value of the objective function. Examples of such methods are simulated annealing (Kirkpatrick *et al.*, 1983; Ingber, 1989), genetic algorithms (Holland, 1975) or particle swarm optimisation

(Kennedy and Eberhart, 1995). The drawback of these strategies is their computational cost, as numerous evaluation of the objective function are required.

Local optimisation methods (Nocedal and Wright, 2006) are a less expensive alternative. The gradient of the objective function is used to iteratively update an initial guess of the model, such that the value of the objective function decreases with iterations. Local optimisation requires the ability of computing the gradient of the objective function with respect to the model parameters. The adjoint-state method (Plessix, 2006; Chavent, 2009) provides a computationally efficient way of performing this derivation.

In practice the objective function is not necessarily convex because of the non-linear relationship between data and model parameters: gradient-based method may converge to a local minima. Therefore an other requirement for local optimisation is the knowledge of an initial model sufficiently close to the true model. Inversion strategies considering first the low-frequency content of the data and progressively incorporating higher-frequencies help mitigating the non-linear behaviour of the objective function and relax the requirement of an accurate initial model (Bunks *et al.*, 1995). Strategies to define convex objective functions will be detailed in sections 1.2 and 1.3.

Multiparameter inversion

Most applications of inversion techniques consider a mono-parameter description of wave propagation with the pressure-wave velocity V_P. Moving to a multi-parameter inversion is challenging (Operto *et al.*, 2013). As pointed out before, due to limited acquisition with sources and receivers at the surface, different classes of parameters are coupled, meaning that they have a similar impact on seismic data. This issue is known as *cross-talk* or *trade-off* between parameter. To overcome this difficulty, a suitable parametrisation is needed. In the acoustic example, the parameter couples can be (V_P, ρ) or (V_P, I_P) depending of the acquisition (Virieux and Operto, 2009; Zhou *et al.*, 2015; Zhou, 2016).

Scale separation

In the following we consider a monoparameter inversion with the P-wave velocity model, noted $c(\mathbf{x})$ from now on, as unknown. $\mathbf{x} = (x, y, z)$ or (x, z) denotes the spatial coordinates. The resolution of the velocity model image that can be recovered is limited by the frequency band of the data (typically 5 to 70 Hz) and the data acquisition setup. The model resolution is better described with spatial frequencies. As illustrated by the well-known sketch of Claerbout (1985) (figure 1.3), two separate ranges of spatial frequencies of the model can be reconstructed from seismic data (Jannane *et al.*, 1989), leading to a scale separation of the velocity model (figure 1.4):

- the smooth slowly-varying component of the model (figure 1.4b), called the *background velocity model*, or *macro-model*, controlling the kinematics of wave-propagation;
- rapid changes in the components of the model, responsible for the reflections, referred to as *reflectivity* (figure 1.4c). This part can be physically interpreted as the Earth discontinuities.

Note that since the publication of Claerbout's book, the frequency gap has been progressively filled by improvement of seismic source design allowing to record lower frequencies in the data and the progress of imaging techniques allowing to recover a more detailed background

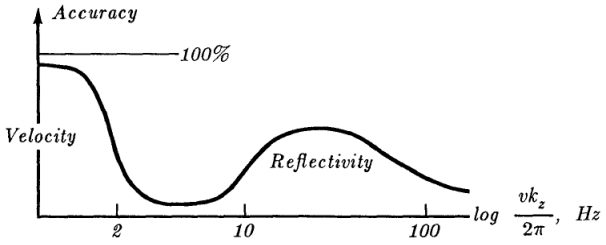
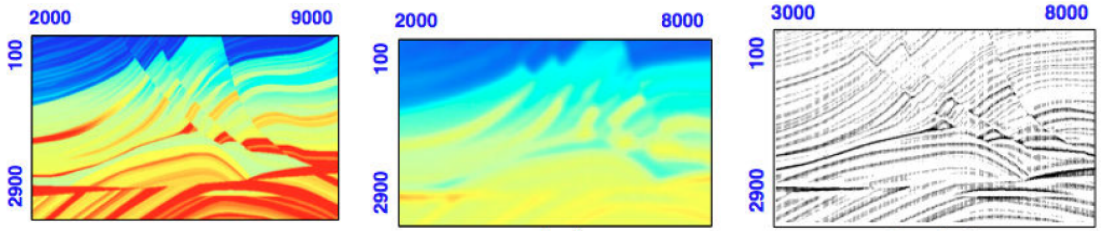


Figure 1.3. – Spatial frequencies that can be resolved from seismic data (from Claerbout, 1985).



(a) Velocity model $c(\mathbf{x})$. **(b)** Background model $c_0(\mathbf{x})$. **(c)** Model perturbation $\delta c(\mathbf{x})$.

Figure 1.4. – Illustration of the usual scale separation for the Marmousi model (Brougois *et al.*, 1990). The complete velocity model $c(\mathbf{x})$ is decomposed into a smooth background part $c_0(\mathbf{x})$ and a perturbation $\delta c(\mathbf{x})$ of the background model c_0 (adapted from Billette, 1998).

velocity model (Nichols, 2012; Lambaré *et al.*, 2014). However, the traditional scale separation remains the theoretical basis of many seismic imaging methods.

The scale separation can be written explicitly

$$c(\mathbf{x}) = c_0(\mathbf{x}) + \delta c(\mathbf{x}), \tag{1.1}$$

where c_0 is the background velocity model and δc is the reflectivity. If $|\delta c| \ll |c_0|$, the reflectivity can be seen as a perturbation of the background velocity model. Under the *Born approximation*, the relationship between primary reflection data and model perturbation δc is linear. Data still depend non-linearly on the background velocity model c_0, though.

The large and short-scale structures of the velocity model can be inverted simultaneously without scale separation, as in Full Waveform Inversion (FWI) presented in section 1.2.1. Alternatively, they can be recovered separately. *Migration* algorithms are designed to convert events recorded in time into a reflectivity map of the model perturbation δc function of depth. Under the Born approximation, the determination of δc is a linear inverse problem. Performing this conversion requires the prior knowledge of the background velocity model. The determination of the long-scale structure of the velocity model is performed with *tomographic* approaches, either in the data-domain (Bishop *et al.*, 1985) or in the image-domain (Symes, 2008). Migration and tomography strategies will be reviewed in sections 1.2.2, 1.2.3 and 1.3.

1.1.6. Postprocessing

Finally the quantitative knowledge of physical properties provided by the inversion is used to determine a model of the subsurface at the reservoir scale. Empirical relationship between V_P and V_S may give indications about the presence of hydrocarbon (Tatham and Stoffa, 1976) and the permeability and porosity of rock materials (Domenico, 1984). This information are used to manage the oil production of the reservoir.

This thesis focuses on the resolution of the inverse problem in the isotropic acoustic approximation of the wave equation. We study a method called Migration Velocity Analysis (MVA), an image-domain technique aiming at recovering the background velocity model. The purpose of the thesis is to investigate the possible inclusion of multiple reflections in this approach, which traditionally considers only primary reflections. The principle and issues of MVA will be detailed in section 1.3. We now review standard time-domain strategies for the resolution of the inverse problem.

1.2. Time-domain methods for the resolution of the inverse problem

We detail in this section time-domain methods addressing the resolution of the inverse problem in the isotropic acoustic approximation of the wave equation, meaning that the unknown is the pressure wave velocity field $c(\mathbf{x})$. Image-domain methods will be described in section 1.3. We first present the FWI strategy, which consider the full recorded traces to recover both the large and short scales of $c(\mathbf{x})$. Then alternative strategies relying on the scale separation assumption (figure 1.4) are detailed. We shall distinguish between methods dedicated to the recovery of the short-scale structures $\delta c(\mathbf{x})$ (section 1.2.2) and those aiming at retrieving the background velocity model $c_0(\mathbf{x})$ (section 1.2.3).

1.2.1. Full Waveform Inversion

A data-fitting procedure

FWI (Virieux and Operto, 2009; Fichtner, 2011) is an iterative procedure considering the complete recorded seismic traces. The associated objective function measures the least-squares misfit between recorded data and simulated data.

$$J_{FWI}[c] = \frac{1}{2}\left\|P^{calc}[c] - P^{obs}\right\|_2^2. \tag{1.2}$$

All type of waves are included in the modelling of the data (direct, diving, reflected, multiply scattered waves). First introduced by Tarantola (1984), the method has gained in popularity with the increase of computer power. In theory different model parameters (velocity, density, attenuation, anisotropy) can be resolved, requiring a modelling engine able to reproduce the physics of wave-propagation as accurately as possible (Warner *et al.*, 2012). However most applications use the acoustic approximation of the wave-equation and are interested in retrieving only the pressure-wave velocity model of the Earth because of the computational cost of elastic modelling and of the challenges of multi-parameter estimation (Operto *et al.*, 2013).

Resolution analysis

As FWI tries to explain the complete data set, it should provide high-resolution images of the subsurface, with both the long and short-scale structures of the velocity model. The contribution of different kinds of waves in the construction of the velocity model can be analysed with the following relationship (Devaney, 1982; Miller *et al.*, 1987),

$$\mathbf{k} = \frac{2\omega}{c_0} \cos\left(\frac{\theta}{2}\right)\mathbf{n}, \tag{1.3}$$

linking the spatial frequency vector or wavenumber $\mathbf{k}$ at point $\mathbf{x}$ to the diffraction angle θ associated to a source-receiver pair, $\mathbf{n}$ being the normalisation of vector $\mathbf{k}$ (figure 1.5). In the shallow part of the subsurface, large diffraction angles obtained for diving waves allow to recover the small wavenumbers, i.e. the large-scale structure of the velocity model. Because of the limited surface offset range, the deeper part of the subsurface is investigated mainly by reflected waves with a small diffraction angle, and only the high-frequency part of the velocity model is recovered. In the deeper part of the subsurface, FWI thus behaves like a non-linear least-squares migration algorithm (Mora, 1988, 1989). Methods extracting the information about the long-scale structure contained in reflected events will be presented in section 1.2.3.

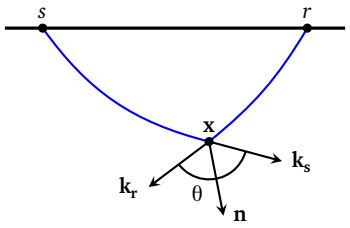


Figure 1.5. – Relationship between the wavenumbers $\mathbf{k_s}$ and $\mathbf{k_r}$ and the opening angle θ at the image point $\mathbf{x}$.

The *cycle-skipping* issue

It is known that the FWI objective function (1.2) suffers from local minima because of the non-linear relationship between model and data (Gauthier *et al.*, 1986). This problem is called *cycle-skipping* and imposes in theory the use of global optimisation procedures (Tognarelli *et al.*, 2015; Galuzzi *et al.*, 2016). These methods are computationally expensive as they require many evaluations of the objective function. Gradient-based methods are less expensive, but they need a starting model close to the solution to avoid converging to a local minimum. More precisely the phase mismatch between data computed with the initial and true model should be less than half the shortest wavelength contained in the data.

To overcome this difficulty, a hierarchical strategy consists of inverting first the low-frequency content of the data and then progressively incorporating higher frequencies, as the basin of attraction of the misfit function is inversely proportional to the central frequency of the data (Bunks *et al.*, 1995; Pratt *et al.*, 1996; Sirgue and Pratt, 2004). The presence of large offsets recording in the surveys allows also to better constrain the background model with transmitted waves (Shipp and Singh, 2002). This hierarchical strategy has been successfully applied on

real data sets in marine (Sirgue *et al.*, 2009; Vigh *et al.*, 2009) and land (Plessix *et al.*, 2010) environments. In typical acquisition geometries, for a target at a few kilometres depth the frequency content of the data should start around 1 Hz to avoid cycle skipping. If low-frequency data or large offsets are not available, other strategies have to be designed to avoid falling in a local minimum.

Alternative techniques modify the definition of the objective function to enlarge the basin of attraction around the correct velocity model. The usual ℓ_2-norm can be replaced by a more convenient distance. Métivier *et al.* (2016) use an optimal transport distance to measure the misfit between seismograms; the basin of attraction of the objective function is extended at the cost of a new optimisation problem dedicated to the calculation of the distance. Other strategies transform the signals of both observed and calculated data to a more convenient domain before subtraction. Shin and Cha (2008, 2009) have studied the use of the Laplace and Laplace-Fourier domain to perform the inversion and showed that an accurate smooth model could be built in these domains. In the Normalised Integration Method (Donno *et al.*, 2013), the transformation consists of integrating the square of the signals and can be interpreted as a measure of the accumulation of energy along the trace. The objective function compares monotone increasing signals and is more convex; however the processing of noise for this method should be investigated. Other methods consider the envelope of seismic data (Wu *et al.*, 2014; Chi *et al.*, 2014) to work on less oscillatory signals. All these transforms allow to mitigate the cycle-skipping issue but produce velocity models with a lower resolution than conventional FWI. However they can be considered as a first inversion step dedicated to the building of an accurate starting model for FWI (Tejero *et al.*, 2015), or for migration algorithms used to recover the short-scale structure of the velocity model.

1.2.2. Linearised waveform inversion

The purpose of migration techniques is to recover a map of Earth discontinuities corresponding to a perturbation of the velocity model. These methods are based on the classic scale separation and assume the knowledge of an estimate of the macromodel. They were historically designed for primary reflections only. Extension to multiple reflections will be presented in section 1.4.4.

The first migration techniques used geometrical construction (Bleistein *et al.*, 2001). In a simple geologic context, an event appearing at time t in a trace corresponding to a source at position $\mathbf{s}$ and a receiver at position $\mathbf{r}$ is due to a diffracting point situated at a location $\mathbf{x}$ such that the sum of traveltimes from $\mathbf{s}$ to $\mathbf{x}$ and from $\mathbf{x}$ to $\mathbf{r}$ is equal to t. All the subsurface points satisfying this criterion define an isochrone curve; in the case of an homogeneous velocity model, it is an ellipse with focus points at $\mathbf{s}$ and $\mathbf{r}$. Repeating the process for all sources and all receivers, the summation of ellipses interferes constructively along the reflectors and destructively anywhere else (Bleistein *et al.*, 2001).

Later the construction of a reflectivity model was reformulated by considering the propagation of complete wavefields and Claerbout (1971) introduced the concept of *Imaging Condition*. Interfaces are defined as the location where a downgoing incident wavefield coincides in time with an upgoing reflected wavefield. In the case of primary reflections in surface acquisition, the downgoing wavefield is the source wavefield and the upgoing wavefield is the reflected wavefield recorded by the receivers. This leads to a three-step procedure:

- propagation of an estimation of the source wavelet to determine the source wavefield $\mathrm{S}(s, \mathbf{x}, t)$ in the medium at all subsurface points and for all times;
- backpropagation of the data recorded at receiver positions at all subsurface points and for all times to determine the receiver wavefield $\mathrm{R}(s, \mathbf{x}, t)$;
- application of an imaging condition to these wavefields S and R to determine an image of the subsurface. Many formulations exist for this imaging condition, the most commonly used is the cross-correlation at zero-lag of the source and receiver wavefields.

The procedure is repeated for each shot point, yielding as many migrated sections as shot points in the survey. The obtained images can be stacked to increase the signal to noise ratio. However, the similarities and discrepancies between these different images can also be used as information. This is the basic principle of MVA techniques presented in Section 1.3.

This migration algorithm has been recognised as the first iteration of the least-squares data-fitting inverse problem (Lailly, 1983; Tarantola, 1984). Under the Born approximation, this inverse problem is linearised. Assuming a fixed estimate of the background velocity model c_0, it consists of determining the model perturbation δc which best reproduces recorded data. The associated scalar misfit function is

$$\mathrm{J}_0[c_0, \delta c] = \frac{1}{2}\left\|\mathrm{P}[\delta c] - \mathrm{P}^{\mathrm{obs}}\right\|^2, \tag{1.4}$$

where $\mathrm{P}^{\mathrm{obs}}$ stands for recorded data and $\mathrm{P}[\delta c]$ represents the Born-modelling of data with the model perturbation δc. As in the case of FWI (equation 1.2), this objective function measures the misfit between observed data and calculated data, except that here only the short-scale part of the velocity model is updated, and the associated inverse problem is linear.

The migration procedure detailed above actually computes the gradient of the objective function (1.4) in $\delta c = 0$,

$$\frac{\partial \mathrm{J}_0}{\partial \delta c}\big[c_0, \delta c = 0\big](\mathbf{x}) = \frac{2}{c_0^3(\mathbf{x})}\int_s \int_t \frac{\partial^2}{\partial t^2}\mathrm{S}[c_0](s, \mathbf{x}, t)\mathrm{R}[c_0](s, \mathbf{x}, t)\,\mathrm{d}t\,\mathrm{d}s, \tag{1.5}$$

where S and R are the source and receiver wavefield, whose values depend on the background velocity $c_0(\mathbf{x})$. Note that the definition of the misfit function (1.4) contains an implicit summation over the sources. Defining the reflectivity as the gradient of this function produces a *post-stack* image.

The migration operator is actually the adjoint of the Born modelling operator and provides only a qualitative image of the subsurface: the position of the reflectors is kinematically consistent with the assumed velocity model and the recorded traveltimes. However the method does not provide a correct estimation of their amplitude. To obtain a *quantitative* estimate of the reflectivity model, the minimisation problem presented above should be solved completely. This *true-amplitude migration* can be achieved with an approximate inverse of the Hessian of the objective function (Beylkin, 1985; Lambaré *et al.*, 1992; Plessix and Mulder, 2004; Kiyashchenko *et al.*, 2007). Alternatively an iterative procedure using gradient-based methods can be defined to minimise the misfit function (Nemeth *et al.*, 1999; Østmo *et al.*, 2002). In chapter 2, an iterative migration procedure will be presented in the context of second-order Born modelling.

The result of the migration process depends strongly on the velocity model used to compute the source and receiver wavefields. With an incorrect velocity model, events may be positioned at the wrong depth and may not be well focused. In the following section, time-domain methods to construct the background velocity model are reviewed.

1.2.3. Tomographic methods: retrieving the large-scale structure of the velocity model

Traveltime tomography

Traveltime tomography techniques are developed with the ray theory, based on the high-frequency asymptotic approximation of the wave equation (Červený, 2005). Wave propagation in the subsurface is described by rays with propagation laws similar to those used in optics. In this formalism, the Green's functions describing the wave propagation can be decomposed into three terms, one accounting for the traveltime, one for the amplitude and one for the source signature. In tomographic methods, events in the data are characterised by their traveltime only. Their amplitude and the finite-frequency nature of the data are not considered, contrary to FWI. The objective function of the associated inverse problem measures the difference between traveltimes picked on a selection of events in the data with traveltimes computed with an estimation of the velocity model c:

$$\mathrm{J}_{\mathrm{TT}}[c_0] = \frac{1}{2}\|\tau[c_0] - \tau^{\mathrm{obs}}\|^2. \tag{1.6}$$

The velocity model is updated iteratively to minimise the traveltime differences. The method requires a first processing stage to pick events, as well as a modelling engine relying on the resolution of the Eikonal equation to compute traveltime maps corresponding to a velocity model (Vidale, 1988; Podvin and Lecomte, 1991; Noble *et al.*, 2014). Depending on the acquisition, different events can be picked.

The first arrivals are mostly used in seismology. They consists of direct and diving waves and are easier to pick than reflections. In active seismic experiments, they are well suited when the acquisition geometry emphasises transmission effects, for example in surface acquisition with large offsets, cross-well acquisition or with buried receivers and sources at the surface (Vi Nhu Ba, 2014). In the case of surface acquisition without large offsets, the depth penetration of these waves restricts the area where the background velocity can be recovered. Information about the deeper part of the model is contained in reflected waves.

Reflected events can be picked as well. In this case the model is parametrised both by a velocity model and a geometric description of the reflectors. Reflection tomography was historically developed in the time domain (Bishop *et al.*, 1985; Farra and Madariaga, 1988). Each reflection in the data is associated to a reflector. An issue is that picking reflected events along a wide range of offset is a difficult task, especially with complex geology and noisy data (Lailly and Sinoquet, 1996). To improve the signal to noise ratio, the data can be stacked if the model is not too complex. Alternatively the analysis can be transposed to the migrated domain (Stork, 1992), where interfaces are easier to pick. Reflectors are picked on migrated sections, then the corresponding migrated events are modelled (or demigrated) to be compared to the observed traveltimes (Jacobs *et al.*, 1992; Grau and Lailly, 1993). Note that with short-offsets only, there is an ambiguity between the velocity and the depth of reflectors (Williamson,

1990). Reflection tomography is more generally known to be an ill-posed inverse problem requiring external constraints. These constraints can be existing well data (Le Stunff and Grenier, 1998), or information about the subsurface provided by other imaging methods, gravimetric or electromagnetic for example (Lines *et al.*, 1988).

Other alternatives to the picking of continuous reflected events are slope tomography methods, and in particular stereotomography (Billette and Lambaré, 1998; Lambaré, 2008; Guillaume *et al.*, 2013). In addition to traveltimes, the slope of locally coherent events are picked on common shot and common receiver panels. The inverted velocity model should then explain both traveltimes and slopes of picked events. In this method there is no need to describe the model with a set of continuous layer interfaces, only the dip of local events is needed. Then the density of picked events is higher. As for traveltime tomography, local picking can be performed in the image domain (Chauris *et al.*, 2002). Then kinematic invariants such as traveltimes and slopes can be recomputed in the time domain.

Traveltime tomography has become a standard in the oil and gas industry for velocity model building since the late nineties (Woodward *et al.*, 2008). Although recent publications have focused mainly about FWI, progresses have been made to better constrain the inversion, include structural information like sharp velocity contrasts and make the method more and more automatic (Lambaré *et al.*, 2014). Improvement of both the method and the recording devices have increased the resolution of velocity model constructed by traveltime tomography (Nichols, 2012). We now review extension of traveltime tomography that go beyond the high-frequency approximation.

Wave-Equation Traveltime Tomography

Wave-Equation Traveltime Tomography (WETT) is an extension of traveltime tomography which takes into account the finite-frequency nature of seismic data. Luo and Schuster (1991) proposed to cross-correlate the traces of first-arrival events in observed and calculated data instead of subtracting their traveltime. Then they minimise the time-lag maximising the cross-correlation. van Leeuwen and Mulder (2010b) indicate that the traveltime error might not coincide with the time-lag maximising the cross-correlation because of errors in the source wavelet used for calculated data. They propose to increase the robustness of the approach with an objective function penalising non-zero time-lag. Compared to FWI, WETT yields velocity models with lower resolution but is less prone to cycle-skipping and can be used to build a starting model for FWI (Wang *et al.*, 2014). The resolution of the method can be increased by replacing correlation by deconvolution (Luo and Sava, 2011). A related strategy is Adaptive Waveform Inversion (AWI) (Warner and Guasch, 2014, 2015) in which the coefficients of a Wiener filter applied to calculated data are determined at each iteration to match observed data. The velocity model is then updated so that the coefficients of the filter amount to a simple zero time lag, which corresponds to the case where calculated data perfectly reproduce observed data.

Reflection Waveform Inversion

Reflection Waveform Inversion (RWI) (Xu *et al.*, 2012; Zhou *et al.*, 2012; Brossier *et al.*, 2015) is an approach similar to FWI developed to extract information about the macromodel from reflected events. It is inspired by the Migration-Based Traveltime Tomography (MBTT) procedure

(Chavent and Jacewitz, 1995; Plessix *et al.*, 1995). It either assumes a scale separation between low and high frequency content of the velocity model (Xu *et al.*, 2012) or relies on a parametrisation with both the velocity and the impedance to naturally facilitate the scale separation (Zhou, 2016). An initial estimate of the macromodel is used to construct a migrated section. Then reflected events are modelled from the migrated section and the background velocity model, and compared to observed reflections in an objective function similar to FWI. With this strategy, transmission wavepaths are constructed between the reflectors and the surface as well as between the reflectors and the receivers, allowing to recover the long-scale part of the velocity model in areas of the subsurface not reached by diving waves. Zhou *et al.* (2015) and Alkhalifah and Wu (2016b) proposed strategies to combine the information extracted from both FWI and RWI about the macromodel.

An alternative strategy using the scale separation is the Differential Waveform Inversion (DWI) approach (Chauris and Plessix, 2013). A migrated section is computed from one shot gather and used to calculate synthetic data for the next shot. The difference with the corresponding observed data is used as information about the errors contained in the macromodel.

Finally van Leeuwen and Mulder (2008a) propose to correlate traces of observed and simulated data separated by a spatial shift and with a temporal lag in order to build coherency panels function of the space and time shifts. The correct velocity model corresponds to focusing of energy at zero spatial and temporal delay. The criteria assessing the quality of the velocity model in this approach as well as in DWI are inspired by image-domain methods presented in the next section.

1.3. Migration Velocity Analysis

We present in this section image-domain methods for the resolution of the inverse problem. This is the class of methods which will be studied in the thesis.

The basic principle of image-domain methods is that seismic data are redundant. Using an initial velocity model, a collection of migrated images can be created with different subsets of the data, for example one image for each shot gather (Al-Yahya, 1989). Migrating with the correct velocity model should result in kinematically coherent images, meaning that reflectors in images produced with different shot-point experiments should be positioned at the same depth (Al-Yahya, 1989). Conversely, discrepancies between different images of the same Earth carry information about the errors in the estimated velocity model. The family of methods using this principle is called Migration Velocity Analysis (MVA).

Historically MVA emerged as an extension of the usual Normal Move Out (NMO) correction procedure. In this method the data are sorted into panels called Common Mid Point Gather, function of the surface offset and the time. With the assumption of horizontal structure of the Earth, each trace represents the reflection produced by the same subsurface point with different source-receiver spacings. Reflections then appear as hyperbolas with a traveltime increasing with offset. The NMO correction consists of finding the (1D-)velocity model that best explains these hyperbolas. The better the velocity model, the better traces stack constructively after correction. Although dipping events can be treated with the Dip Move Out correction, this method is limited to the case of simple velocity model structures. It has been transposed to the

image domain by Al-Yahya (1989) to handle more complex media.

Many different MVA approaches have been proposed. They all analyse primary reflections and rely on the separation of scale and the Born approximation. We will review them using the formalism introduced by Symes (2008). MVA methods can be described in four steps:

- migration of the data to an image-domain parameterised with an additional parameter representing the redundancy of seismic data. For example in 2D, the migrated volume $\delta c(x,z,h)$ is parameterised by lateral position x, depth z, and the redundant parameter h. Then the subsurface image and the data have the same dimension. The latter indeed depends on the source horizontal coordinate, the receiver horizontal coordinate and time. This is a key point of MVA strategies: even in an incorrect background velocity model, no information is lost during migration of the data, in the sense that data can be re-modelled in the same background velocity model using the migrated volume;

- definition of a coherency or focusing criterion. It depends on the choice of the redundant parameter and states that the value of the reflectivity along the additional parameter carries the information about the velocity model. Note that stretching effects and limited acquisition geometries are not necessarily taken into account in this principle. The analysis is performed on panels called Common Image Gathers (CIGs) representing a section of the reflectivity volume as a function of depth and of the extra-parameter for a fixed lateral position;

- definition of an objective function. Contrary to the data domain where observed data provide a natural reference for assessing the quality of an estimated velocity model, there is no obvious reference in the image domain. Instead, the objective function evaluates if the focusing criterion is satisfied. It is defined by the general expression

$$\mathrm{J}_{\mathrm{MVA}}[c_0] = \frac{1}{2}\left\| \mathrm{A}\,\delta c[c_0](x,z,h)\right\|^2, \tag{1.7}$$

  where operator A is called annihilator, and is chosen such that $\mathrm{J}_{\mathrm{MVA}}$ is extremal for the correct background velocity model. To allow minimisation with gradient-based methods, the objective function should be smooth and free of local minima over a large range of velocities. The partial derivative of the function with respect to the image, $\mathrm{A}^{\mathrm{T}}\mathrm{A}\delta c$, is commonly called *image residual*.

- computation of the gradient of $\mathrm{J}_{\mathrm{MVA}}$ with respect to c_0. This is usually done with the adjoint-state method (Plessix, 2006) for efficiency reasons. The background velocity model is then updated using non-linear local optimisation methods (Nocedal and Wright, 2006).

In the following, we present two formulations of MVA: in the *surface-oriented* approach, the redundant parameter is related to the acquisition, contrary to the *depth-oriented* approach where it is introduced during the construction of the migrated image. Then we will focus on the depth-oriented formulation. We define the objective function in the image-domain and detail the resolution of the associated inverse problem. The general difficulties and limitations of the method, appearing even in the case of primary reflections only, are presented. Finally we illustrate the principle of MVA on a simple 1D case and show that the presence of multiples

interpreted as primary leads to an incorrect background velocity update. This motivates the need to deal with multiples in the MVA framework.

1.3.1. Surface-oriented MVA

A natural choice for the redundant parameter is the shot number or shot position (Symes and Kern, 1994; Huang and Symes, 2015). The idea is that images obtained from several experiments should be identical and independent of the position of the source. With this formulation, CIGs represent the image of the subsurface obtained with different subsets of the data. On these CIGs, events should be horizontal for the correct velocity model. Several authors suggest using the surface offset (distance between source and receiver) instead of the shot number (Zhang *et al.*, 2010) as common offset sections provide large illumination, although it is less natural for wave-equation based modelling. An example of an image volume indexed by the surface offset is represented in figure 1.6.

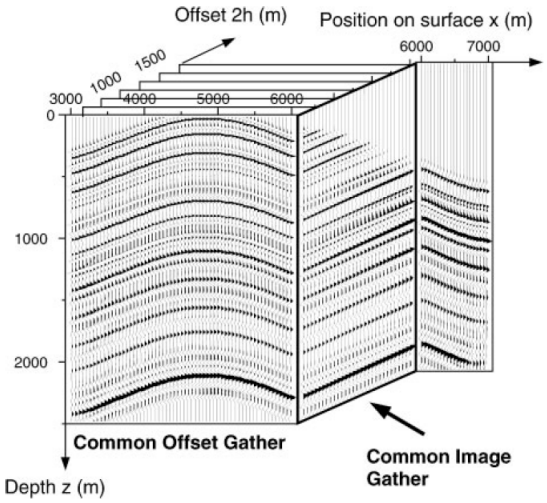


Figure 1.6. – Example of CIG with the surface offset as redundant parameter (taken from Chauris *et al.*, 2002).

If the velocity model is incorrect, events in CIGs are not horizontal any more. In simple models, they curve upward for a too slow velocity model and downward for a too low velocity model. Then two strategies can be defined. In the case of the *Semblance* criterion (Chavent and Jacewitz, 1995), the similarity between images produced with different subsets of the data is measured by stacking over the extension parameter and the objective function measures the energy of the stack. For the correct velocity model, the images sums up constructively and the energy of the stacked image is maximal. The objective function is free of local minima over a large range of velocity but may exhibit oscillations away from the correct model (Chauris and Noble, 2001).

With the Differential Semblance Optimisation (DSO) (Symes and Carazzone, 1991) strategy, the objective function computes the derivative of the image with respect to the redundant

parameter to measure the flatness of events in CIGs (Chauris and Noble, 2001). It has been shown, at least for simple 1D models, that the DSO functional has better convexity properties than the one defined by the semblance principle (Stolk and Symes, 2003; van Leeuwen and Mulder, 2010a). Note that the DSO functional was initially defined as a regularisation term for FWI (Symes and Kern, 1994).

The estimation of the derivative is sensible to coherent noise and thus requires filtering before evaluation of the objective function (Chauris and Noble, 2001). The recognition of continuous events across the offset range might be tedious as well. Then Chauris *et al.* (2002) proposed to pick the slope of locally coherent events in CIGs and common offset sections to assess the quality of the velocity model. They show that this strategy is an equivalent in the image domain of the stereotomography procedure described before.

However as each individual image is obtained with only a subset of the data, kinematic artefacts may appear when complex wavepaths are considered (Xu *et al.*, 2001; Stolk and Symes, 2004; Zhang *et al.*, 2010). In the depth-offset extended model presented in the next paragraph, each component of the migrated volume is constructed from the whole data set, the image is thus better constrained and is less prone to the apparition of artefacts (Stolk and de Hoop, 2005).

1.3.2. Depth-oriented MVA

In the depth formulation of MVA, the redundant parameter is independent of the acquisition and is introduced during the construction of the migrated image. The image domain is said to be *extended* and the *extension parameter* can be a spatial (Rickett and Sava, 2002; Shen *et al.*, 2003) or temporal (Sava and Fomel, 2006; Yang and Sava, 2011) delay. Alternatively a scattering angle can be considered (Sava and Fomel, 2003; Biondi and Symes, 2004). The spatial delay (figure 1.7) is commonly referred to as *depth-offset* or *subsurface-offset*. In 2D, the general extended cross-correlation formula with the subsurface offset $\mathbf{h} = (h_x, h_z)$ and the time-lag Δt (Sava and Vasconcelos, 2011) is

$$\delta c[c_0](\mathbf{x}, \mathbf{h}, \Delta t) = \frac{2}{c_0^3(\mathbf{x})} \int_s \int_t \frac{\partial^2}{\partial t^2} \mathrm{S}[c_0](s, \mathbf{x}-\mathbf{h}, t-\Delta t)\mathrm{R}[c_0](s, \mathbf{x}+\mathbf{h}, t+\Delta t)\,\mathrm{d}t\,\mathrm{d}s. \quad (1.8)$$

The section of this image at $\mathbf{h} = 0$ and $\Delta t = 0$ corresponds to the image of the physical reflectivity obtained with the classical migration formula (1.5) after summation over all sources and receivers. The introduction of the delays compensates for errors contained in the velocity model and allows to capture information that is not present in the usual migrated section. Energy focusing at non-zero values of the extension parameter carries information about the background velocity model. In the depth-extended domain, the coherency criterion used to assess the quality of the velocity model is the focusing of energy at zero-space lag and zero time-shift.

In practice, after computation of the source and receiver wavefields, a correlation is performed for each value of the subsurface-offset and the time-lag. To keep the migrated image size and the computational time within reasonable limits, the extension is usually made along a single extension parameter only. The subsurface-offset is usually chosen (Shen and Symes, 2008). In the case of surface acquisition and horizontal structures, most of the information about the velocity model is contained along the horizontal component of the depth-offset. Therefore in the

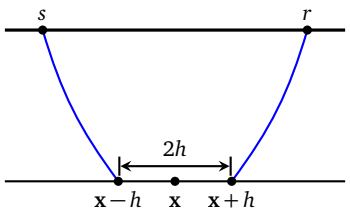


Figure 1.7. – Definition of the subsurface-offset spatial delay.

following, the subsurface-offset will be considered horizontal only, $\mathbf{h} = (h, 0)$, as represented in figure 1.7.

CIGs then represent the migrated section at fixed lateral position and depend on the depth z and on the subsurface-offset h. Energy focuses around zero-offset for the correct background velocity model and spreads over non-zero offset in an inaccurate model. Events have a downward (upward) curvature for a too low (too high) velocity model (Mulder, 2014) (see figure 1.9a for an example). Note that this is the opposite for the surface-offset case.

The Semblance and DSO strategies used in surface-oriented MVA have equivalences in the depth-oriented approach. The Semblance principle consists of maximising the energy around zero-offset. Note that maximising the energy of the image at $h = 0$ corresponds to the surface-oriented Semblance criterion. In the depth-oriented DSO formulation, the objective function should penalise defocused energy. This is the strategy used in this study.

1.3.3. Choice of a misfit function and inversion of the velocity model

Usually, the objective function of DSO is constructed by multiplying the image by the value of the subsurface offset (Shen and Symes, 2008; Mulder, 2008):

$$\mathrm{J}_{\mathrm{DSO}}[c_0] = \frac{1}{2}\left\|h\delta c[c_0](\mathbf{x}, h)\right\|^2. \tag{1.9}$$

This is the definition used in this study. Alternatively, a mixed formulation involving both the semblance and the differential semblance can also be formulated (Shen and Symes, 2008; Mulder, 2014)

$$\mathrm{J}[c_0] = \frac{1}{2}\left\|h\,\delta c[c_0](\mathbf{x}, h)\right\|^2 - \frac{\alpha}{2}\left\|\delta(h)\delta c[c_0](\mathbf{x}, h)\right\|^2. \tag{1.10}$$

with a positive scalar coefficient α to be determined. This formulation allows to use the total information contained in the migrated image, as the energy at zero-offset discarded by the DSO term is used by the semblance term. In this formulation, away from the minimum, the optimisation is driven by the DSO term while the robustness of the semblance close to the solution is used in the last iterations.

The MVA objective function is non-linear and minimisation is performed with local optimisation schemes. The calculation of the gradient is the main computational effort of the method. It is therefore performed with the adjoint-state method (Lameloise *et al.*, 2014; Yang and Sava, 2015). As will be explained in the following of this section, this gradient is not always smooth although we want to update the background part of the velocity model. Therefore a smoothing

is commonly applied to the gradient, at least in the first iterations (Lameloise and Chauris, 2016). In theory the inversion should account for the Hessian of the MVA objective function, but its computation is not affordable in practice. So far only an estimation of its diagonal computed as the product of the Hessian matrix with a unit vector has been used (Liu *et al.*, 2014b; Shen and Symes, 2015).

1.3.4. Limitations of MVA

Despite an attractive formulation, not so many applications on real data have been published (Chauris and Noble, 2001; Mulder and ten Kroode, 2002; Alkhalifah, 2005; Shen and Symes, 2008; Mulder, 2014; Weibull *et al.*, 2012a,b; Lameloise, 2015). We review some of the difficulties of MVA which prevent its use in realistic industrial contexts.

Computational cost

To limit the computational cost and memory requirements of the method, the extension of the model is made only along the horizontal component of the subsurface-offset, and its vertical component as well as the time-lag parameter are not considered. Even in this simplified setting, MVA remains computationally expensive. This is the main limitation for the extension of MVA techniques to 3D-applications, where two extra dimensions are in principle added to the model space to match the data size.

A migration step should be performed at each iteration on the background velocity model, involving the calculation of the source and receiver wavefields and their cross-correlation. This cross-correlation should be performed for each value of the extension parameter. To reduce the computation time necessary for cross-correlation, Yang and Sava (2015) propose to compute CIGs only at a selection of image points instead of the whole image. van Leeuwen *et al.* (2015) propose to construct CIGs with only a random choice of traces.

A related issue is the maximum value of the subsurface-offset $h_{\max}$ that should be considered. In practice it is chosen empirically, depending on the maximum value of the surface offset and the depth of the reflectors. If a too low value is chosen, some information may be lost during migration. If it is too high, the cross-correlations for the large values of h increase the computational cost without adding any information. As the background velocity model improves with iterations, energy should be more and more focused in CIGs and the value of $h_{\max}$ could decrease. Using this idea, Fu and Symes (2015) propose a multiscale/multigrid strategy starting with low frequency data and a coarse sampling of the subsurface-offset axis. At each iteration the data misfit is computed in two cases, first considering all the subsurface-offsets until $h_{\max}$, second using only half of the subsurface-offsets range. This criteria is used to update the value of $h_{\max}$ at each iteration. In parallel, the sampling of the subsurface offset-range is increased as higher frequency are included in the data.

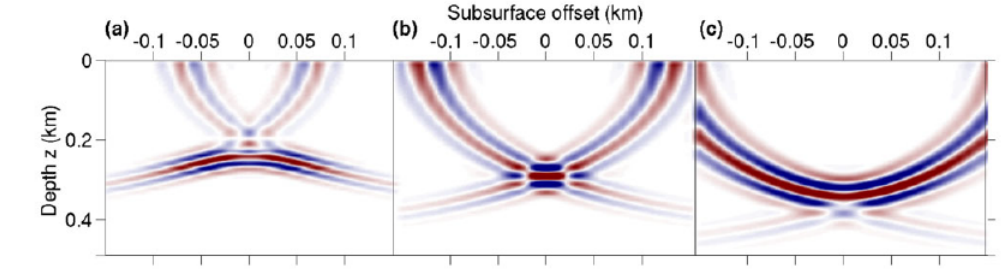
Note that if this techniques may decrease the computational cost of cross-correlation, the cost due to propagation remains expensive, especially on fine grids.

Although computational efficiency is an important aspect of the method, this issue will not be specifically addressed in the study, CIGs being calculated with a fixed range and sampling of the extension axis.

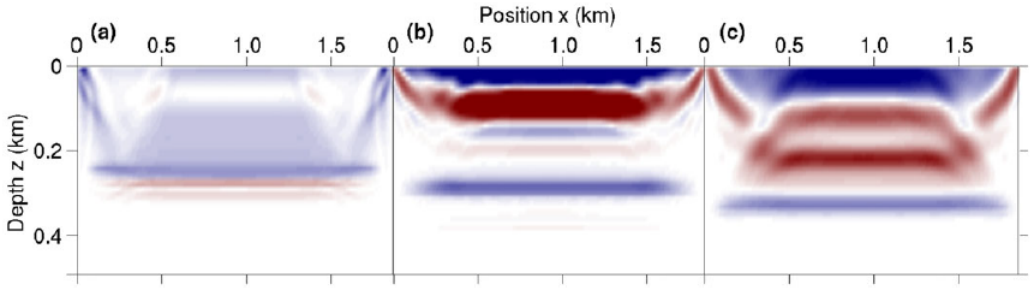
Accuracy of migrated images

The reflectivity model used in MVA techniques is the result of a classic migration algorithm. As already mentioned, this image is the first gradient of the objective function minimising the misfit between observed and calculated data and is produced using the adjoint of the Born modelling operator. Several authors suggest that this is not a sufficiently accurate solution of the minimisation problem. Yang *et al.* (2013) report that under complex structures like salt dome, uneven illumination results in defocused energy even for the correct velocity model. As a remedy, they propose to incorporate illumination as a weight in the MVA objective function, such that defocusing due to poor illumination does not influence the velocity update.

Even for very simple models, CIGs are perturbed with migration artefacts (figure 1.8a) related to the limited source and receiver coverage of the acquisition setting (Mulder, 2014). These



(a) Common Image Gathers.



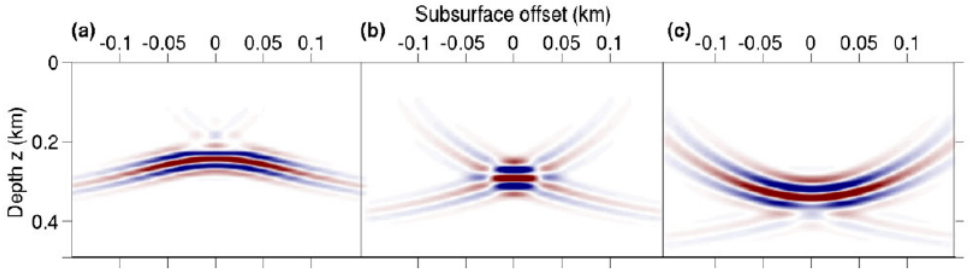
(b) Gradient of the MVA objective function.

Figure 1.8. – (a) CIGs and (b) associated gradients of the MVA objective function computed with classical migration for a too low (left), correct (middle), and too high (right) initial velocity model with a single horizontal reflector. Blue, white and red colours correspond to negative, null and positive values, respectively (from Lameloise *et al.*, 2014). Homogeneous update above the reflector would be expected in a tomographic approach.

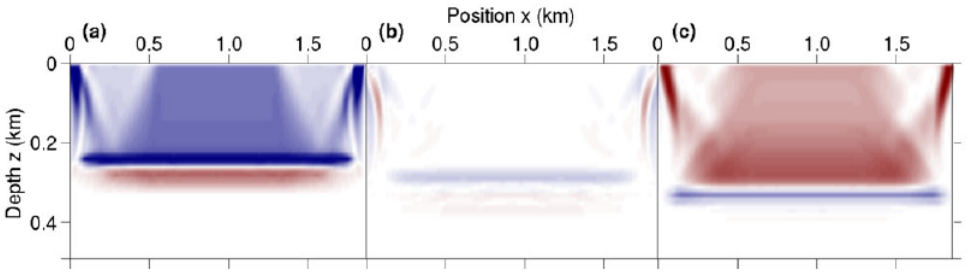
artefacts always have an upward curvature, independently of the velocity used during migration and do not focus for the correct velocity. As a result, the MVA objective function is not minimal for the correct velocity model but for a model with slightly higher values (Lameloise *et al.*, 2014) and its gradient does not provide a satisfactory update (figure 1.8b).

As these events are steeper than the ones associated to the true reflector, Weibull and Arntsen (2011) and Chauris *et al.* (2015) propose to apply a z-derivative to the CIGs to attenuate migration

artefacts. Alternatively, Mulder (2014) suggests the application of tapers to smooth the truncation in source and receiver coverage. A more sophisticated approach consists of replacing the classic migration by a quantitative migration (Lameloise and Chauris, 2014; Hou and Symes, 2015). Lameloise *et al.* (2014) extend the ray+Born inversion approach of Lambaré *et al.* (1992) to the extended model and define migration weights such that the Hessian of the migration objective function is almost diagonal. In practice the weights allow to compensate for uneven illumination in the CIGs. Migration artefacts are correctly attenuated (figure 1.9a) and the gradient of the objective function improved (figure 1.9b). However, this approach is limited to the use of ray



(a) Common Image Gathers.



(b) Gradient of the MVA objective function.

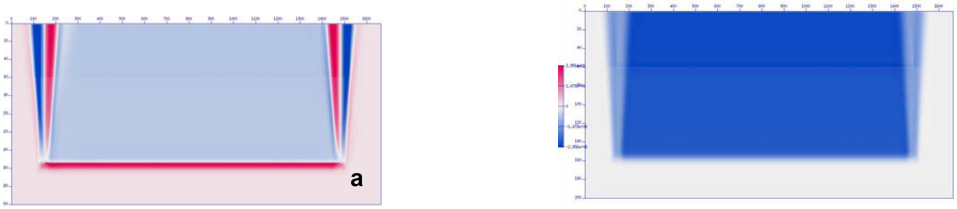
Figure 1.9. – Same as figure 1.8 with a quantitative (ray-based) migration instead of a classical (adjoint) migration (from Lameloise *et al.*, 2014).

theory in the computation of the CIGs. In this thesis, it will be extended to wave-equation based operators (Hou and Symes, 2015, 2017). We will show that this strategy involves the application of a z-derivative to the migrated image, similar to what was proposed by Weibull and Arntsen (2011) and Chauris *et al.* (2015), thus giving a formal justification to their strategy.

Gradient artefacts

A well-known issue pointed out by Fei and Williamson (2010), Vyas and Tang (2010) and Chauris and Lameloise (2014) is the presence of strong oscillations in the gradient of the MVA objective function with respect to the background velocity (figure 1.10a). They are known as "gradient artefacts". They occur at discontinuities in the reflectivity model and along the reflectors and prevent the MVA gradient to be used as a velocity update without a prior smoothing stage.

Fei and Williamson (2010) propose to add a phase shift of 90° to the image residual with the application of a h-derivative before the computation of the gradient of the MVA objective



(a) Without horizontal contraction. **(b)** With horizontal contraction.

Figure 1.10. – (a) Illustration of the "gradient artefacts" on a homogeneous model with a single horizontal reflector. Oscillations appear at the truncation of the reflectivity. (b) The "horizontal contraction" approach allows to remove these spurious oscillations (from Fei and Williamson, 2010).

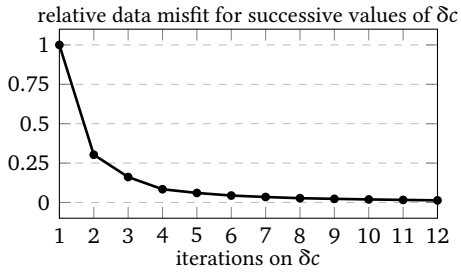
function. This trick greatly improves the quality of the velocity update (figure 1.10b) but is not fully understood. Shen and Symes (2015) recognise in this modification a warping technique: the derivative with respect to h applied to the image residual can be seen as a contraction of CIGs in the h direction, this contracted CIGs having their energy more focused at zero subsurface-offset. Therefore Shen and Symes (2015) rename this technique "horizontal contraction". A disadvantage of this method is that the modified velocity update is not the gradient of an objective function any more. Moreover such a technique does not properly work in the case of low velocity anomalies (Shen and Symes, 2015).

Stability of the MVA gradient

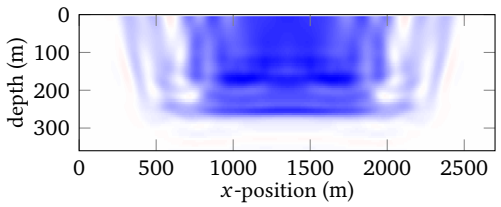
Another numerical difficulty in the computation of the MVA gradient has been recently pointed out (Huang, 2016). To obtain an accurate reflectivity model free of migration artefacts, one could treat migration as an inverse problem and carry on the minimisation of the associated objective function (1.4) iteratively. The resulting reflectivity model is used as input of the MVA objective function (1.9). This nested optimisation strategy (see also figure 1.22) will be extensively analysed in this thesis.

As recently observed by Huang (2016), an issue with this approach is the apparent instability of the gradient of the MVA objective obtained after each inner iteration on the reflectivity: "the theoretical relation between the error in solving the inner problem and the error in the gradient computation has not been established. The reason is that the small ℓ_2 error in the data misfit does not imply a small error in the WEMVA operator." (Huang, 2016, p. 100).

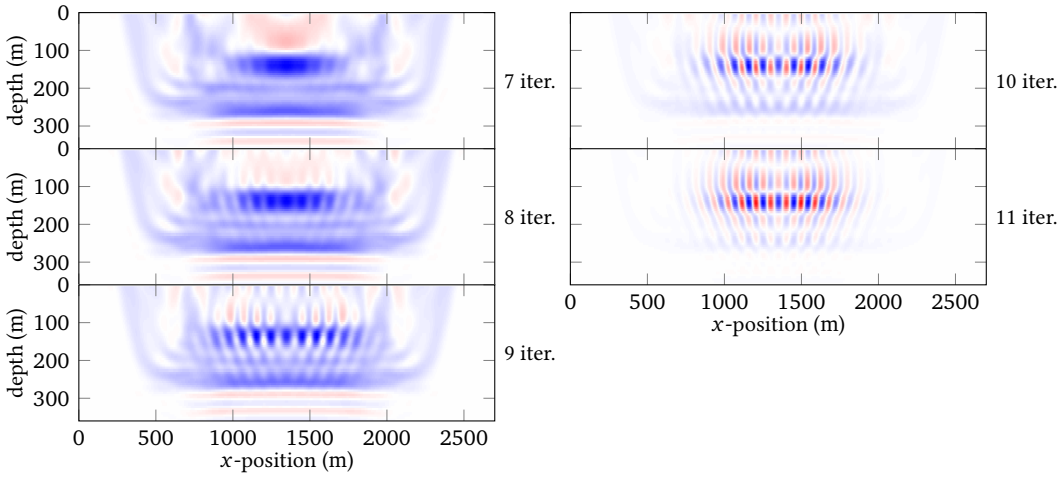
We illustrate this issue with a numerical result (figure 1.11) extracted from chapter 3. The model considered here is similar to those shown in figures 1.8 to 1.10. A single horizontal reflector is located in a homogeneous background velocity model and the MVA gradient is computed in a too low homogeneous medium. Thus we expect a homogeneous negative gradient above the reflector position (figure 1.11b). A good solution of the inner problem is found after 5 iterations. The subsequent reflectivity models provide very similar data misfit (figure 1.11a). However using these successive reflectivity models as input to the MVA objective function leads to very different MVA gradients (figure 1.11c), whose similarity to the expected velocity update decreases exponentially with the number of iterations (figure 1.11d). This issue will be further illustrated in chapters 3 and 4 and possible solutions will be proposed in chapters 3 and 5.



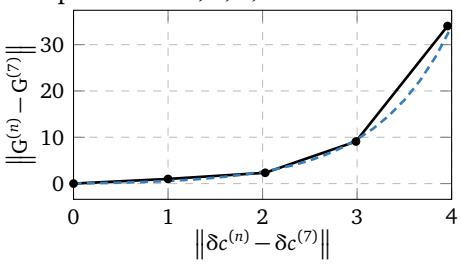
(a) Objective function representing data misfit for the successive value of δc across iterations.



(b) An example of expected gradient of the MVA objective function. Blue, white and red colours correspond to negative, null and positive values, respectively.



(c) Actual MVA gradient computed after 7, 8, 9, 10 and 11 iterations on the reflectivity δc.



(d) Distance between the gradient obtained at iterations 7, 8, 9, 10 and 11 and the gradient obtained at iteration 7, function of the same distance between the reflectivity models. The dahsed blue line represents the curve $y = 0.23(e^{1.25x} - 1)$.

Figure 1.11. – Iterative resolution of the migration inverse problem and computation of the associated MVA gradient, showing the sensitivity of the gradient to minor changes in the reflectivity model.

Sensitivity to noise

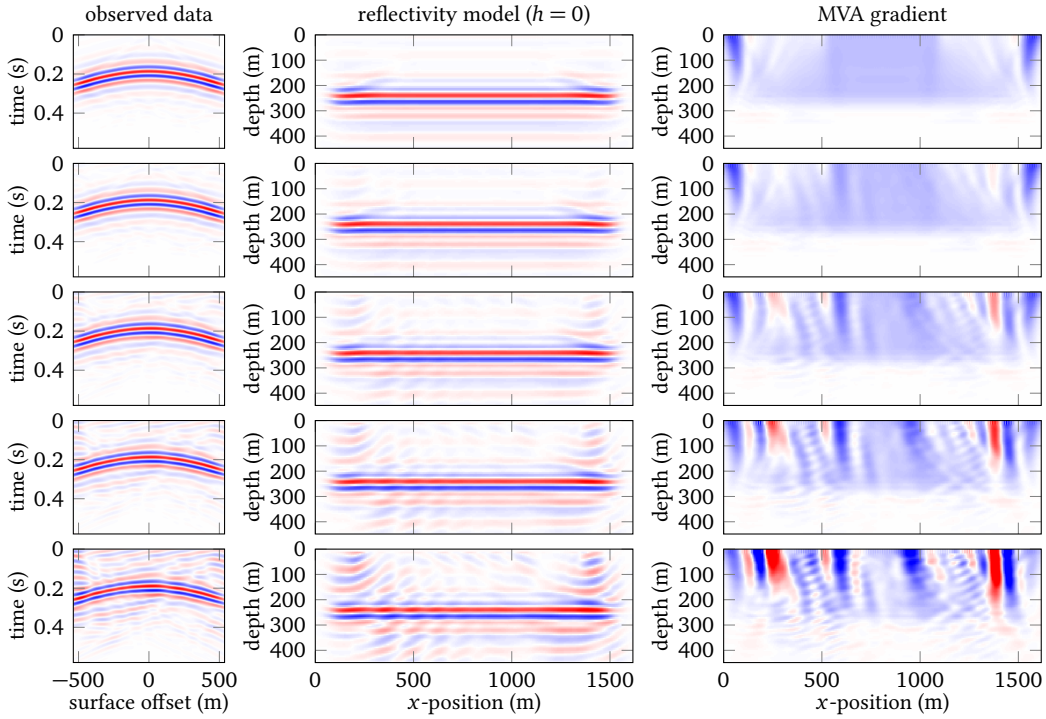
We now illustrate the sensitivity of the MVA gradient to noise in observed data (figure 1.12). We use an example similar to the one of the preceding paragraph. We first compute the reflectivity model and the associated MVA gradient with noise-free observed data. The method used to obtained these images will be detailed in chapter 4. The MVA gradient is homogeneous and negative above the reflector. Increasing the noise on observed data leads to an increased noise level in the reflectivity image, but the reflector is still clearly distinguishable. The migration operator is linear in the noise nevel, and so is the impact on the reflectivity image (figure 1.12b). On the contrary, the background velocity update is more altered by noise in observed data: artefacts with an incorrect sign appear and the gradient is not homogeneous any more. It differs from the reference model with a polynomial behaviour (here, a power 1.13 as illustrated in figure 1.12b) when the noise level increases, so this is not as extreme as the instability shown in figure 1.11.

Beyond primary reflected waves

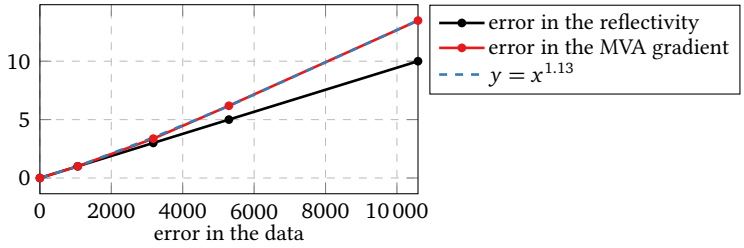
All the theoretical aspects of the method are developed for primary reflections. However, the background velocity updates could benefit from the inclusion of other events with different illumination in the MVA procedure.

Recently, the extension to transmitted events has been investigated (Shen and Symes, 2013; Chauris *et al.*, 2013; Lameloise *et al.*, 2015; Lameloise and Chauris, 2016). Lameloise and Chauris (2016) define *image functions* for transmitted events, equivalent to the extended migrated sections for reflected waves. They are constructed with the same extended cross-correlation imaging condition, except that the full recorded wavefield including reflected and transmitted waves is backpropagated. The corresponding CIGs are sensitive to the velocity and exhibit the same focusing behaviour: energy is focused at zero subsurface-offset for the correct velocity model and spreads over non-zero offsets otherwise. The conventional MVA objective function penalising defocused energy can be used and the corresponding velocity updates are smooth and do not suffer from the "gradient artefacts" observed for reflected waves.

Multiple reflections are still an issue for MVA. If they are not removed from the data, they mislead the velocity analysis as the methodology developed for primary reflection does not explain the kinematics of multiple reflections. Compared to the primary it may be mistaken for, a multiple travels in the upper part of the model. With the hypothesis of increasing velocity with depth, multiples then tend to favour lower apparent velocities. Hence they need to be removed from the data in a preprocessing stage or in the migrated sections with a muting in the $f - k$ domain (Mulder and ten Kroode, 2002; Li and Symes, 2007), for example. However the removal of multiple reflection is challenging and residual multiple energy may survive these filtering processes. As a remedy, Mulder and van Leeuwen (2008) introduce a bias in the MVA objective function to favour higher velocities, thus compensating for multiples. This formulation relies on the assumption that multiples favour lower velocity, which does not hold for models where the velocity does not increase with depth.



(a) Observed data (left), reflectivity section at $h = 0$ (middle) and MVA gradient (right) recovered with an increasing level of noise (from top to bottom). Blue, white and red colours correspond to negative, null and positive values, respectively.



(b) Error in the recovered reflectivity model and in the associated MVA gradient, function of the level of noise introduced in observed data.

Figure 1.12. – Influence of noise in observed data in the reflectivity model recovered by migration and on the associated MVA gradient.

1.3.5. Illustration of the principle and difficulties of MVA on a simple example

To conclude this part, we illustrate in this paragraph the basic principle of MVA and some issues due to the presence of multiples. For the sake of simplicity, we consider a pure 1D example (figure 1.13). Although it may not be representative for the complexity of a 2D acquisition, it shares many properties with the 2D and 3D cases that will be considered in the following chapters. This 1D example will be used again in chapters 4 and 5.

The input is reflected observed data (figure 1.13b) pre-computed in an exact homogeneous velocity model with a single reflector located at 400 m depth. MVA techniques aim at recovering the background velocity model. Here we start with an initial too low homogeneous background velocity, such that we expect the method to provide a positive update. The method is decomposed in two steps. First observed data are converted into a reflectivity image function of the subsurface by *migration*. This image (e) is obtained as the zero-lag cross-correlation of two wavefields (c) and (d) computed by propagating the source wavelet (a) emitted at $z = 0$ and by backpropagating observed data (b) recorded at $z = 0$. As the initial velocity model is too low, the reflector is located above its true position.

The so-called "image residual" (g) (which can be seen as an equivalent of the more natural data residual in data-domain methods) is obtained by multiplying the reflectivity image by $|z - z_{ex}|^2$, square of the distance to the true reflector depth. Hence the norm of the image residual $\|(z - z_{ex})\delta c\|^2$ is minimum when the reflector is located at its correct position. The objective of MVA is to find a background velocity model minimising the norm of the image residual. Note that this 1D case is somehow artificial as it requires the prior knowledge of the exact reflector position, which is an unknown in practice. Moreover, the approach is valid only for a single reflector. The 2D approach detailed at the beginning of this section is more general and does not require prior knowledge of the subsurface, but its principle is very similar.

The derivation of the velocity updates involves computing two new wavefields obtained as the interaction of the source (c) and receiver wavefields (d) with the image residual (g) (Lameloise *et al.*, 2014; Yang and Sava, 2015). The summation of the zero-lag cross-correlations of these new wavefields with the source and receiver wavefields gives the opposite of the velocity update ((h) and (k)). We obtain a constant value above the reflector altered with unwanted oscillations around the reflector which can not be easily attenuated by smoothing. This is one of the issues of MVA which has been discussed in section 1.3.4.

Finally, we repeat the same experiment with observed data containing both the primary reflection and a surface multiple (figure 1.14) to illustrate the potential issues caused by multiples. Hence, the multiple is treated as a primary and is migrated to twice the depth of the primary in the image domain (figures 1.14e and 1.14g). As a consequence, the image residual would not be minimal for the correct velocity. There are two differences with the preceding case on the velocity update (figures 1.14h and 1.14k): it has a non-zero value below the reflector, where no information can be recovered in theory, and the sign of the update is incorrect. This phenomena will be observed in the 2D case as well in chapter 4 and illustrates the need for a proper inclusion of multiples in MVA, as well as in other inversion strategies.

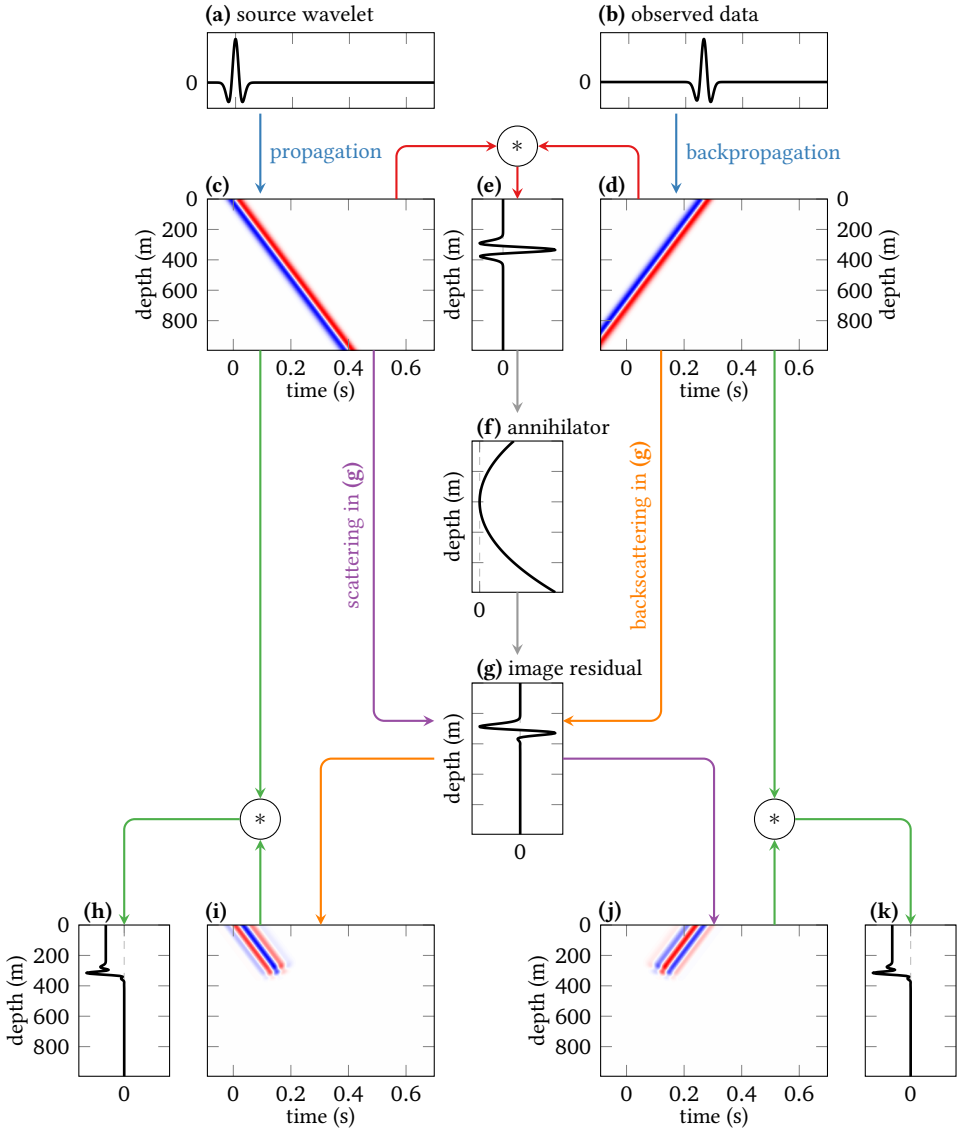


Figure 1.13. – Principle of MVA in a simple 1D case. The left column describes the migration process. The source wavefield (c) is constructed by propagating the source (a) in the medium. Data observed at the surface (b) are backpropagated to construct the receiver wavefield (d). Correlation of the source and receiver wavefield produces the model perturbation image (e). Energy away from the position of the true reflector (400 m) is penalised by an annihilator (f) to produce a modified migrated image (g) called "image residual". The latter is used to construct scattered and backscattered wavefields ((j) and (i), respectively). The (opposite of the) velocity update is made of two contributions ((h) and (k)), obtained by correlating (c) and (i) on one side, and (d) and (j) on the other side.

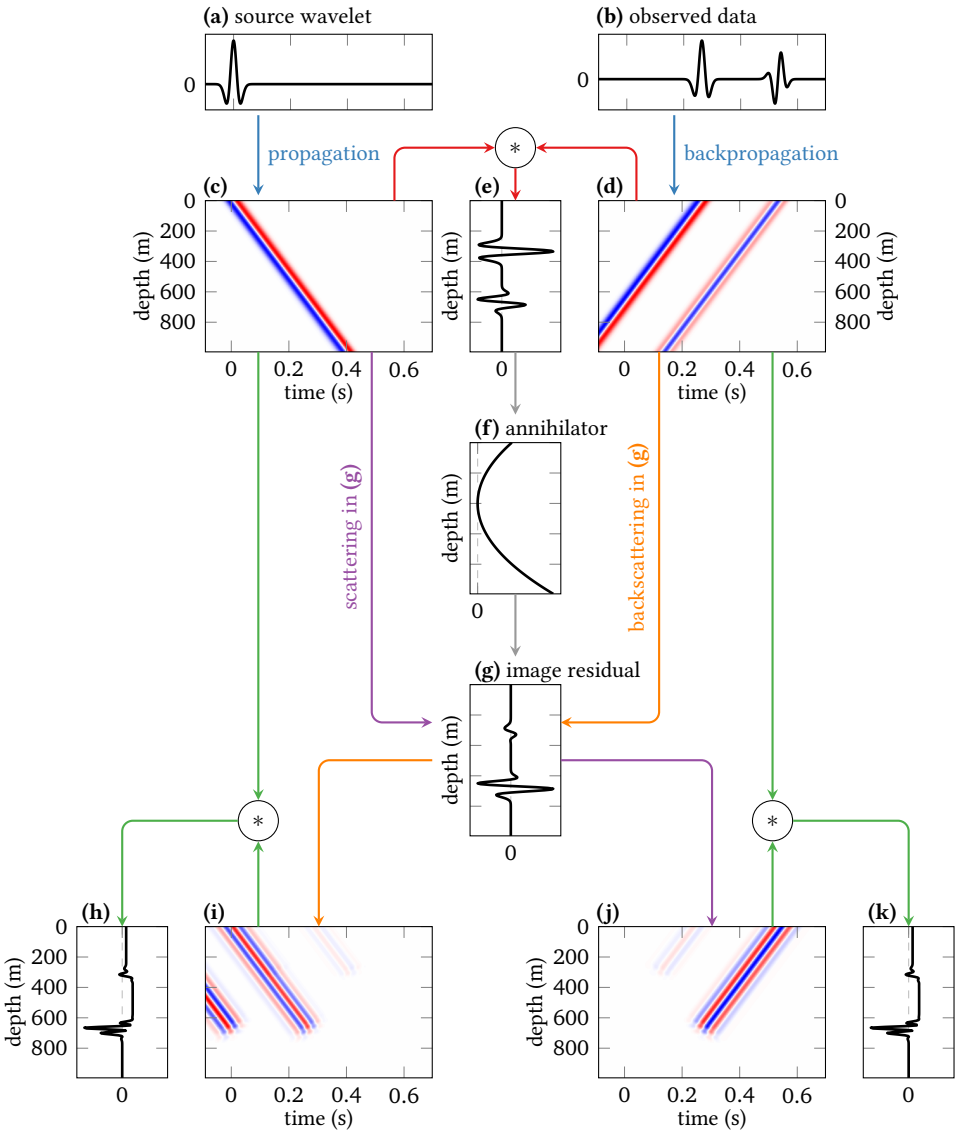


Figure 1.14. – Same as figure 1.13, but with a surface multiple in observed data.

Conclusion
The objective of the thesis is to investigate the inclusion of multiple reflections in the usual depth-oriented MVA. A strategy is developed, first in the case of primaries only, to make the method more robust against migration artefacts and spurious oscillations in the gradient. The method is then extended to the case of multiples.

To further motivate the inclusion of multiple reflection in MVA, we explain in the next section why their removal is challenging and what benefits they may bring to imaging procedures. In particular we report on recent extensions of migration techniques to multiples, which may help to define a MVA procedure handling multiples.

1.4. Multiple reflections

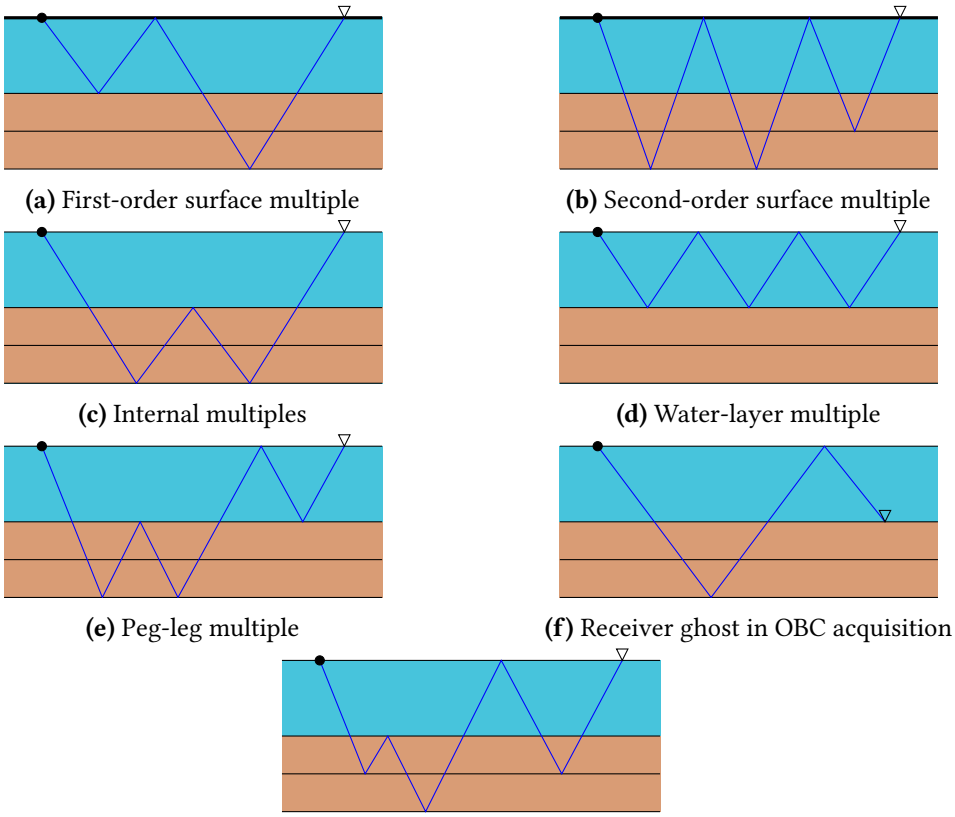
1.4.1. Physics of multiples

As defined by Verschuur (2013), primaries are waves that undergo only one upward reflection, whereas multiple reflections are characterised by at least one downward reflection. Because of spherical divergence and of the loss of energy occurring at each reflection, multiples effectively visible in the data are associated to strong reflectors. This is typically the case in marine acquisition where the interface between the air and the sea acts as a mirror. The water bottom is also a strong reflector. An other example is the chalk layer in the North Sea located between high-velocity layers, creating two high impedance contrasts which are strong multiples generators (Reinier *et al.*, 2012).

A first classification of multiples refers to the location where the shallowest downward reflection occurs. If it is the surface, the multiples are called *surface-related multiples* (figure 1.15a). Note that the upgoing multiple reflection can be reflected at the surface again, leading to higher-order surface multiples (figure 1.15b). In the following, a first-order surface-multiple refers to the multiple event undergoing a single downward reflection at the surface, and a second-order surface-multiple as the corresponding event for two downward reflections, etc. When the shallowest downward reflection occurs at an interface between two layers in the subsurface, the term *internal multiple* is used (figure 1.15c). Similarly, several order of internal multiples can be defined.

In the case of marine acquisition, the reverberations occurring in the water column are called *water-layer multiples* (figure 1.15d). Multiples undergoing at least one reflection below the water bottom and a reverberation in the water-layer are referred to as *peg-leg multiples* (figure 1.15e). In practical marine acquisition, both source and receivers are located a few meters below the sea surface. Then one part of the energy emitted by the source goes directly to the subsurface, while an other part reflects at the sea surface, leading to a surface multiple called *source ghost*. The same phenomenon occurs at the receiver side and is called *receiver ghost*. These kinds of multiples are also observed in the case of OBC acquisition (figure 1.15f). Naturally more complex paths in the subsurface which do not fit in the previous categories are possible (figure 1.15g).

One can also distinguish between *long-period* and *short-period* multiples. Long-period multiples (figure 1.16a) appear in the data record as separate events from their relative primaries. Visually the same succession of events is repeated at regular intervals of time (figure 1.17). This is the case with water-layer reverberations recorded in a deep-water environment. Oppositely,



(a) First-order surface multiple

(b) Second-order surface multiple

(c) Internal multiples

(d) Water-layer multiple

(e) Peg-leg multiple

(f) Receiver ghost in OBC acquisition

(g) A more complex multiple

Figure 1.15. – Different kinds of multiple reflections (similar to Verschuur, 2013, pp. 8-9). The black point and the white triangle represent source and receiver positions, respectively.

short-period multiples (figure 1.16b) are produced by reverberation within small layers such as chalk. These reverberations overlap with one another and with the original primary, resulting in a single event with a different wavelet signature in the recorded data.

(a) Long-period multiple **(b)** Short-period multiple

Figure 1.16. – Long-period and short-period multiples (see also Verschuur, 2013, p. 10).

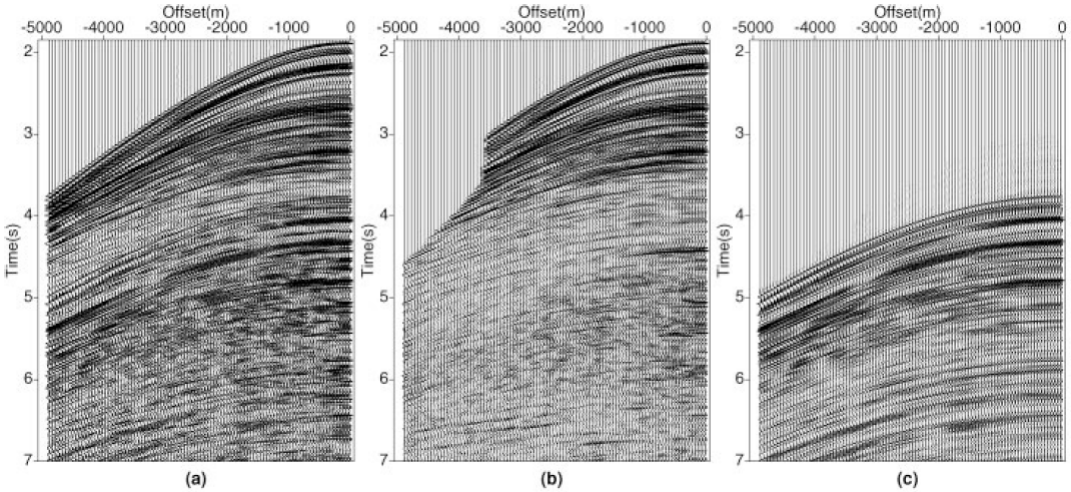


Figure 1.17. – (a) A real marine data set containing both primaries and multiples is decomposed into (b) primaries only and (c) multiples only (from Trad *et al.*, 2003).

Many imaging procedures designed for primary reflections require multiples to be removed from data. On the contrary, some methods presented in section 1.4.4 include them as complementary information. In any case, one needs the ability to model multiple reflections. Standard multiple modelling techniques are now reviewed.

1.4.2. Modelling of multiples

Surface-related multiples can be included in modelling by replacing the usual absorbing boundary condition at the surface by a free-surface. Alternatively, multiples can be predicted using primary reflections. First-order surface related-multiples can indeed be considered as the (primary-)response of the Earth to an areal source made of the primary reflections reaching the surface. Then first-order multiples can be obtained by re-injecting the primaries into a primaries modelling code. If the model of the first layer is known (for example the water layer

in marine acquisition), multiples can be predicted by adding a fictive roundtrip to the data through this layer with wavefield extrapolation (Wiggins, 1988, 1999). Then primaries are transformed in first-order multiples, first-order multiples in second-order multiples, and so on. The surface reflection coefficient should be taken into account (the coefficient -1 is the simplest approximation).

A data-driven alternative requiring no model of the subsurface is possible. In this approach, a first-order surface-multiple is obtained by autoconvolution of the primary reflection. The prediction is exact except for the source wavelet which is present twice in the multiple due to the autocorrelation. A shape correction has then to be applied. Similarly the second-order surface-multiple can be obtained by convolution of the primary with the first-order multiple, and so on. In the 1D case and with a Dirac source, the total response including all order of surface-multiples can be described as an infinite series:

$$d(t) = \widehat{\mathrm{P}_0}(t) + \widehat{\mathrm{P}_0}(t) * \widehat{\mathrm{P}_0}(t) + \widehat{\mathrm{P}_0}(t) * \widehat{\mathrm{P}_0}(t) * \widehat{\mathrm{P}_0}(t) + \cdots \tag{1.11}$$

where $*$ represent the convolution product and $\widehat{\mathrm{P}_0}(t)$ the primary impulsive response of the Earth. For the 2D and 3D case, a summation over the reflection point at the surface has to be added. More generally, the multiples contained in a data set can be generated by convolving the total data set with the impulse primary response (Verschuur *et al.*, 1992; Weglein *et al.*, 1997; Dragoset *et al.*, 2010)

$$\mathrm{M}(\mathbf{x}_s, \mathbf{x}_r, t) = \sum_{\mathbf{x}_k} \widehat{\mathrm{P}_0}(\mathbf{x}_k, \mathbf{x}_r, t) * d(\mathbf{x}_s, \mathbf{x}_k, t), \tag{1.12}$$

where $\mathbf{x}_k$ is the reflection point at the surface.

The interaction of internal multiples with the subsurface can be described by the Lippmann series (Lippmann, 1956; ten Kroode, 2002), originally used in quantum mechanics. The Born approximation used to model primaries is actually the first-order approximation of the series. Alternatively the Bremmer series can be considered (Bremmer, 1951; de Hoop, 1996). Berkhout (2014a) describes a Full Wavefield Modelling (FWMod) strategy which takes as input a reflectivity model and one-way propagation operators from one depth-level to the following based on the velocity model. The modelling is performed recursively: in the first roundtrip, the downgoing wavefield emitted by the source and the primary reflections are computed. The latter interact with the reflectivity model to generate both internal and surface first-order multiples in the second roundtrip. The process is repeated for higher-order multiples and stopped when their amplitude is too weak.

1.4.3. Removal of multiples

Multiples are sometimes easy to recognise in the data, for example long-period multiple appear as a repetition of the primaries pattern. The dip of primaries and multiples events arriving at the same time in the data can also be inconsistent. In more complicated cases, multiples may easily be mistaken for primaries. Moreover, usual migration algorithms are designed to account for primary reflections only. Hence multiples falsely interpreted as primaries can lead to spurious events in the final migrated image and incorrect interpretation of the geology (figure 1.18).

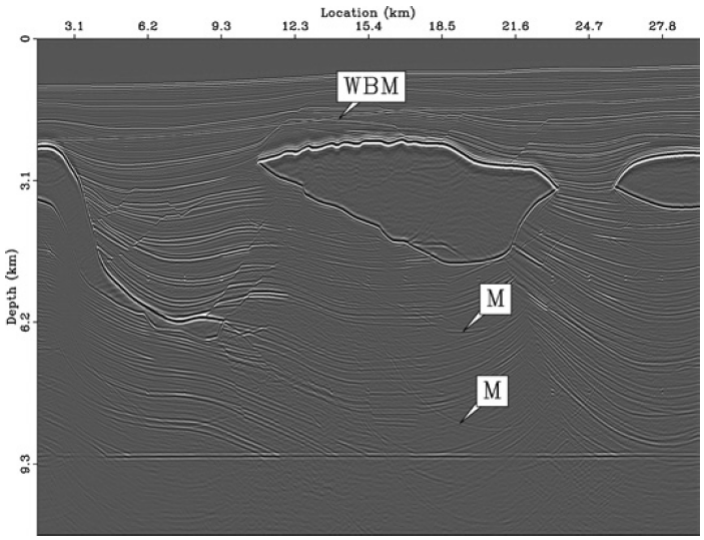


Figure 1.18. – Synthetic example of classic migration performed on data contaminated with multiples. WBM is a spurious events caused by a Water Bottom Multiple while M refers to surface related multiples repeating the bottom structure of the salt dome (from Sava and Guitton, 2005).

Numerous strategies to remove multiple reflections from seismic data have been designed. We review briefly the main families here and explain why multiple attenuation remains a challenging issue. The reader is referred to Verschuur (2013) for an extensive review.

Radon transform

The first family of multiple removal takes advantage of the fact that a multiple and a primary arriving at the same time in the data have not "seen" the same velocities in the Earth. With the hypothesis of a velocity increasing with depth, a multiple travels in shallower and slower layers of the subsurface. Then primaries and multiples do not exhibit the same move-out on Common Mid Point (CMP) gathers. Using a NMO-correction with the correct velocity, primaries are flattened but multiples are not. A strategy using these move-out discrepancies transform the CMP gathers to a space where primaries and multiples are easily separable. Multiples are muted and the inverse transform is performed to output CMP gathers without primaries. The double Fourier transform mapping CMP gathers to the f-k domain (Ryu, 1982) is an example of such a transformation, but the most popular choice is the parabolic (Hampson, 1986; Kabir and Marfurt, 1999) or hyperbolic (Foster and Mosher, 1992) Radon transform. The method assumes that primaries and multiples in CMP gathers (with or without NMO-correction) can be described by different parabolic or hyperbolic functions. Then in the Radon domain, they should appear as focused events. After muting of the multiples, the primaries are reconstructed with the inverse Radon transform. In practice, artefacts appear in the Radon domain and primaries and multiples are not well separated, leading to inaccurate multiple attenuation. A remedy consists of redefining this strategy as an inversion aiming at finding the model in the Radon domain that best fits the original data after inverse Radon transform. A sparsity constraint in

the Radon domain is added to ensure primaries and multiples are easily separable (Sacchi and Ulrych, 1995; Trad *et al.*, 2003). However the method still has difficulties in case of complex geology when primaries and multiples do not exhibit enough move-out difference or when they cannot be described by parabolic or hyperbolic functions. To better account for the propagation effect in complex media, the same analysis can be transposed to the image-domain where the move-out discrepancies are analysed in CIGs function of the surface offset (Duquet and Marfurt, 1999) or the scattering angle (Sava and Guitton, 2005). The velocity model is assumed to be relatively accurate, so that primaries are flat in the angle-domain CIGs, contrary to multiples. Multiples are then easily recognisable and a similar strategy using the Radon transform is used to discriminate between flat and curved events. As in the data-domain the possible overlap of primaries and multiples at zero-angle may harm the attenuation of multiples. The main hypothesis in this approach is the knowledge of a reliable velocity model.

Adaptive subtraction

The second family of multiple removal techniques is a two-step procedure. First the data set is used to compute a prediction of the multiples following one of the method presented in section 1.4.2. Then the prediction is subtracted from the data to yield the primary estimation.

This second step is difficult because the prediction of multiples is in practice never perfect. In wavefield extrapolation methods, inaccurate prediction may be caused by an incorrect model of the layer in which the multiple reflection is simulated. Multiples predicted by convolution of the data with the primary response contain the source wavelet twice. Moreover, predictions based on a 2D-model assumption do not take into account the 3D-propagation effects. Therefore there may be inaccuracies in the phase, amplitude, and wavelet of predicted events (Abma *et al.*, 2005) and simple subtraction of the prediction to the original data does not yield a good estimation of the primaries. The prediction has to be accommodated to the recorded multiples in a process called *adaptive subtraction*, which consists of applying a filter to the multiples before subtraction. The filter is determined by minimising the energy of the differences between the original data and the filtered multiples in a least-squares sense. If primaries and multiples are overlapping in the data domain, minimising the energy can result in distortion of the primaries and/or residual multiple in the final result (Nekut and Verschuur, 1998). Alternatives consider replacing the least-squares criterion by the ℓ_1-norm (Guitton and Verschuur, 2004), among other possibilities (Batany *et al.*, 2016). The subtraction can also be performed in a domain where primaries and multiples are less likely to interfere such as the curvelet domain, Radon domain, or frequency domain (Sacchi and Ulrych, 1995; Donno *et al.*, 2010; Batany, 2016).

A very-well known strategy based on multiple prediction and adaptive subtraction is the Surface-Related Multiple Elimination (SRME) method (Verschuur *et al.*, 1992). It is based on the observation that a surface-multiple can be decomposed in several primary reflections connected by a reflection point at the surface. If the source and receiver coverage of the acquisition is dense enough, these primaries are contained in the data and can be used to predict the multiples. SRME is implemented as an iterative method aiming at improving an initial estimation of the primaries (Berkhout and Verschuur, 1997). This estimation is convolved with the data to produce a prediction of the multiples. Then an adaptive subtraction is performed to remove the estimated multiples from the total data, yielding an improved estimation of the primaries. The first estimate of the primaries can be the output of another multiple elimination technique or

simply the complete data. This process converges very fast in practice and only a few iterations are needed (Berkhout and Verschuur, 1997). One major advantage of this method is that it is fully data-driven and requires no model of the subsurface. Conversely, the limitations of the method are related to the data acquisition. A requirement is that all the primaries composing the multiples should be recorded in the data. This may not be the case in typical marine acquisition, where the first offsets near the source are missing for practical reasons. This represents also an issue for the extension to 3D marine acquisition where the direction perpendicular to the streamers is sampled with only a few lines. In practice, these limitations require a method to reconstruct the missing primaries, for example by interpolation of the data (Kabir and Verschuur, 1995).

Estimation of Primaries by Sparse Inversion (EPSI) (van Groenestijn and Verschuur, 2009) addresses the issue of inaccurate multiple prediction issue by recasting the prediction and the adaptive subtraction into a single inversion procedure. An objective function compares observed data to predicted primaries and predicted multiples. The corresponding iterative process should update the estimation of the primaries as well as the source wavelet. EPSI provides enhancements compared to SRME but is computationally more expensive.

Note that SRME is designed to deal with surface-related multiples only. A direct adaptation to internal multiples would consist of wavefield extrapolation to calculate the data that would be recorded at fictive sources and receivers located on the boundary generating the internal multiples. The process should be repeated for each layer and has the disadvantage of requiring an accurate velocity model to process the wavefield extrapolation. Jakubowicz (1998) proposes an alternative formulation in which the redatuming uses the primary reflections contained in the data, discarding the need for a subsurface model. Alternatively an inverse scattering series can be used to predict all internal multiples in a fully data-driven manner (Weglein *et al.*, 1997; ten Kroode, 2002)

1.4.4. Using multiples as valuable information

A lot of effort has been put to develop multiple removal techniques because migration and velocity analysis techniques were designed for primary reflections only. However in the last decade, a different approach considering multiples as signal rather than noise has gained popularity. Multiples have travelled at least twice through the subsurface and therefore contain a lot of valuable information that may not be present in primary reflections. For example they may travel in area of the subsurface not illuminated by primaries like beneath salt domes (figure 1.19). Moreover a surface multiple may originate from a virtual source not present in the survey, for example when short-offsets are missing in the acquisition, thus providing illumination with smaller angle reflections, different wavenumbers for imaging and better vertical resolution than primaries (figure 1.20). In particular the cross line of a 3D marine acquisition is coarsely sampled and the so-called *acquisition footprint* is still visible in the data. Long *et al.* (2013) show that this artefacts can be mitigated with the use of multiple reflections. A correct use of multiple reflections could then possibly relax the need of dense source and receiver coverage at the surface. The sensitivity of multiples to the velocity model has been studied by some authors to detect small velocity changes in time-lapse experiments (Snieder, 2002; Verschuur and Staal, 2014).

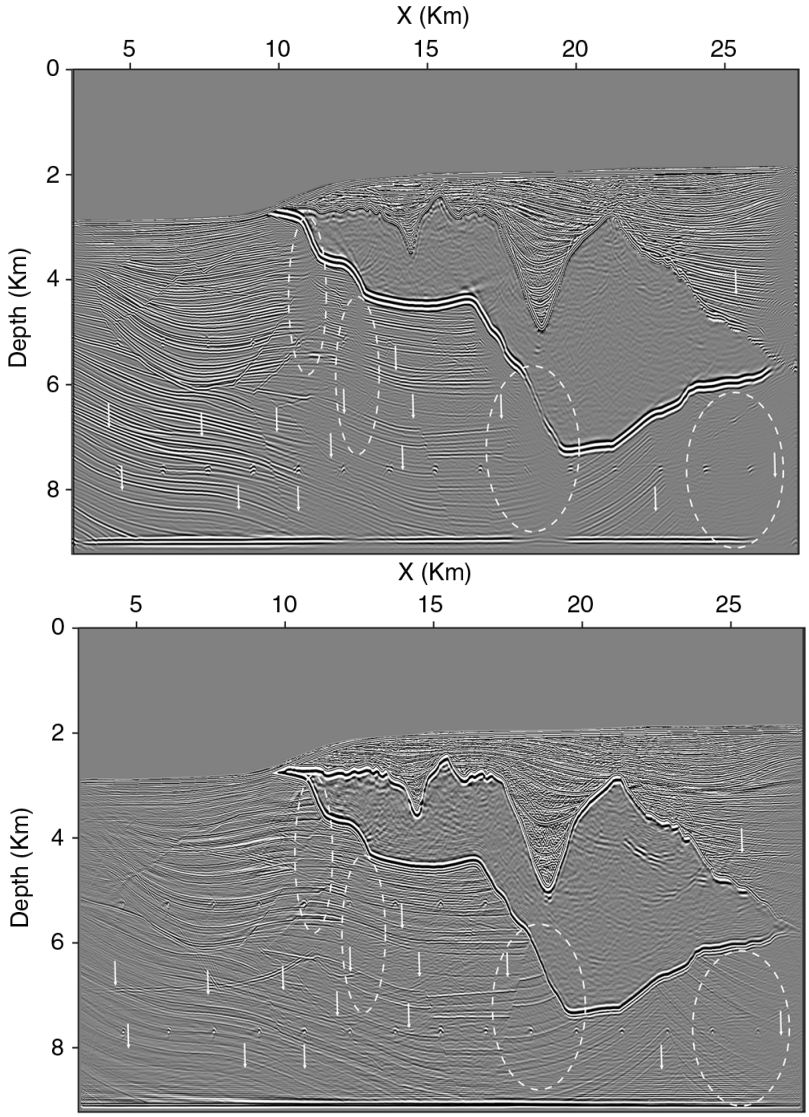


Figure 1.19. – Example of conventional migration using primaries only (top) and using primaries and multiples (bottom) on the Sigsbee2B synthetic data set. Note the extended illumination provided by multiples below the salt dome in the areas marked with the white ellipses (from Liu *et al.*, 2011).

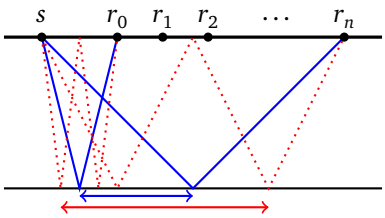


Figure 1.20. – Illumination of a flat reflector by primary reflections (blue, solid) and first-order surface-related multiples (red, dotted). Multiples allow a wider coverage of the reflector with smaller scattering angles.

1.4.5. Imaging with multiples

A simple way of using multiple reflections consists of transforming them into pseudo-primaries which can be imaged with classic migration algorithms. The transform can be achieved with seismic interferometry (Schuster *et al.*, 2004), which states that the response recorded at a point B to a fictive source located in A can be obtained by cross-correlating the passive measurements made in A and B. When A and B are two receivers located at the surface, the cross-correlation of the traces recorded in A and B results in the primary reflection linking the two points (Sheng, 2001; Shan, 2003; Schuster *et al.*, 2004; Jiang *et al.*, 2005, 2007). Alternatively Berkhout and Verschuur (2003, 2006) deconvolve the data with an estimation of the primaries (obtained with SRME for example) to transform first-order surface multiple into pseudo-primaries, second-order surface multiple into pseudo-first-order surface multiple, and so on... Then a new application of the SRME method isolates the pseudo-primaries from higher-order multiples and a classic migration algorithm can be applied.

Conversely the usual migration algorithms can be adapted to multiple reflections. A first attempt was made by Reiter *et al.* (1991) who used Kirchhoff migration to separately image primary reflections as well as ghost reflection against the surface in OBC data. Their approach assumes that multiples and primaries are separated in time in the data and is thus restricted to deep-water acquisition. More recently this constraint was released by the development of sensors able to separate up and downgoing energy, thus discriminating between upgoing primaries and downgoing receiver ghosts (Muijs *et al.*, 2007; Whitmore *et al.*, 2010).

The usual imaging condition for primary reflection can be transposed to multiple reflections. It still consists of correlating a source and a receiver wavefield, except that the source wavefield is the forward propagation of the entire data set and the receiver wavefield the backward propagation of the same data set without primary reflections (Berkhout and Verschuur, 1994; Guitton, 2002; Liu *et al.*, 2014a). Compared to the usual primary imaging condition, the source wavefield does not originate at a single point but is an areal shot with virtual sources located at every receiver position. An advantage of this method is that there is no need to estimate the shape of the source wavelet as the new pseudo-source is directly the data recording. With this methodology, two images can be constructed, one from the migration of primaries, and one from the migration of multiples.

However, a correct separation of the primaries and the multiples is required and the final image is contaminated with *cross-talk* artefacts (figure 1.21). These spurious events are caused

by downgoing energy interfering with an unrelated upgoing energy, for example a primary with a second-order surface multiple. To avoid these artefacts, each order of multiple should be isolated to produce a separate migrated image.

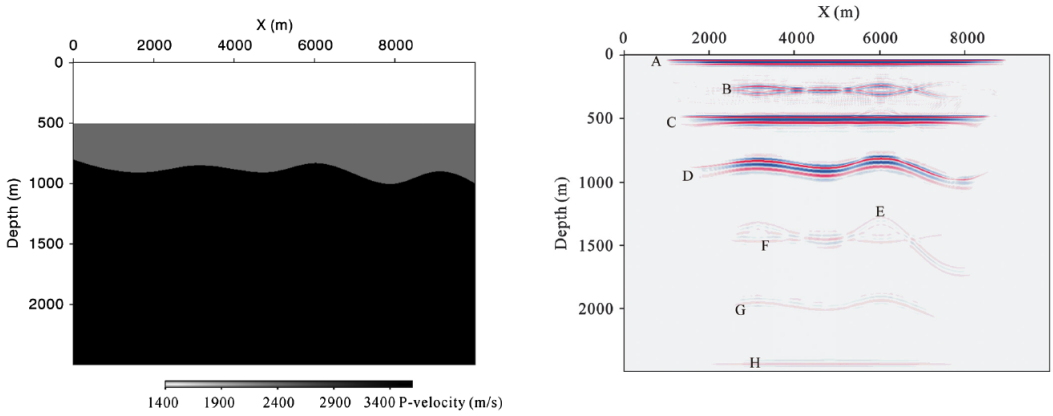


Figure 1.21. – Example of cross-talk artefacts. The exact velocity model is displayed on the left panel. The right panel shows the image reconstructed from the migration of multiple reflections. C and D correspond to the true reflectors. All other events including A and B are cross-talk artefacts (from Liu *et al.*, 2011).

As a remedy to the *cross-talk* artefacts, the use of the deconvolution imaging condition instead of the cross-correlation has been studied (Muijs *et al.*, 2007; Whitmore *et al.*, 2010), however this method has limited success in complex media (Poole *et al.*, 2010; Tu *et al.*, 2013). More recent studies use a least-squares inversion approach based on fitting the data reconstructed with the estimated reflectivity model and observed data (Brown and Guitton, 2005; Verschuur and Berkhout, 2011; Wong *et al.*, 2014; Zhang and Schuster, 2014; Tu and Herrmann, 2015). When modelling data with reflectivity image contaminated with cross-talk artefact, extra-reflections will appear in synthetic data. These will be back-projected in the reflectivity update to attenuate the cross-talk artefacts.

If internal multiples are properly modelled, they can be included in the inversion procedure. This is the approach of the Full Wavefield Migration (FWM) technique (Berkhout, 2012, 2014b; Soni and Verschuur, 2014) where the modelling is performed with the Full Wavefield Modelling procedure presented in section 1.4.2. The amplitude differences between modelled and observed data is converted into reflectivity update.

More recently, new imaging techniques have been derived from the resolution of the Marchenko equation (Wapenaar *et al.*, 2014). As in reverse-time migration, Marchenko imaging consists of correlating a downgoing and an upgoing wavefield at each point of the subsurface. However in reverse-time migration, the downgoing wavefield is approximated by direct propagation in a non-reflective media of the source wavelet to the image point, and the upgoing wavefield is obtained in the same way by backpropagation of the residuals. In Marchenko imaging, all the internal multiples reflections encountered by the source and receiver wavefield are accounted for, so that the wavefields correlated at each subsurface location correctly include internal multiples (Behura *et al.*, 2014). As a consequence, Marchenko imaging does not generate cross-talk.

These wavefields are obtained from surface recordings and an estimation of the direct arrival recorded at each subsurface location by iteratively solving the Marchenko equation (Rose, 2002; Wapenaar *et al.*, 2011). Note that a velocity model is needed for the direct arrival estimation. Singh *et al.* (2015) extend this procedure to use simultaneously primaries, internal multiples and surface multiples.

1.5. Motivations and thesis outline

1.5.1. Motivations: towards a more robust MVA and extension to multiple reflections

Real seismic data contain both primaries and multiple reflections. Much effort has been devoted to the removal of multiple reflections to provide migration algorithms with data containing only single-scattered events. As a consequence, Migration Velocity Analysis techniques have been designed for single-scattered events only. Recent developments showed that transmitted waves could be incorporated in MVA (Shen, 2013; Lameloise and Chauris, 2016). Extending MVA to multiple reflections would have two advantages. First this would remove the difficult pre-processing stage of multiple attenuation, second the information contained in multiple reflections may be used to update the background velocity model in areas not illuminated by primaries.

So far multiples have been considered as noise in MVA techniques (Mulder and ten Kroode, 2002; Li and Symes, 2007), and only a few proposals have been made to include them as valuable signal. van Leeuwen and Mulder (2008b) study the behaviour of multiples in the data-domain correlation methodology proposed in van Leeuwen and Mulder (2008a). They propose to update alternatively the reflectivity and the velocity model to both maximise the correlation in the data domain and minimise the misfit between observed and calculated data. Another strategy in the data domain is Joint Migration Inversion (JMI) (Staal and Verschuur, 2012; Berkhout, 2014c; Staal, 2015). It is an extension of the Full Wavefield Migration technique presented in section 1.4.4 which includes internal and surface multiples in the usual data misfit objective function. The amplitude of data residuals is used to update the reflectivity model and the phase is used to update the velocity model. Both models are updated simultaneously at each iteration. Note that this approach does not use an extended-reflectivity. To our knowledge, only two proposals have been made to include multiples in usual image-domain methods. Nasyrov *et al.* (2008, 2009) propose an original strategy in which primaries and first-order surface-related multiples are migrated separately. They consider an objective function measuring the similarity of the two images, so that the discrepancies between the image obtained with primaries and the image obtained with multiples are used to update the velocity model. Alternatively, the sensitivity of the Marchenko imaging procedure to the background velocity has been very recently investigated by Díaz *et al.* (2016) who compute extended reflectivity images in subsurface-offset and angle domain with internal and surface multiples. The dependence of the Marchenko wavefields to the background velocity is not as explicit as in the case of RTM, thus a method for the derivation of a velocity update should still be investigated.

Even in the case of primaries only, MVA still faces difficulties as discussed in section 1.3. Therefore before introducing multiples, a robust method to deal with primary reflection has to

be determined. Lameloise *et al.* (2014) have shown that the issue of migration artefacts can be mitigated by introducing weights in the migration procedure. As a consequence, the gradient of the DSO objective-function is greatly improved. However their approach rely on ray theory and does not extend easily to multiple reflections. The extension to wave-equation based operators presented in chapter 4 is inspired from ideas by ten Kroode (2012) and Hou and Symes (2015) who have developed a pseudo-inverse of the extended Born modelling operator. In chapter 4, we propose a slightly modified version and more importantly a coupling with velocity analysis (Hou and Symes, 2016b).

In this study I first investigate a modification of the standard MVA procedure consisting of performing the migration step iteratively to determine a better solution of the migration inverse problem than the one provided by standard migration. The method is not limited to ray theory and can handle multiple reflections. The analysis is restricted to first-order surface-related multiples modelled with a second-order Born approximation. An extended reflectivity section is constructed to minimise the data misfit between calculated and observed data including both primaries and first-order surface-multiples. At the convergence, the iterative procedure is expected to yield CIGs free of migration and cross-talk artefacts, thus allowing to use standard MVA, here subsurface-offset DSO.

This results in a nested optimisation procedure (figure 1.22). In the inner loop, the extended-reflectivity is determined as a minimisation of the data misfit, while in the outer loop the velocity model is updated to minimise defocused energy in the optimal reflectivity. There are two main differences with the JMI approach. First here an extended-reflectivity is considered. This allows to capture all the information contained in the data residual, even in the case of an inaccurate initial velocity model. Second the update of the velocity is based on defocusing in the extended-domain and not on data-misfit. Besides, the JMI approach aims at recovering both the reflectivity and the velocity which are updated simultaneously at each iteration whereas here a nested optimisation approach is studied, whose main objective is the determination of a correct background velocity model.

In the synthetic examples, we will first test our method on primaries only data, and then on data containing both primaries and first-order surface multiples. We will focus on the first gradient of the outer loop. The objective is that the result obtained with primaries and multiples in observed data should be close to the one that would be obtained with observed data free of multiples, meaning that multiples are correctly interpreted by the iterative migration algorithm and do not hinder the velocity analysis procedure. To reach this objective, a number of essential elements have to be developed, successively detailed in chapters 2 to 5. Some of them are also useful for a stable MVA in the case of primaries only: in particular, we pay attention at the issue pointed out by Huang (2016) regarding the stability of the MVA gradient computed with a fixed background velocity model and successive values of the reflectivity model. Each chapter concludes with a limitation, justifying further developments. The final formulation is obtained in chapter 5.

1.5.2. Thesis outline

The thesis is organised as follows:

- Chapter 2 is dedicated to the iterative migration procedure, corresponding to the inner

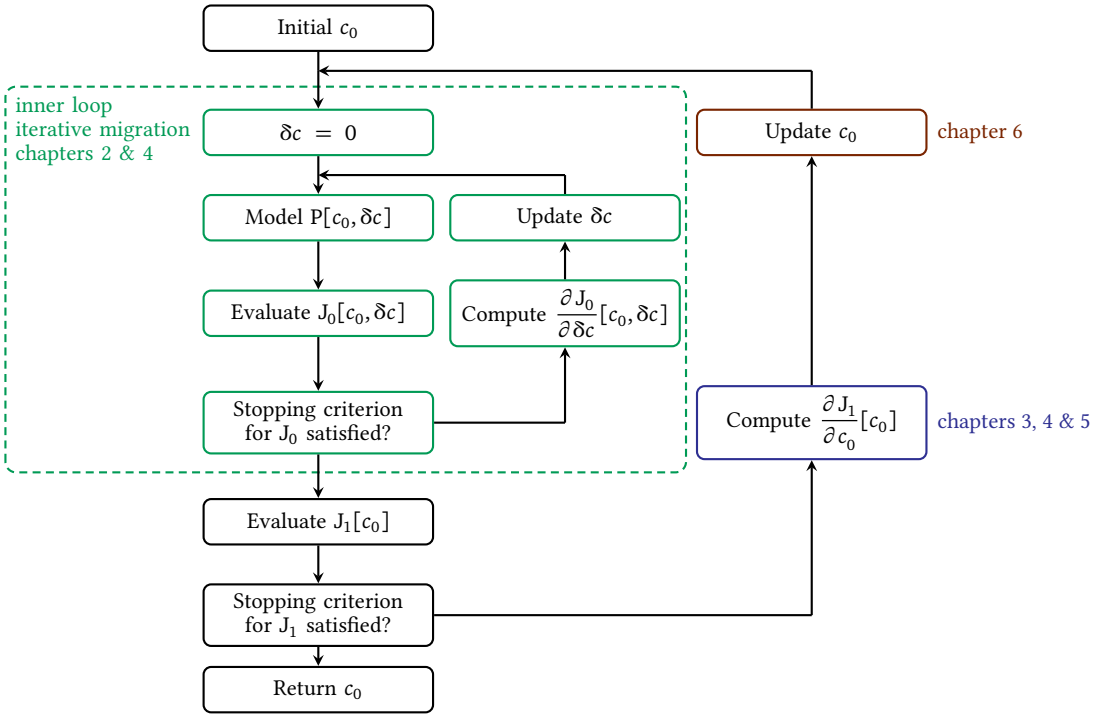


Figure 1.22. – Sketch showing the algorithm of MVA with iterative migration. In the inner loop, we optimise the reflectivity δc for a given c_0 in order to minimise the objective function J_0 measuring the data misfit $\|P[c_0, \delta c] - P^{obs}\|$. In the outer loop, we update the velocity model to minimise the objective function J_1 measuring defocused energy in the CIGs of δc.

loop of figure 1.22. The modelling of multiples with a second-order Born approximation is introduced and compared to a finite difference modelling with a free-surface condition. An efficient derivation of the gradient of the migration objective function using the adjoint state method is presented. We show that even in the presence of multiples, iterative migration yields clean CIGs free of migration and cross-talk artefacts and therefore improve the properties of the associated MVA objective function.

- In chapter 3, two methods for computing an approximate gradient of the MVA objective function after iterative migration are investigated. In the first case, we compute an exact gradient of an approximate objective function. In a second approach, we suppose that iterative migration has converged and compute the approximate gradient of the ideal MVA objective function. This method is preferred as it has a simpler implementation. It consists of solving a so-called "adjoint problem", which is a linear system very similar to iterative migration. We highlight the importance of introducing regularisation during migration to obtain coherent velocity updates. We discuss also here instabilities observed in the gradient and a first possibility to attenuate them. The second possibility is developed in chapter 5. The approach remains expensive, as two iterative schemes have to be solved. Chapter 4 aims at obtaining a faster convergence.

- In chapter 4, an approximate inverse of the extended-Born modelling operator is presented. This direct inversion formula is designed for primary reflections only and is free of ray quantities. It is a variant to the method introduced by Hou and Symes (2015). The new aspect is the coupling with MVA. The weights are introduced as a preconditioner to accelerate the resolution of iterative migration and of the adjoint problem, even in the presence of multiples. In the primary-only case, we compare the results of direct and iterative inversions: we show that although they provide very close reflectivity images, the associated gradients are not similar. We also underline the difficulty to solve the adjoint problem compared to iterative migration (Huang, 2016). This justifies additional developments discussed in chapter 5.

- In chapter 5, the behaviour of iterative migration velocity analysis is studied on a pure 1D primary-only case to better understand the issues presented in chapter 4. A major advantage of the 1D analysis is the possibility to have explicit operators and to run a large number of iterations. We propose a simple modification of the MVA objective function consisting of applying a "filter" to the final CIGs before penalising defocused energy. The filter is function of the background velocity model. In the primaries only case, this modified approach improves the convergence of the adjoint problem and yields gradients close to those obtained after direct inversion, both in 1D and 2D. In the case of multiples, we show that introducing classical regularisation on the migration during the migration step is still essential.

- In chapter 6, the approach is applied on 2D synthetic data sets. The robustness of the method is discussed with observed data computed with a different modelling engine from the one used for the inversion, for example with the introduction of a variable density, or when a different source wavelet is used for inversion and for observed data.

- In chapter 7, the main conclusions of the thesis are summarised and we discuss remaining issues in the perspective of real data applications.

1.5.3. Realisations and contributions

Starting from an existing finite-difference acoustic forward propagation code, I have built an iterative migration Fortran90 code computing extended CIGs by iterative minimisation of the data misfit with non-linear optimisation techniques. I implemented the two methods for the computation of the background velocity update presented in chapter 3, as well as the inversion formula described in Chauris and Cocher (2017) used here as a preconditioner. The implementation of adjoint operators has been verified with the dot-product test (Claerbout, 2014, p. 28) and gradient derived with the adjoint-state method (Plessix, 2006) have been compared with finite-difference computations.

My main contributions are

- the construction of CIGs in the subsurface-offset domain with consistent inclusion of first-order surface-related multiples;

- the extensive study of the construction of the gradient of the MVA objective function after iterative migration, as detailed in chapters 3 and 5;
- the comparison of direct and iterative inversion schemes for the construction of background velocity updates and the proposition to filter final CIGs in the iterative case to stabilise the method.

These developments allow a better understanding of the behaviour of MVA after direct or iterative inversions. We have proposed strategies for a more robust scheme, applicable in the case of surface-related multiples but also of primaries only.

Part of the results have been presented in

- E. Cocher and H. Chauris, 2014. Iterative Migration to Remove the Imprint of Multiples on the Reflectivity. SEG/KOC Workshop: Seismic multiples – Are they signal or noise?, Kuwait City, Kuwait;
- E. Cocher, H. Chauris and C.-A. Lameloise, 2015. Imaging with Surface-Related Multiples in the Subsurface-Offset Domain. 77th EAGE Conference & Exhibition, We N101 11;
- H. Chauris and E. Cocher, 2014. Iterative Migration for Velocity Analysis in the Presence of Surface-Related Multiples. Workshop on using multiples as signal for imaging, 2014 SEG Annual Meeting;
- H. Chauris, C.-A. Lameloise and E. Cocher, 2015. Inversion Velocity Analysis - The Importance of Regularisation. 77th EAGE Conference & Exhibition, WS05–A02;
- C.-A. Lameloise, H. Chauris and E. Cocher, 2015. Automatic Migration Velocity Analysis Applied to Direct Waves in a Crosswell Configuration. 77th EAGE Conference & Exhibition, We P1 05;

or submitted to publication,

- H. Chauris and E. Cocher, 2017. From Migration to Inversion Velocity Analysis. Geophysics, **82**(3), S207–S223;
- E. Cocher, H. Chauris and R.-É. Plessix, 2017a. Seismic Iterative Migration Velocity Analysis: Two Strategies to Update the Velocity Model. Computational Geoscience. Submitted (minor revisions);
- E. Cocher, H. Chauris and R.-É. Plessix, 2017b. Towards a Stable Iterative Migration Velocity Analysis Scheme. Geophysical Journal International. In preparation.

Chapter 2.

Iterative migration

Contents

Résumé du chapitre 2

Ce chapitre décrit l'algorithme de migration itérative et l'illustre sur un exemple simple dans le cas de primaires seuls, puis dans le cas de primaires et de multiples de surface du premier ordre. La migration itérative fait l'hypothèse d'une séparation d'échelle du modèle de vitesse en un *macro-modèle* lisse et en un modèle de *réflectivité* hautes fréquences. Pour une valeur fixée du macro-modèle, il s'agit alors de déterminer un modèle de réflectivité qui minimise une fonction objective mesurant l'écart entre données observées et données recalculées. Cette fonction coût comprend également un terme de régularisation mesurant la norme ℓ_2 de la réflectivité. Le macro-modèle est fixe au cours de l'optimisation et peut être erroné. Afin d'évaluer la qualité

du macro-modèle utilisé, le modèle de réflectivité est étendu par l'introduction de l'offset en profondeur horizontal comme paramètre additionnel. Les erreurs dans le macro-modèle se traduisent par la présence d'énergie pour des valeurs non nulles de l'offset en profondeur dans le modèle de réflectivité final. Cette information sera utilisée au chapitre 3 pour remettre à jour le macro-modèle.

Dans le cas de primaires seuls, les données calculées sont modélisées sous l'approximation de Born et dépendent linéairement du modèle de réflectivité. Dans le cas où les multiples de surface du premier ordre sont également considérés, l'approximation de Born du second ordre est utilisée et la relation entre données et réflectivité n'est plus linéaire. Cette modélisation est comparée à une modélisation par différences finies sans approximation de Born et avec une condition de surface libre utilisant la méthode d'image pour modéliser les multiples de surface. L'approximation de Born est satisfaisante pour de faibles contrastes de vitesse. En présence de plus forts contrastes, l'amplitude d'évènements correspondants à de grands angles de réflexions est surestimée. Cela peut consister une limitation de cette approche en vue d'applications à des données réelles. Finalement une modélisation efficace des primaires et des multiples dans le cas d'un modèle de réflectivité étendu avec l'offset en profondeur est introduit. Pour les multiples, j'ai choisi la solution la plus générale avec l'introduction un offset en profondeur à chaque point de réflexion.

La minimisation de la fonction coût associée à la migration itérative est réalisée avec des méthodes d'optimisation locale (Nocedal et Wright, 2006) utilisant le gradient de la fonction coût pour remettre à jour le modèle de réflectivité. Le calcul du gradient, obtenu par la méthode de l'état adjoint (Plessix, 2006), est détaillé. Dans le cas linéaire de primaires seuls, la minimisation de la fonction objective est équivalente à la résolution d'un problème linéaire. Dans les exemples numériques, ce système est résolu itérativement avec l'algorithme du gradient conjugué linéaire. Lorsque des multiples sont considérés, des techniques d'optimisation non linéaire doivent être employées, ici le gradient conjugué non linéaire utilisant la formule de Polak-Ribière et la recherche linéaire de Moré et Thuente (1994).

La migration itérative est illustrée dans le cas simple d'un unique réflecteur horizontal et d'un macro-modèle homogène. Dans un premier temps, les données observées (synthétiques) et calculées ne contiennent que des réflexions primaires. Un macro-modèle homogène mais plus lent que le modèle exact, est utilisé. À la première itération, les artefacts de migration décrits par Mulder (2014) et Lameloise *et al.* (2014) apparaissent dans les CIGs et le minimum de la fonction coût évaluée pour des macro-modèles homogènes n'est pas atteint pour la bonne valeur de vitesse. En poursuivant les itérations, les artefacts de migration sont progressivement atténués et le minimum décalé vers la bonne valeur de modèle de vitesse. Le test est répété dans le cas où les multiples de surface du premier ordre sont ajoutés aux données observées. À la première itération, les multiples interprétés comme des primaires conduisent à de nouveaux artefacts dans les CIGs à une profondeur double de la position du vrai réflecteur. À l'itération suivante, ces artefacts génèrent dans les données calculées de nouveaux évènements (primaires et multiples) absents des données observées, qui conduisent à leur tour à une remise à jour du modèle de réflectivité. Ainsi les artefacts dus aux multiples, tout comme les artefacts de migration sont progressivement atténués.

2.1. Introduction

This chapter introduces an iterative migration scheme (Nemeth *et al.*, 1999; Østmo *et al.*, 2002) with a specific focus on surface-related multiples. A scale separation between a smooth *background velocity model* and a rapidly-varying *reflectivity model* is assumed. The purpose of migration is to determine a reflectivity model which allows to faithfully reproduce observed data, assuming a fixed initial background velocity model. The latter may be inaccurate, therefore the reflectivity model space is extended with an additional parameter, the subsurface offset. A reflectivity model explaining observed data can then be obtained even in an incorrect velocity model. Here we want to determine a reflectivity model able to explain correctly both primaries and first-order surface multiples. The evaluation of the quality of the background velocity model (MVA part) is investigated in chapter 3.

Migration is formulated here as an inverse problem which consists of minimising an objective function measuring the misfit between observed and simulated data. This corresponds to the inner loop of the scheme described in figure 1.22. In the forward problem, data corresponding to a reflectivity model are modelled with a second-order Born approximation: primaries are linear in the reflectivity, but multiples are not. The inverse problem is solved iteratively with a gradient-based method, the reflectivity model being updated with the gradient of the least-squares objective function.

Starting from a zero model, the first gradient does not provide a reliable solution to the inverse problem. First, multiples incorrectly interpreted as primaries result in spurious events in the reflectivity model which do not correspond to a physical reflector. Second, even in the case of primaries only, migration artefacts appear on Common Image Gathers (CIGs), defined as panels displaying the reflectivity along the depth and the subsurface-offset for fixed lateral positions (Lameloise *et al.*, 2014; Mulder, 2014). CIGs are later used to evaluate the quality of the background velocity model. Migration artefacts are thus an issue for the update of the background velocity model. We show that iterative migration allows to attenuate both migration artefacts and spurious events caused by cross-talk.

We begin by describing the forward problem consisting of modelling primaries and first-order surface-related multiples with a second-order Born-approximation. The results are compared with a classical finite-difference modelling with a free-surface condition to investigate the ability of the Born approximation to deal with high velocity contrasts, which generate strong multiples. We then show how the modelling is performed in the extended-domain with the horizontal subsurface-offset as extension parameter. The inverse problem associated to iterative migration and the corresponding objective function are presented. It is solved with non-linear gradient-based optimisation techniques and an efficient way of computing the gradient of the objective function is introduced. Finally, we show simple synthetic examples. First we consider primary reflections only to show the influence of migration artefacts (migration smiles) in CIGs on the shape of the MVA objective function. We show how iterative migration can mitigate the problems caused by migration artefacts. The same tests are run with surface multiples and the attenuation of cross-talk artefacts as well as migration artefacts with iterations is illustrated.

2.2. Notations

We consider a surface acquisition and a 2D Earth model where a subsurface point is noted $\mathbf{x} = (x, z)$. We define

- the **model space** $\mathscr{M}$ of physical model parametrised by $\mathbf{x}$. For example the background velocity model $c_0(\mathbf{x})$;
- the **extended model space** $\mathscr{E}$ of variables function of the subsurface point $\mathbf{x}$ and the subsurface offset considered strictly horizontal $\mathbf{h} = (h, 0)$, for example the extended reflectivity $\xi(\mathbf{x}, h)$;
- the **observed data space** $\mathscr{D}_{\text{obs}}$ parametrised by the source horizontal coordinate s, the receiver horizontal coordinate r and the angular frequency ω (or time t) in the frequency domain formulation. Observed data are noted $\mathrm{P}^{\text{obs}}(s, r, \omega)$. Although we use the time domain for the implementation, equations will be written in the frequency domain in the following to simplify calculations;
- the **data space** $\mathscr{D}$. We note $\mathrm{P}[c_0, \xi](s, \mathbf{x}, \omega)$ the wavefield calculated with the estimated reflectivity model ξ. Contrary to P^{obs}, P is known everywhere in the subsurface. We define the operator $\mathrm{M} : \mathscr{D} \mapsto \mathscr{D}_{\text{obs}}$ which, applied to a simulated wavefield $\mathrm{P}[c_0, \xi](s, \mathbf{x}, \omega)$, selects its values at receivers' locations. Hence MP can be compared to P^{obs} during iterative migration.

Table 2.1 summarises the notations used for the different spaces and their size. For more details on the choice of the data and model spaces, the reader is referred to Blazek *et al.* (2013) and Symes (2014). Note that with the introduction of the subsurface offset, the extended model $\mathscr{E}$ and the observed data space $\mathscr{D}_{\text{obs}}$ have the same dimension, which allows to alternate between these two spaces without losing information.

space	notation	size
model space	$\mathscr{M}$	$n_z \times n_x$
extended model space	$\mathscr{E}$	$n_z \times n_x \times n_h$
observed data space	$\mathscr{D}_{\text{obs}}$	$n_t \times n_s \times n_r$
data space	$\mathscr{D}$	$n_t \times n_s \times n_z \times n_x$

Table 2.1. – Data and model spaces involved in this study and size of their elements. n_z, n_x and n_h are the number of grid points considered in the extended model space. n_t is the number of time samples. n_s and n_r are the number of sources and receivers considered in the acquisition. In the data space, n_x can be restricted to a zone around the shot gather including at least all the receivers.

The inner product of two vectors ξ and χ of $\mathscr{E}$ is noted

$$\langle \xi \,|\, \chi \rangle_{\mathscr{E}} = \int_{\mathbf{x}} \int_{h} \xi(\mathbf{x}, h) \chi(\mathbf{x}, h) \, \mathrm{d}h \, \mathrm{d}\mathbf{x}, \tag{2.1}$$

and the associated norm is $\|\xi\|_{\mathscr{E}} = \sqrt{\langle \xi \,|\, \xi \rangle_{\mathscr{E}}}$. Similarly we define the inner-product $\langle \cdot \,|\, \cdot \rangle_{\mathscr{D}}$ and the norm $\|\cdot\|_{\mathscr{D}}$ in the data space. The norm of the data residuals R, defined as the difference between observed data and the value of simulated data at receivers' positions,

$$\mathrm{R}[c_0, \xi](s, r, \omega) = \Big[\mathrm{MP}[c_0, \xi]\Big](s, r, \omega) - \mathrm{P}^{\mathrm{obs}}(s, r, \omega), \tag{2.2}$$

is given by

$$\|\mathrm{R}\|_{\mathscr{D}_{\mathrm{obs}}} = \sqrt{\int_s \int_r \int_\omega \Big|\mathrm{R}(s, r, \omega)\Big|^2 \,\mathrm{d}\omega \,\mathrm{d}r \,\mathrm{d}s}. \tag{2.3}$$

Finally we call F^* and F^{-1} the adjoint and inverse operators of an operator F.

2.3. Modelling under the second-order Born approximation

We detail in this section how primaries and multiples are modelled with a second-order Born approximation in the extended domain. We first review the usual Born approximation for primary reflections in the non-extended case.

2.3.1. First-order Born approximation

In the constant-density acoustic approximation, wavefield propagation obeys the following wave-equation

$$\frac{(\mathrm{i}\omega)^2}{c(\mathbf{x})^2}\mathrm{P}(s, \mathbf{x}, \omega) - \Delta \mathrm{P}(s, \mathbf{x}, \omega) = \Omega(\omega)\delta(\mathbf{x} - s), \tag{2.4}$$

for a source located at $\mathbf{s} = (s, 0)$ with a wavelet $\Omega(\omega)$. The velocity model is noted $c(\mathbf{x})$.

The Born approximation is a linearisation of this wave-equation and relies on the scale separation presented in the introduction. We write

$$c(\mathbf{x}) = c_0(\mathbf{x}) + \delta c(\mathbf{x}), \tag{2.5}$$

where c_0 and δc stands for the background velocity model and the model perturbation respectively. Accordingly we separate the data P into $\mathrm{P} = \mathrm{P}_0 + \delta\mathrm{P}$. P_0 corresponds to the data modelled in the smooth background c_0 and $\delta\mathrm{P}$ is the perturbation caused by the model perturbation δc.

We suppose that the model perturbation is small with respect to the background velocity model, that is $\delta c \ll c_0$. Neglecting second-order term $\delta\mathrm{P}\delta c$, P_0 and $\delta\mathrm{P}$ are solution of

$$\begin{cases} \mathscr{L}_0 \mathrm{P}_0(s, \mathbf{x}, \omega) = \Omega(\omega)\delta(\mathbf{x} - s), & \text{(2.6a)} \\ \mathscr{L}_0 \delta\mathrm{P}(s, \mathbf{x}, \omega) = (\mathrm{i}\omega)^2 \xi(\mathbf{x}) \mathrm{P}_0(s, \mathbf{x}, \omega), & \text{(2.6b)} \end{cases}$$

where we have defined

- the wave operator $\mathscr{L}_0$ as

$$\mathscr{L}_0 = \frac{(\mathrm{i}\omega)^2}{c_0^2(\mathbf{x})} - \Delta. \tag{2.7}$$

  Note that $\mathscr{L}_0$ is defined with the smooth background velocity $c_0(\mathbf{x})$;

- the reflectivity $\xi(\mathbf{x})$ as

$$\xi(\mathbf{x}) = \frac{2\delta c(\mathbf{x})}{c_0^3(\mathbf{x})}. \tag{2.8}$$

The coefficient $2/c_0^3(\mathbf{x})$ is included in the definition of $\xi(\mathbf{x})$ to further simplify calculations, although the physical reflectivity property is described by the dimensionless quantity $\xi(\mathbf{x}) = \delta c(\mathbf{x})/(2c_0(\mathbf{x}))$. Note that the reflectivity defined in equation (2.8) is not extended yet. The case of the extended model will be presented in section 2.3.4.

The data perturbation δP corresponds to the reflected wave due to the reflectivity ξ. In all the following, it will be noted P_1.

The solutions of equation (2.6) involve Green's functions G_0 defined as the solution of the wave-equation with an impulsive source

$$\mathscr{L}_0 G_0(s, \mathbf{x}, \omega) = \delta(\mathbf{x} - s). \tag{2.9}$$

Then one can verify that

$$\begin{cases} P_0(s, \mathbf{x}, \omega) = G_0(s, \mathbf{x}, \omega)\Omega(\omega), & \text{(2.10a)} \\ P_1(s, \mathbf{x}, \omega) = \Omega(\omega) \displaystyle\int_y (i\omega)^2 G_0(\mathbf{s}, \mathbf{y}, \omega)\xi(\mathbf{y})G_0(\mathbf{y}, \mathbf{x}, \omega)\, d\mathbf{y}, & \text{(2.10b)} \end{cases}$$

are solutions of equations (2.6a) and (2.6b), respectively.

2.3.2. Reflection at the free surface

To consider first-order surface-related multiples, we need to model the reflection of upgoing primaries at the surface. The latter is known to act as a mirror and a first guess is to model multiples by applying the primary modelling workflow to a new areal source made of primaries recorded at the surface multiplied by a constant coefficient $R = -1$. Numerical comparison with a finite differences modelling and a free-surface condition suggests that an additional coefficient $2i\omega/c_0$ should be applied as well. We explain the physical meaning of this coefficient on a pure 1D case in appendix A and assume that it is valid in the general 2D case, too.

In the following, we note P_2 the wavefield resulting from the reflection of the primaries P_1 at the surface. The associated wave-equation is

$$\mathscr{L}_0 P_2(\mathbf{s}, \mathbf{x}, \omega) = M_s P_1(\mathbf{s}, \mathbf{x}, \omega), \tag{2.11}$$

where operator $M_s : \mathscr{D} \mapsto \mathscr{D}$ is defined by

$$M_s P_1(\mathbf{s}, \mathbf{x}, \omega) = (-1)\delta(\mathbf{x} - \mathbf{x}_{\text{surf}}) \frac{2i\omega}{c_0(\mathbf{x})} P_1(\mathbf{s}, \mathbf{x}, \omega), \tag{2.12}$$

and includes the additional reflection coefficient discussed in appendix A.

We test the accuracy of this approximation, in particular at non-zero offset, with the following numerical example. We consider a Ricker with a maximum frequency of 40 Hz as the

source wavelet (figure 2.1). The source and the receivers are buried in the subsurface assumed homogeneous with a velocity of 3000 m/s (figure 2.2a). We first simulate the propagation of the source with a classical second-order accurate space and time finite-difference scheme. Absorbing boundary conditions with Perfectly Matched Layer (PML) (Bérenger, 1994; Komatitsch and Martin, 2007) are implemented on each edge of the model except on the surface where a free-surface condition is implemented with the method of images (Levander, 1988; Robertsson, 1996) (figure 2.2c, left). Then we use the same code with PMLs on each edge of the model to solve equations (2.10a) and (2.11) (figure 2.2c, middle). The difference between the two results increases with offset (figure 2.2c, right). The amplitudes accuracy decays with offsets (figure 2.2b) as the reflection coefficient (-1) at the surface is set independent of the incident angle, but more importantly the traveltime remains accurate. Thus equation (2.11) stands as a good approximation to model the surface reflection.

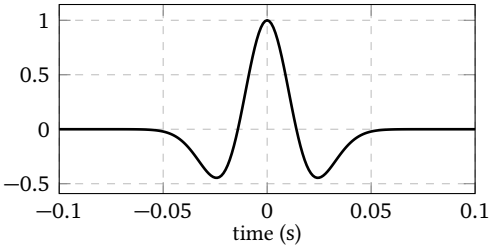


Figure 2.1. – The Ricker wavelet with maximum frequency of 40 Hz.

2.3.3. Second-order Born approximation

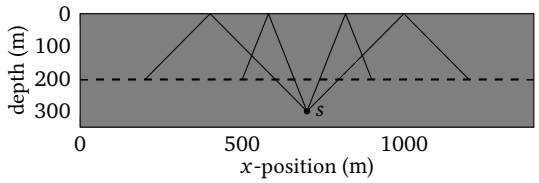
First-order surface multiples P_3 are modelled with the same equation as primary reflections P_1 (equation 2.10b), except that the incident wavefield is P_2 instead of P_0. Then, given a background velocity model $c_0(\mathbf{x})$ and a (non-extended) reflectivity model $\xi(\mathbf{x})$, data modelling is performed by solving four wave-equations

$$\begin{cases} \mathscr{L}_0 P_0(s,\mathbf{x},\omega) = \Omega(\omega)\delta(\mathbf{x}-s), & \text{(2.13a)} \\ \mathscr{L}_0 P_1(s,\mathbf{x},\omega) = (i\omega)^2 \xi(\mathbf{x}) P_0(s,\mathbf{x},\omega), & \text{(2.13b)} \\ \mathscr{L}_0 P_2(s,\mathbf{x},\omega) = M_s P_1(s,\mathbf{x},\omega) & \text{(2.13c)} \\ \mathscr{L}_0 P_3(s,\mathbf{x},\omega) = (i\omega)^2 \xi(\mathbf{x}) P_2(s,\mathbf{x},\omega), & \text{(2.13d)} \end{cases}$$

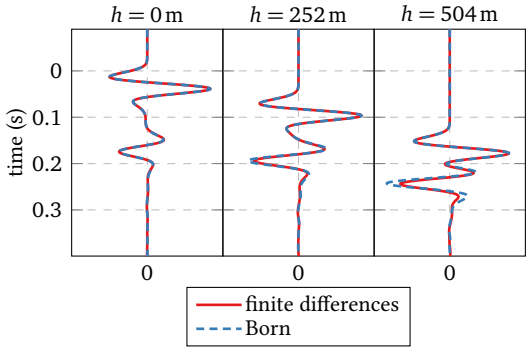
where M_s is defined in equation (2.12).

Comparison with a finite-difference modelling and a free-surface condition

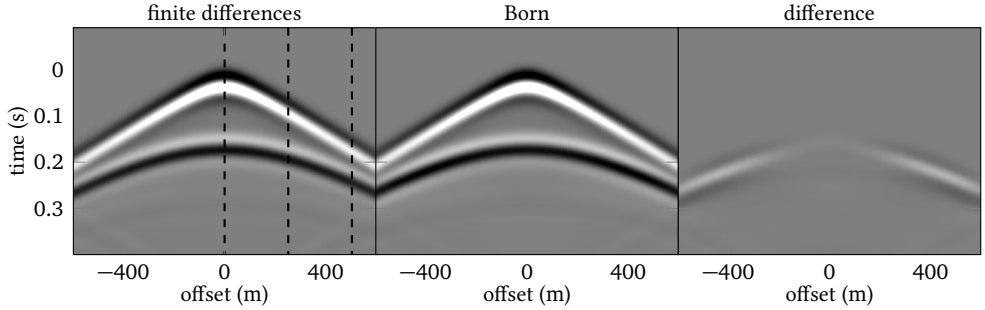
We have tested the validity of the Born approximation with respect to the free surface. In theory, the Born approximation is valid for small velocity contrasts. We want to study its behaviour in the case of high contrasts which may cause strong multiples. We compare data modelling with the second-order Born approximation described by equations (2.13) with a finite-difference modelling and a free-surface condition on four examples. We consider a background velocity of 1500 m/s and two values for the velocity perturbation: 1600 m/s and 2500 m/s, corresponding



(a) Position of the source (black dot) and the receivers (along the dashed lined).



(b) Traces extracted at the position indicated by the dashed lines in (c)



(c) Data recorded at receiver positions when the modelling is performed with finite differences (left) and with the Born approximation (centre). The right panel display their difference. The same colour scale is used for the three plots.

Figure 2.2. – Accuracy of the second-order Born approximation with respect to the free-surface reflection.

to a limited ($100/(1500+1600) = 0.03$) and a larger ($1000/(1500+1600) = 0.25$) velocity contrast. For each value, we consider two cases

- a horizontal reflective layer with homogeneous velocity below the interface located at $z_0 = 200\,\text{m}$ (figures 2.3 and 2.4). In this case, the background velocity model used for second-order Born modelling is obtained by smoothing the exact slowness model $\sigma(\mathbf{x})$ defined as $\sigma(\mathbf{x}) = 1/c_0(\mathbf{x})$ and the velocity perturbation is calculated as ;
- a horizontal "diffractive interface", with a velocity perturbation localised at depth z_0 and a homogeneous velocity ($1500\,\text{m/s}$) below and above the interface (figures 2.5 and 2.6).

The source and the receivers are located near the surface. Ghosts at the source and the receivers sides are created by the free-surface in the finite-difference simulation. This is accounted for in the Born modelling by a time shift in the source wavelet and in the computed data

$$\begin{cases} \Omega_{\text{ghost}}(t) = \Omega(t) - \Omega\Big(t + 2\dfrac{z_s}{c_0}\Big), & \text{(2.14a)} \\ \mathrm{P}_{\text{ghost}}(s,r,t) = \mathrm{P}(t) - \mathrm{P}\Big(t + 2\dfrac{z_r}{c_0}\Big). & \text{(2.14b)} \end{cases}$$

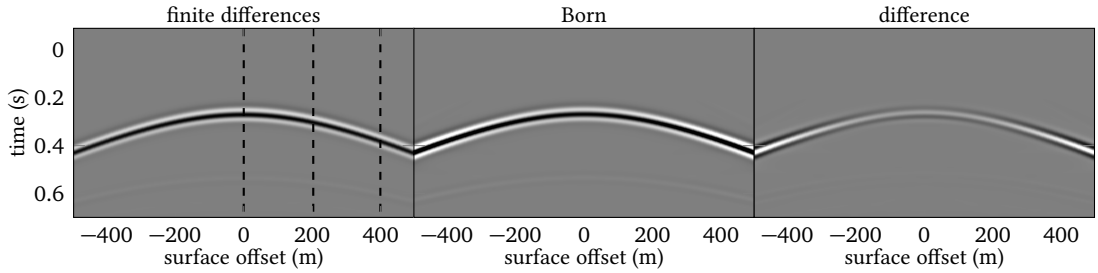
Once more, this approximation is valid for vertical wave propagation only.

Both in the reflection and diffraction cases, multiples are accurately modelled when the velocity contrast is small. In the diffraction case, the kinematics of multiples is respected even with a high velocity contrast, although the amplitudes are overestimated (figure 2.6). In the reflection case with a higher velocity contrast , the Born approximation does not model the kinematics of primaries as accurately, and it is worse in the case of multiples (figure 2.4). This is due to the propagation in a smooth model that modifies the velocities compared to the propagation in a blocky velocity model. Note however that the time shift between the two data sets is very similar at zero and far offset. Besides, there is not a unique choice for the definition of the smooth model. It is obtained in these examples by smoothing the blocky slowness model, but other definitions may lead to better results, at least at zero offset. In practice, there is no need to smooth an exact model: the MVA strategy aims actually at determining both a background velocity model and a reflectivity model allowing to reproduce observed data with a second-order Born approximation.

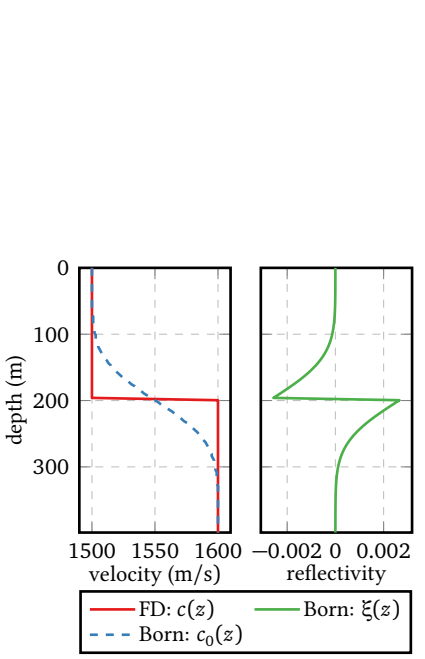
Eventually, we note that reflection multiples (figures 2.3 and 2.4) appear as a scaled version of the primaries with an opposite wavelet, contrary to diffraction multiples (figures 2.5 and 2.6) which have a different wavelet from the primaries.

Non-linearity of multiples with respect to the reflectivity model

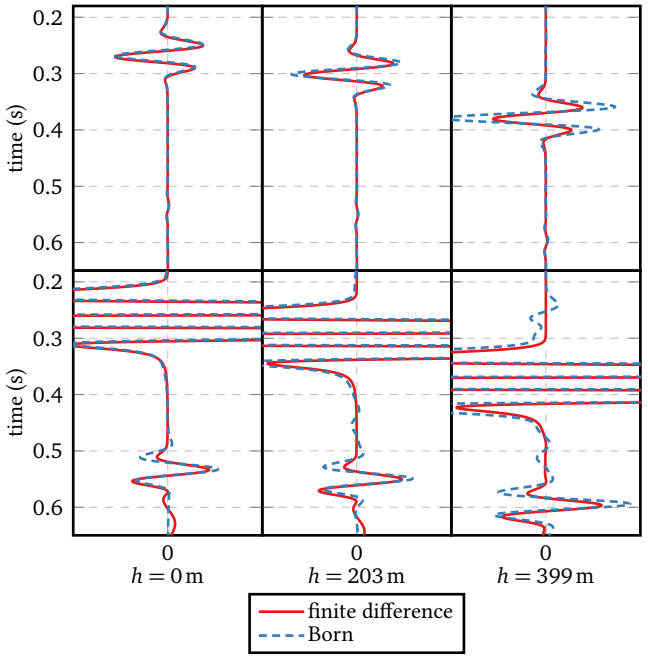
Finally, we illustrate the linear dependence of primaries and non-linear dependence of multiples to the reflectivity model in the second-order Born approximation with a simple example (figure 2.7). We consider two different reflectivity models ξ_A and ξ_B consisting of a single reflector located at 200 m depth for model A and 300 m for model B. A single shot is fired from the middle point in the surface. The source wavelet is a Ricker with a maximum frequency of 40 Hz (figure 2.1). We compute the response recorded by receivers at the surface, respectively $\mathrm{P}_1[\xi_A]$ and $\mathrm{P}_1[\xi_B]$ (figure 2.7, first and second columns). Then we compare the response obtained with the linear combination of models $\xi_C = \xi_A + 2\xi_B$ (figure 2.7, third column) with the linear combination of the individual data sets $\mathrm{P}_1[\xi_A] + 2\mathrm{P}_1[\xi_B]$ (figure 2.7, fourth column). In the



(a) Data recorded at receiver positions when the modelling is performed with finite differences (left) and with a free-surface condition and a second-order Born approximation (centre). The right panel display their difference. The same colour scale is used for the three plots.

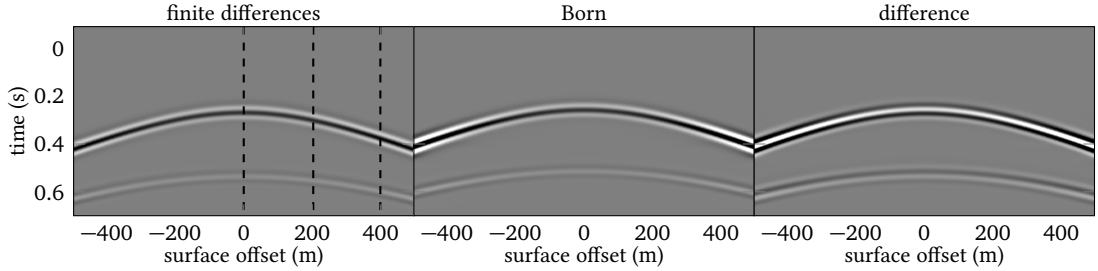


(b) 1D-models used for finite-differences and Born modelling.

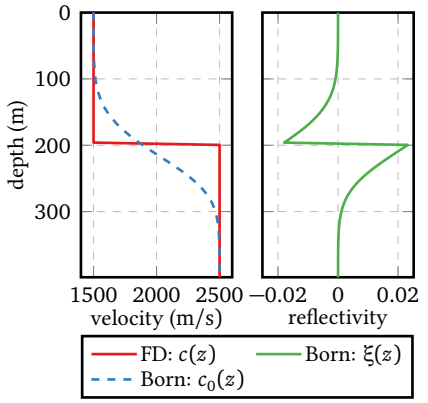


(c) Traces extracted at the positions indicated by the dashed lines in (a). The bottom row is a zoom of the top row allowing to compare multiples modelled with both approaches.

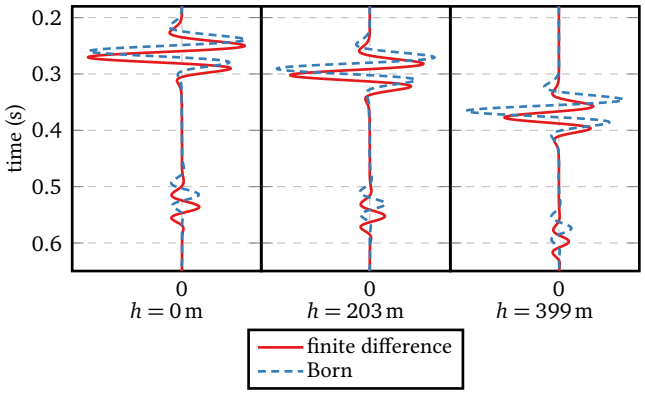
Figure 2.3. – Comparison of finite-differences modelling and free-surface modelling with a second-order Born approximation in the reflective case.



(a) Data recorded at receiver positions when the modelling is performed with finite differences (left) and with a free-surface condition and a second-order Born approximation (centre). The right panel display their difference. The same colour scale is used for the three plots.

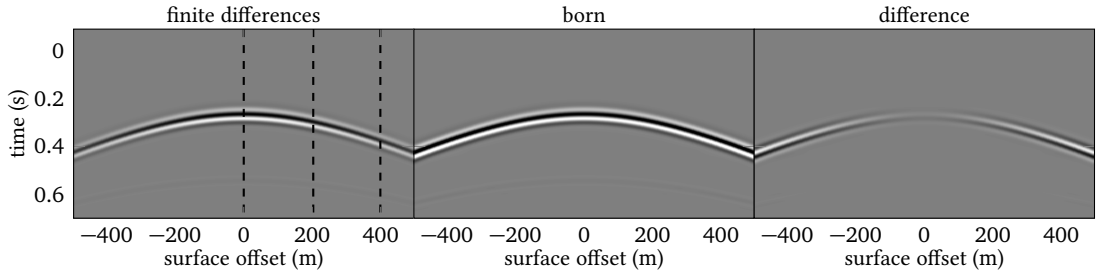


(b) 1D-models used for finite-differences and Born modelling.

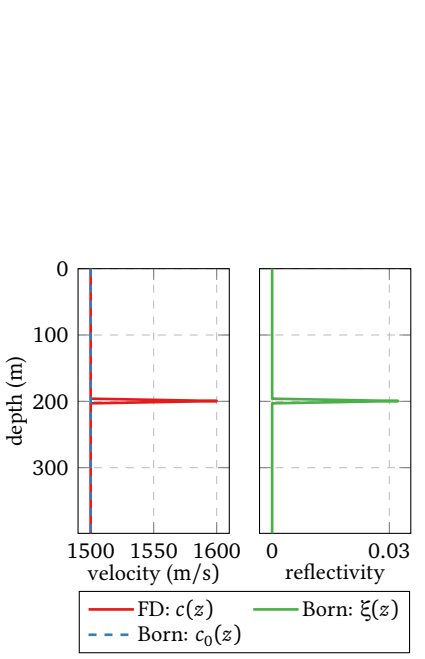


(c) Traces extracted at the positions indicated by the dashed lines in (a).

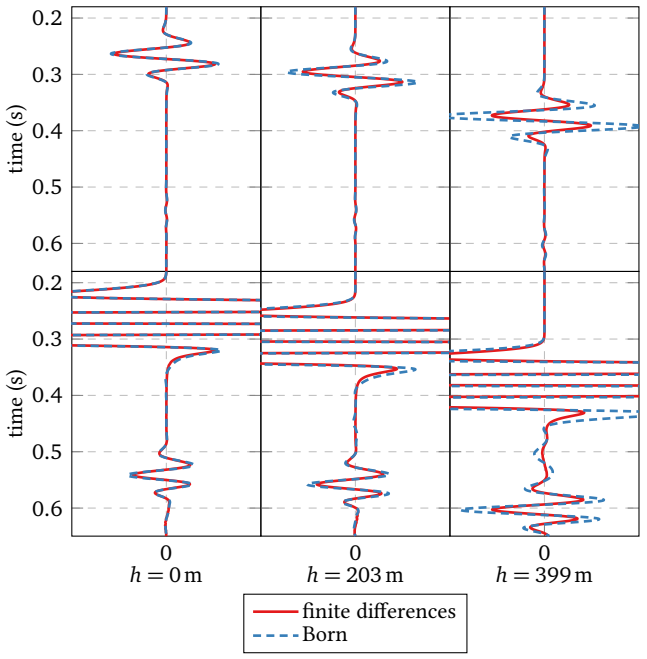
Figure 2.4. – Same as figure 2.3, but with a stronger velocity contrast.



(a) Data recorded at receiver positions when the modelling is performed with finite differences (left) and with a free-surface condition and a second-order Born approximation (centre). The right panel display their difference. The same colour scale is used for the three plots.

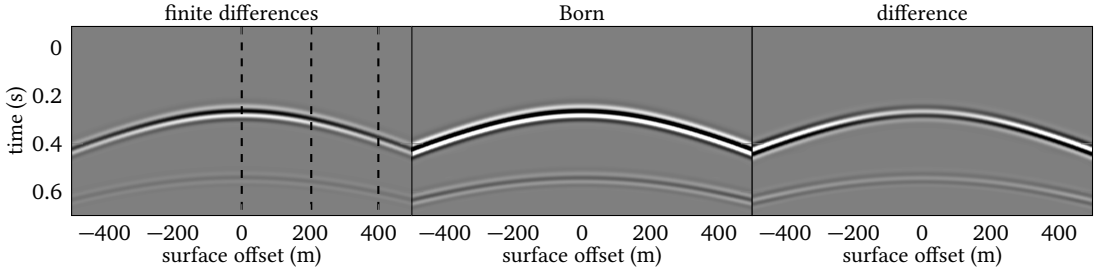


(b) 1D-models used for finite-differences and Born modelling.

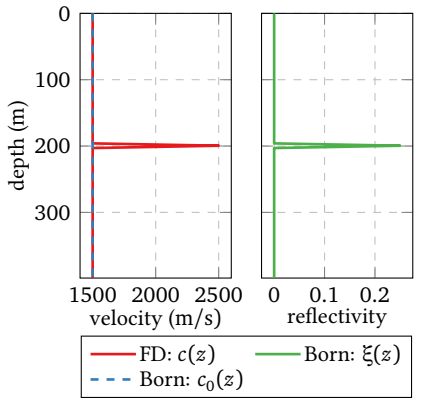


(c) Traces extracted at the positions indicated by the dashed lines in (a). The bottom row is a zoom of the top row allowing to compare multiples modelled with both approaches.

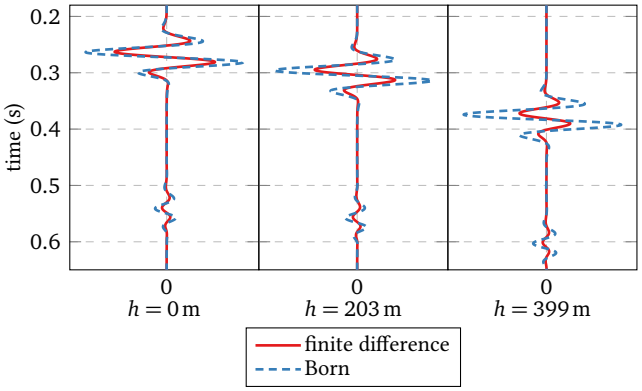
Figure 2.5. – Comparison of finite-differences modelling and free-surface modelling with a second-order Born approximation in the diffractive case.



(a) Data recorded at receiver positions when the modelling is performed with finite differences (left) and with a free-surface condition and a second-order Born approximation (centre). The right panel display their difference. The same colour scale is used for the three plots.



(b) 1D-models used for finite-differences and Born modelling.



(c) Traces extracted at the positions indicated by the dashed lines in (a).

Figure 2.6. – Same as figure 2.5, but with a stronger velocity contrast.

case of primaries only (figure 2.7, middle row) these data sets are similar, which confirms the linear relationship between velocity perturbation and primary reflections obtained under the first-order Born-approximation. In the case of multiples two remaining events are visible, the multiple which reflects twice on the deeper reflector and the one which reflects once at each reflector. As expected, the relationship between data and model perturbation is thus non linear under the second-order Born approximation.

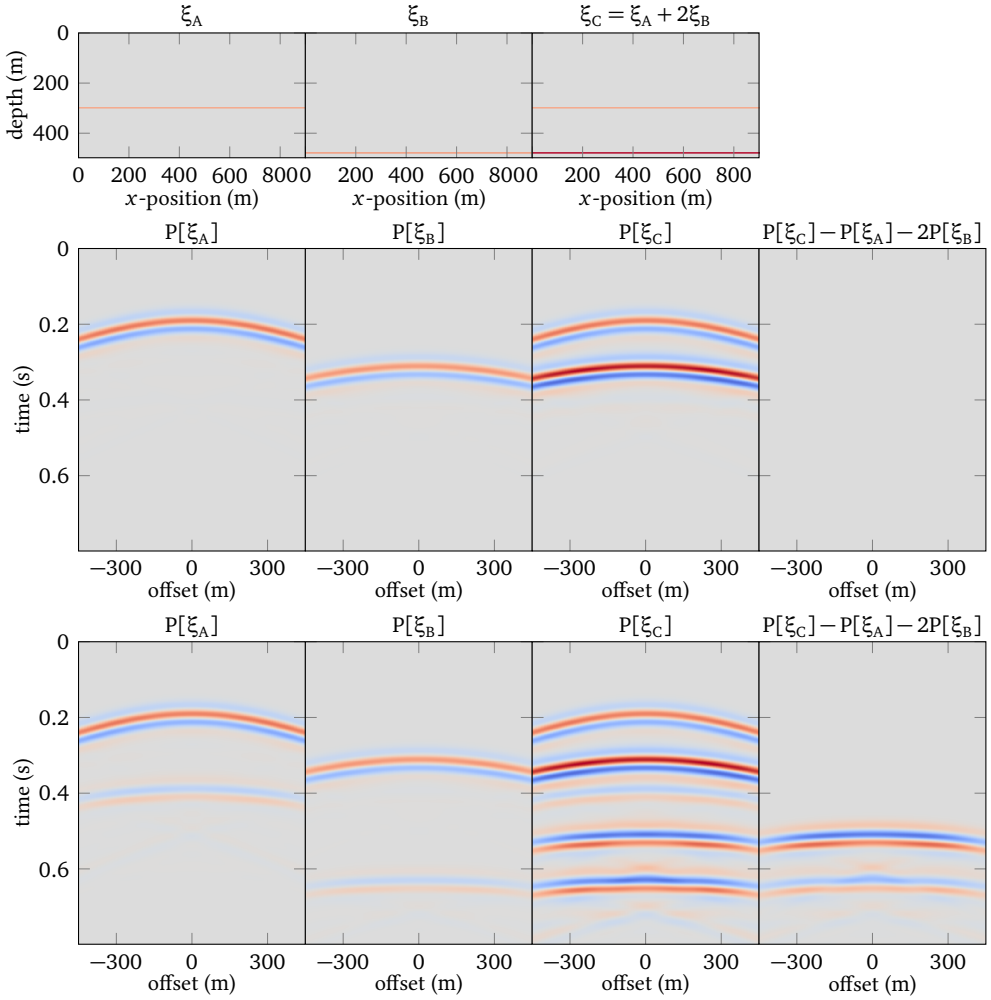


Figure 2.7. – Illustration of the linearity of primaries and non-linearity of multiples to the model perturbation. Two reflectivity models ξ_A and ξ_B and their combination ξ_C are considered (top row). Data are modelled in the case of primary reflections only (middle row) and in the case of primaries and first-order surface multiples (bottom row). The most right column shows the difference between data modelled in ξ_C and the combination of the data sets computed with ξ_A and ξ_B.

2.3.4. Introduction of the horizontal subsurface offset

We now extend the second-order Born modelling procedure to the extended domain. The space of physical models $\mathscr{M}$ is extended with the subsurface offset $\mathbf{h}$, considered strictly horizontal $\mathbf{h} = (h, 0)$. The extended reflectivity $\xi(\mathbf{x}, h)$ now depends on this extra parameter and lives in $\mathscr{E}$, and is defined as

$$\xi(\mathbf{x}, h) = \frac{2\delta c(\mathbf{x}, h)}{c_0^3(\mathbf{x})}. \tag{2.15}$$

The modelling of primary reflections now includes a spatial delay at depth as illustrated on figure 2.8. Equation (2.10b) defining P_1 is then transformed into

$$\mathrm{P}_1(s, \mathbf{x}, \omega) = (\mathrm{i}\omega)^2 \Omega(\omega) \int_{\mathbf{y}} \int_h \mathrm{G}_0(\mathbf{s}, \mathbf{y} - h, \omega) \xi(\mathbf{y}, h) \mathrm{G}_0(\mathbf{y} + h, \mathbf{x}, \omega) \, \mathrm{d}h \, \mathrm{d}\mathbf{y}, \tag{2.16}$$

with a new integral over h. In the case of first-order surface-related multiples, several definitions are possible (figure 2.9). We consider the case where a spatial delay is introduced at each reflection point (figure 2.9a), as it is a more general case and leads to similar formulations for the definition of P_1 and P_3. The modelling of first-order surface-related multiples now reads

$$\mathrm{P}_3(s, \mathbf{x}, \omega) = (\mathrm{i}\omega)^2 \int_{\mathbf{z}} \int_k \mathrm{G}_0(m, \mathbf{z} - k, \omega) \xi(\mathbf{z}, k) \\ \mathrm{G}_0(\mathbf{z} + k, \mathbf{x}, \omega) \, \mathrm{d}k \, \mathrm{d}\mathbf{z} \int_{m \in \partial\Omega} (-1) \frac{2\mathrm{i}\omega}{c_0(m)} \mathrm{P}_1(s, m, \omega) \, \mathrm{d}m, \tag{2.17}$$

where m is a point located at the surface $\partial\Omega$. As before, the coefficient $2\mathrm{i}\omega / c_0(\mathbf{x})$ is introduced to properly simulate the reflection at the free-surface.

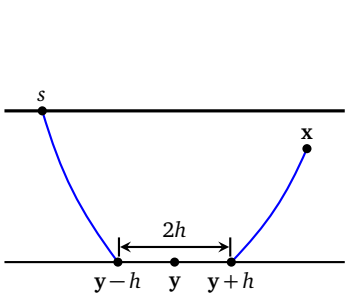


Figure 2.8. – Modelling of primary reflections with the horizontal subsurface-offset h defined as a spatial delay at depth.

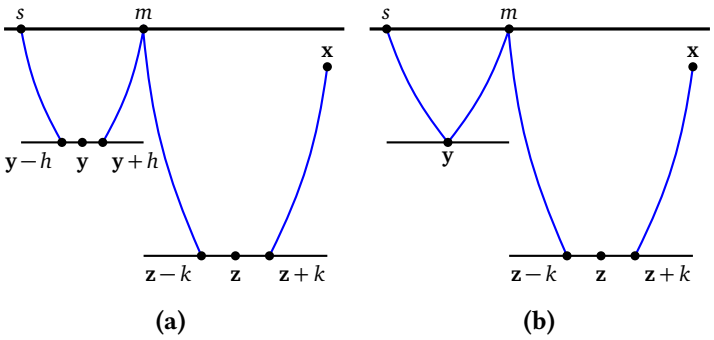


Figure 2.9. – Two possible schemes for the modelling of first-order surface-related multiples with the horizontal subsurface-offset. A spatial delay is introduced either at each subsurface point (a) or at a single image point (b).

The implementation of equations (2.16) and (2.17) in their current formulation is not straightforward. Applying the wave-operator $\mathscr{L}_0$ to this formula allows to derive a more convenient

modelling scheme:

$$\begin{cases} \mathscr{L}_0 \mathrm{P}_0(s,\mathbf{x},\omega) = \mathrm{K}_{\mathrm{P}_0}(s,\mathbf{x},\omega) = \Omega(\omega)\delta(\mathbf{x}-s), & (2.18a) \\ \mathscr{L}_0 \mathrm{P}_1(s,\mathbf{x},\omega) = \mathrm{K}_{\mathrm{P}_1}(s,\mathbf{x},\omega) = \displaystyle\int_h (\mathrm{i}\omega)^2 \mathrm{P}_0(s,\mathbf{x}-2h,\omega)\xi(\mathbf{x}-h,h)\,\mathrm{d}h, & (2.18b) \\ \mathscr{L}_0 \mathrm{P}_2(s,\mathbf{x},\omega) = \mathrm{K}_{\mathrm{P}_2}(s,\mathbf{x},\omega) = \mathrm{M}_s \mathrm{P}_1(s,\mathbf{x},\omega) & (2.18c) \\ \mathscr{L}_0 \mathrm{P}_3(s,\mathbf{x},\omega) = \mathrm{K}_{\mathrm{P}_3}(s,\mathbf{x},\omega) = \displaystyle\int_h (\mathrm{i}\omega)^2 \mathrm{P}_2(s,\mathbf{x}-2h,\omega)\xi(\mathbf{x}-h,h)\,\mathrm{d}h, & (2.18d) \end{cases}$$

which involves the usual wave equation with modified source terms $\mathrm{K}_{\mathrm{P}_1}$ and $\mathrm{K}_{\mathrm{P}_3}$ that include the integral over the subsurface offset. Thus the introduction of the subsurface offset does not change the wave-equation to be solved, but only the source terms $\mathrm{K}_{\mathrm{P}_1}$ and $\mathrm{K}_{\mathrm{P}_3}$. Note that there is no additional loop over h, except in the source terms. The number of wave-equations to be solved remains the same.

To simplify further expressions, we introduce the vector $\mathrm{S} \in \mathscr{D}$ which is zero everywhere in space except at the location of the seismic source: $\mathrm{S}(s,\mathbf{x},\omega) = \Omega(\omega)\delta(\mathbf{x}-s)$. We also define the bilinear operator $\mathrm{K}^-[u,\chi] : \mathscr{D} \times \mathscr{E} \mapsto \mathscr{D}$ for a vector u of $\mathscr{D}$ and a vector χ of $\mathscr{E}$ as

$$\mathrm{K}^-[u,\chi](s,\mathbf{x},\omega) = \int_h (\mathrm{i}\omega)^2 u(s,\mathbf{x}-2h,\omega)\chi(\mathbf{x}-h,h)\,\mathrm{d}h. \tag{2.19}$$

This allows the modelling scheme to be written in a more compact way, which will be used in following calculations,

$$\begin{cases} \mathscr{L}_0 \mathrm{P}_0(s,\mathbf{x},\omega) = \mathrm{S}(s,\mathbf{x},\omega), & (2.20a) \\ \mathscr{L}_0 \mathrm{P}_1(s,\mathbf{x},\omega) = \mathrm{K}^-[\mathrm{P}_0,\xi](s,\mathbf{x},\omega), & (2.20b) \\ \mathscr{L}_0 \mathrm{P}_2(s,\mathbf{x},\omega) = \mathrm{M}_s \mathrm{P}_1(s,\mathbf{x},\omega), & (2.20c) \\ \mathscr{L}_0 \mathrm{P}_3(s,\mathbf{x},\omega) = \mathrm{K}^-[\mathrm{P}_2,\xi](s,\mathbf{x},\omega). & (2.20d) \end{cases}$$

We have described how to solve the forward problem of the inner inverse problem. We now introduce the corresponding objective function and the resolution of the inverse problem, consisting of determining ξ from observed data for a given background velocity model $c_0(\mathbf{x})$.

2.4. Optimisation strategy

Iterative migration is an inverse problem posed in the extended domain $\mathscr{E}$. This is the inverse problem we consider in this chapter. It is solved as an unconstrained optimisation problem. In this section we present the objective function to be minimised and the optimisation strategy which will be used in numerical examples. We eventually explain how the gradient of the objective function is computed.

2.4.1. Definition of an objective function for iterative migration

The objective function J_0 of iterative migration reads

$$\mathrm{J}_0[c_0,\xi] = \mathrm{J}_{\mathrm{Migr}}[c_0,\xi] + a_\varphi \varphi[\xi]. \tag{2.21}$$

Iterative migration aims at determining an extended reflectivity model ξ minimising J_0. The estimated background velocity c_0, which might be inaccurate, is kept fixed during the minimisation of J_0. The objective function J_0 is made of two contributions,

- the first term evaluates the misfit between data computed with the reflectivity model ξ and observed data. It is expressed as

$$J_{\text{Migr}}[c_0, \xi] = \frac{1}{2}\left\| M\big(P_1[c_0, \xi] + P_3[c_0, \xi]\big) - P^{\text{obs}} \right\|^2_{\mathscr{D}_{\text{obs}}} \tag{2.22}$$

  The projection operator $M : \mathscr{D} \mapsto \mathscr{D}_{\text{obs}}$ selects the value of the wavefields P_1 and P_3 at the position of receivers before comparison with observed data;

- the second term is a regularisation function whose weight is controlled by the scalar α_φ. In this chapter, we will consider simple regularisation functions such as the ℓ_2 norm $\varphi[\xi] = \frac{1}{2}\|\xi\|^2_{\mathscr{E}}$, but other terms may easily be introduced.

2.4.2. Linear case of primaries only

In the numerical example section, we begin by applying iterative migration in a simple case without multiples. The first-order Born modelling can be represented by a linear operator $F : \mathscr{E} \mapsto \mathscr{D}_{\text{obs}}$. To simplify the following equations, we assume that the projection operator M is included in the definition of F. With the ℓ_2 norm as regularisation, the objective function (2.21) now reads

$$J_0[c_0, \xi] = \frac{1}{2}\left\| F[c_0]\xi - P^{\text{obs}} \right\|^2_{\mathscr{D}_{\text{obs}}} + \alpha_\varphi \varphi[\xi]. \tag{2.23}$$

Its minimisation is equivalent to the resolution of the linear system

$$\big(F^T F + \alpha_\varphi I\big)\xi = F^T P^{\text{obs}}, \tag{2.24}$$

which is obtained by zeroing the gradient of the objective function (2.23). An efficient method to solve this system is the linear conjugate gradient algorithm (Nocedal and Wright, 2006, p. 112). Note that in practice an F and F^T will not be expliclty computed and an equivalent formulation will be used (section 2.5.1).

2.4.3. Non-linear optimisation in the non-linear case of multiple reflections

With multiples, the modelling operator is not linear any more and non-linear optimisation techniques have to be used. With gradient-based methods, the reflectivity model at iteration $(n+1)$ is updated from the reflectivity model at iteration (n) following

$$\xi^{(n+1)} = \xi^{(n)} + \alpha^{(n)} d^{(n)}, \tag{2.25}$$

where $d \in \mathscr{E}$ is called *descent direction* and the positive scalar α is called *step size* or *step length* determined by a procedure called *linesearch*. The process is initialised with $\xi^{(1)} = 0$ and ξ is updated until a convergence criterion is satisfied, for example when the value of the objective function or the norm of its gradient goes below a given threshold. In the numeric examples

shown at the end of this chapter, we set a maximum number of iterations N, so that the final result is $\xi^{(N+1)}$. d and α are determined such that the value of J_0 decreases at each iteration, that is $J_0(\xi^{(n+1)}) < J_0(\xi^{(n)})$.

The most simple choice for the descent direction is given by the opposite of the gradient of the objective function, that is $d^{(n)} = -g^{(n)}$, with $g^{(n)}$ the gradient of J_0 computed at $\xi^{(n)}$,

$$g^{(n)} = \frac{\partial J_0}{\partial \xi}[c_0, \xi^{(n)}], \tag{2.26}$$

but alternative strategies providing faster convergence exist. The optimal strategy is a trade-off between the number of iteration needed to reach a satisfactory minimisation of the objective function and the numerical cost for determining d and α. We briefly review standard non-linear optimisation strategies in appendix B. The reader is referred to Nocedal and Wright (2006) for an extensive review.

In all the following numerical examples, we use the linesearch procedure of Moré and Thuente (1994), which ensures that the strong Wolfe conditions (see section B.2 and figure B.1) are satisfied. The procedure requires the evaluation of the value and the gradient of the objective function at successive trial points $\xi^{(n)} + \alpha_i d^{(n)}$, until a satisfactory value for α_i is found. This may be expensive if numerous step size have to be tested. However we observed in practice that this optimal value $\alpha^{(n)}_{\text{opt}}$ is usually found in two or three trial steps. Besides as we set $\xi^{(n+1)} = \xi^{(n)} + \alpha^{(n)}_{\text{opt}} d^{(n)}$, the linesearch procedure already provides the value of $J_0(\xi^{(n+1)})$ and $g^{(n+1)}$ at the next iteration. So the additional cost of a linesearch procedure satisfying the strong Wolfe conditions remains affordable.

We use the non-linear conjugate-gradient technique to determine the descent direction,

$$d^{(n)} = -g^{(n)} + \beta^{(n)} d^{(n-1)}, \tag{2.27}$$

where the scalar β is given by the formula Polak-Ribière formula

$$\beta^{(n)} = \frac{\left\langle g^{(n)} \middle| g^{(n)} - g^{(n-1)} \right\rangle_{\mathscr{E}}}{\left\langle g^{(n-1)} \middle| g^{(n-1)} \right\rangle_{\mathscr{E}}} \tag{2.28}$$

A step length satisfying the strong Wolfe conditions ensures that d is effectively a descent direction (Nocedal and Wright, 2006, p. 122).

2.5. Derivatives of J_0

One of the key ingredients of local optimisation methods is the gradient of the objective function. Its computation is the most computationally expensive step of the optimisation procedure and we present here an efficient way of deriving the gradient of J_0 based on the adjoint state method. We illustrate on an example the different contributions of the gradient to the reflectivity update. Eventually, the shape of the Hessian matrix is studied on a simple 1D-case.

2.5.1. Computation of the gradient of J_0

In the objective function defined in equation (2.21), the contribution to the gradient due to the regularisation is straightforward. When the regularisation term is defined as the ℓ_2-norm of ξ,

we have

$$\varphi[\xi] = \frac{1}{2}\|\xi\|^2_{\mathscr{E}}; \tag{2.29}$$

$$\frac{\partial \varphi}{\partial \xi}[\xi] = \xi. \tag{2.30}$$

In this section, we focus on the gradient of $\mathrm{J}_{\mathrm{Migr}}$. It can be expressed as

$$\frac{\partial \mathrm{J}_{\mathrm{Migr}}}{\partial \xi} = \left[\frac{\partial \mathrm{P}_1}{\partial \xi}[c_0, \xi^{(n)}] + \frac{\partial \mathrm{P}_3}{\partial \xi}(c_0, \xi^{(n)})\right]^{\mathrm{T}} \mathrm{M}^{\mathrm{T}}\Big(\mathrm{M}(\mathrm{P}_1[c_0, \xi^{(n)}] + \mathrm{P}_3[c_0, \xi^{(n)}]) - \mathrm{P}^{\mathrm{obs}}\Big). \tag{2.31}$$

The quantities $\partial \mathrm{P}_1/\partial \xi$ and $\partial \mathrm{P}_3/\partial \xi$ are called Fréchet derivatives and correspond to the partial derivatives of specific data with respect to model parameters. Their computation is expensive, and we prefer to use a more efficient gradient computation method, the adjoint state method (Plessix, 2006). This method will directly compute the product of the Fréchet derivatives with the residuals rather than explicitly deriving the two terms.

An augmented functional called Lagrangian is defined with constraints added to the objective function. Each *state variable* P_0, P_1, P_2 and P_3 is associated to an *adjoint variable* λ_0, λ_1, λ_2 and λ_3. The λ_i are vectors of $\mathscr{D}$ and are used as Lagrange multipliers of the state equations (2.20)

$$\begin{aligned}\overline{\mathrm{J}_{\mathrm{Migr}}}(\xi^{(n)}, \mathrm{P}_0, \mathrm{P}_1, \mathrm{P}_2, \mathrm{P}_3, \lambda_0, \lambda_1, \lambda_2, \lambda_3) = \mathrm{J}_{\mathrm{Migr}}(\xi) &- \big\langle \lambda_0 \,\big|\, \mathscr{L}_0 \mathrm{P}_0 - \mathrm{S}\big\rangle_{\mathscr{D}} \\ &- \big\langle \lambda_1 \,\big|\, \mathscr{L}_0 \mathrm{P}_1 - \mathrm{K}^{-}[\mathrm{P}_0, \xi]\big\rangle_{\mathscr{D}} \\ &- \big\langle \lambda_2 \,\big|\, \mathscr{L}_0 \mathrm{P}_2 - \mathrm{M}_s \mathrm{P}_1\big\rangle_{\mathscr{D}} \\ &- \big\langle \lambda_3 \,\big|\, \mathscr{L}_0 \mathrm{P}_3 - \mathrm{K}^{-}[\mathrm{P}_2, \xi]\big\rangle_{\mathscr{D}},\end{aligned} \tag{2.32}$$

where we have dropped the dependencies to $(s, \mathbf{x}, \omega)$ for simplicity. The derivative of the Lagrangian with respect to ξ can be written

$$\frac{\mathrm{d}\overline{\mathrm{J}_{\mathrm{Migr}}}}{\mathrm{d}\xi} = \frac{\partial \overline{\mathrm{J}_{\mathrm{Migr}}}}{\partial \xi} + \sum_{i=0}^{3} \frac{\partial \mathrm{P}_i}{\partial \xi} \frac{\partial \overline{\mathrm{J}_{\mathrm{Migr}}}}{\partial \mathrm{P}_i} + \sum_{i=0}^{3} \frac{\partial \lambda_i}{\partial \xi} \frac{\partial \overline{\mathrm{J}_{\mathrm{Migr}}}}{\partial \lambda_i}, \tag{2.33}$$

where the overlined variable $\overline{\mathrm{J}_{\mathrm{Migr}}}$ refers to the extended Lagrangian (2.32). We decide to set the derivatives $\partial \overline{\mathrm{J}_{\mathrm{Migr}}}/\partial \lambda_i$ to zero, which ensures that the P_i are solutions of the state equations (2.20). Similarly we set the derivatives $\partial \overline{\mathrm{J}_{\mathrm{Migr}}}/\partial \mathrm{P}_i$ to zero. This way, the computation of the Fréchet derivatives is not required and we obtain the *adjoint equations* defining the adjoint variables λ_i

$$\begin{cases} \mathscr{L}_0^* \lambda_3 = \mathrm{M}^{\mathrm{T}}\big(\mathrm{M}(\mathrm{P}_1 + \mathrm{P}_3) - \mathrm{P}^{\mathrm{obs}}\big) & \text{(2.34a)} \\ \mathscr{L}_0^* \lambda_2 = \mathrm{K}^{+}[\lambda_3, \xi] & \text{(2.34b)} \\ \mathscr{L}_0^* \lambda_1 = \mathrm{M}^{\mathrm{T}}\big(\mathrm{M}(\mathrm{P}_1 + \mathrm{P}_3) - \mathrm{P}^{\mathrm{obs}}\big) + \mathrm{M}_s^* \lambda_2 & \text{(2.34c)} \\ \mathscr{L}_0^* \lambda_0 = \mathrm{K}^{+}[\lambda_1, \xi], & \text{(2.34d)} \end{cases}$$

where operator $\mathrm{K}^{+}[u, \chi] : \mathscr{D} \times \mathscr{E} \mapsto \mathscr{D}$ is defined for $u \in \mathscr{D}$ and $\chi \in \mathscr{E}$ by

$$\mathrm{K}^{+}[u, \chi](s, \mathbf{x}, \omega) = \int_h (\mathrm{i}\omega)^2 u(s, \mathbf{x} + 2h, \omega) \chi(\mathbf{x} + h, h)\, \mathrm{d}h, \tag{2.35}$$

and can is an adjoint operator of K^- (equation 2.19):

$$\left\langle u \,\middle|\, \mathrm{K}^-[v,\chi] \right\rangle_{\mathscr{D}} = \left\langle \mathrm{K}^+[u,\chi] \,\middle|\, v \right\rangle_{\mathscr{D}} \tag{2.36}$$

for $(u,v) \in \mathscr{D}$ and $\chi \in \mathscr{E}$.

Eventually the gradient is given by the partial derivative of the Lagrangian with respect to ξ:

$$\frac{\partial \mathrm{J}_{\mathrm{Migr}}}{\partial \xi} = \frac{\mathrm{d}\overline{\mathrm{J}_{\mathrm{Migr}}}}{\mathrm{d}\xi} = \frac{\partial \overline{\mathrm{J}_{\mathrm{Migr}}}}{\partial \xi} = \mathrm{Q}[\mathrm{P}_0, \lambda_1] + \mathrm{Q}[\mathrm{P}_2, \lambda_3] \tag{2.37}$$

where operator $\mathrm{Q}[u,v] : \mathscr{D} \times \mathscr{D} \mapsto \mathscr{E}$ is an extended cross-correlation defined for $(u,v) \in \mathscr{D}$ by

$$\mathrm{Q}[u,v](\mathbf{x},h) = \int_s \int_\omega (\mathrm{i}\omega)^2 u^*(s,\mathbf{x}-h,\omega) v(s,\mathbf{x}+h,\omega)\,\mathrm{d}\omega\,\mathrm{d}s. \tag{2.38}$$

It is related to K^- and K^+ by

$$\left\langle \mathrm{Q}[u,v] \,\middle|\, \chi \right\rangle_{\mathscr{E}} = \left\langle \mathrm{K}^-[u,\chi] \,\middle|\, v \right\rangle_{\mathscr{D}} = \left\langle u \,\middle|\, \mathrm{K}^+[v,\chi] \right\rangle_{\mathscr{D}}. \tag{2.39}$$

The adjoint variables are solutions of equations similar to the state variables. They involve the adjoint of the wave operator, corresponding to a back-propagation in the time domain. Note that the computation of λ_0 is not necessary and the computation of the gradient is less then twice as costly as the computation of the value of the objective function (table 2.2).

	primaries only		primaries and multiples	
evaluate J_0	2	$(\mathrm{P}_0, \mathrm{P}_1)$	4	$(\mathrm{P}_0, \mathrm{P}_1, \mathrm{P}_2, \mathrm{P}_3)$
evaluate J_0 and $\frac{\partial \mathrm{J}_0}{\partial \xi}$	3	$(\mathrm{P}_0, \mathrm{P}_1, \lambda_1)$	7	$(\mathrm{P}_0, \mathrm{P}_1, \mathrm{P}_2, \mathrm{P}_3, \lambda_3, \lambda_2, \lambda_1)$

Table 2.2. – Number of wave-equations to be solved to evaluate the value and the gradient of the objective function at a point $\xi \in \mathscr{E}$. The number indicated here should be multiplied by the number of sources position considered in the acquisition.

We now explain the physical meaning of the different parts of the gradient.

2.5.2. Interpretation of the gradient

Noting that the source terms of the equations defining λ_1 and λ_3 (equations 2.34a and 2.34c) have a common term, we can split the adjoint variable λ_1 into two contributions

$$\lambda_1 = \lambda_3 + \lambda_4, \tag{2.40}$$

with λ_4 solution of

$$\mathscr{L}_0^* \lambda_4 = \mathrm{M}_s^* \lambda_2. \tag{2.41}$$

so that the gradient of J_{Migr} has three contributions

$$\frac{\partial J_{\text{Migr}}}{\partial \xi} = \underbrace{Q[P_0, \lambda_3]}_{g_1} + \underbrace{Q[P_0, \lambda_4]}_{g_2} + \underbrace{Q[P_2, \lambda_3]}_{g_3}. \tag{2.42}$$

To illustrate the contributions of each term to the gradient, we consider a simple example. A single horizontal reflector is located at 600 m in a homogeneous velocity model (3000 m/s) (figure 2.10a). In the trial reflectivity ξ^t, a reflector is located at depth 400 m (figure 2.10b). The modelling of both observed data (figure 2.10c) and calculated data is performed with the correct velocity model. Four events are visible in the residuals (figure 2.10d) defined as the difference between data calculated with the trial reflectivity and observed data. Each part of the gradient convert each event to the model space.

- the first contribution (figure 2.11a) noted g_1 in equation (2.42) interprets the four events in the residuals as primaries, resulting in four events in the gradient. Among them, only one corresponds to the true reflector located at 600 m depth. The three other events are cross-talk artefacts.

- the second and third contributions (figures 2.11b and 2.11c) look very similar. They both interpret the four events in the residuals as multiples. The second term g_2 images a multiple on the source side, while the third term g_3 images a multiple on the receiver side. They can be understood as follows: among the two reflections defining the multiples, one is produced by the trial reflectivity ξ^t (figure 2.10b) and the gradient tries to position the second reflector which allows to reconstruct the events in the data. In our example, the interpretation of the two primaries as multiples results in two other cross-talk artefacts. Finally the interpretation of multiples as multiples leads to two other events, but none of them is positioned at the depth of the reflector, because they are deduced from assumption that a first reflection occurred in ξ^t which is an incorrect model.

This example actually considers an extreme case, where the second and third part of the gradient do not bring any useful information and only mislead the reflectivity update. In practice, the inversion starts at $\xi = 0$, where only the first part of the gradient is non-zero. It results in two kinds of events: cross-talk artefacts and reflectors whose positions are kinematically consistent with the primaries in the data. The latter allow a correct interpretation of multiples at following iterations by the second and third term in the gradient.

We have presented optimisation strategies for iterative migration and an efficient way to compute the gradient of the associated objective function. We now present applications on a simple synthetic example.

2.6. Synthetic examples

2.6.1. Primaries only in a homogeneous medium

In the first example, we consider a 450 m depth and 1620 m large model discretised along a 6 m × 6 m grid. The reflector is 300 m deep and the exact background velocity model is

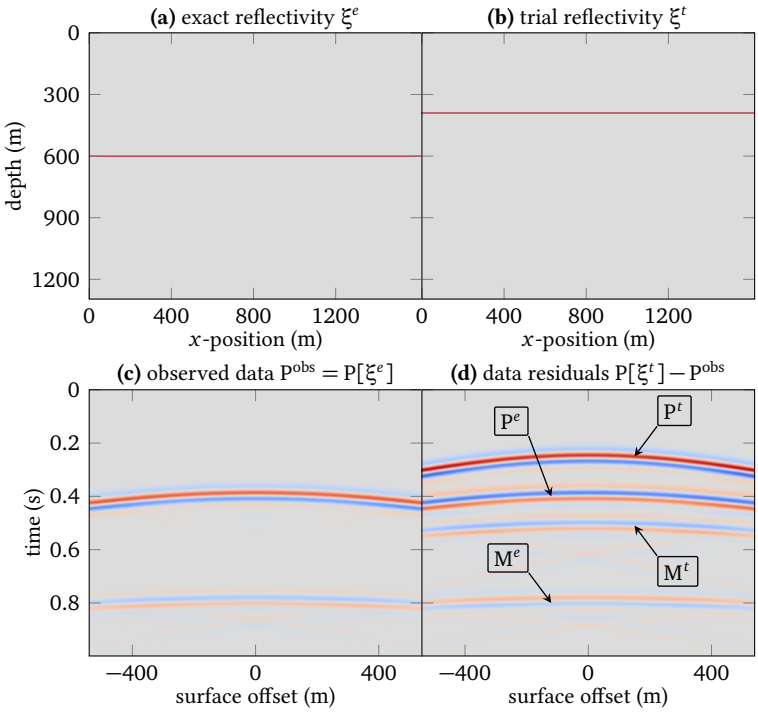


Figure 2.10. – Model constructed to interpret the three parts of the gradients of J_0 as defined in equation (2.42). (a) Exact reflectivity model ξ^e; (c) Corresponding observed data P^{obs}; (b) Trial reflectivity model ξ^t; (d) Residuals ($P[\xi^t] - P^{obs}$). P^e and M^e are the primary and multiple corresponding to the true reflector, P^t and M^t the primary and multiple corresponding to the trial reflectivity.

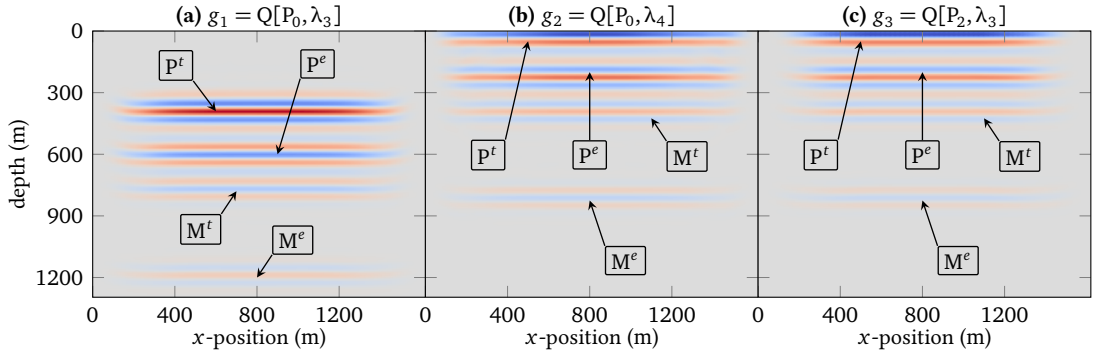


Figure 2.11. – Section at $h = 0\,\text{m}$ of the three contributions of the gradient (equation 2.42) computed at the point ξ^t as defined in figure 2.10. (a) $g_1 = Q[P_0, \lambda_3]$ is the interpretation of the events in the residuals as primary reflections; (b) $g_2 = Q[P_0, \lambda_4]$ and (c) $g_3 = Q[P_2, \lambda_3]$ are the interpretation of the same events as multiple reflections. The labels indicate the corresponding events in the data residual (figure 2.10d).

3000 m/s. A shot is fired every 24 m and the maximum surface offset is ±540 m. Then a single reflected event is visible on observed data (figure 2.12a). For the optimisation process, we use the ℓ_2-norm as regularisation function and the linear conjugate-gradient algorithm. 10 iterations are performed in a too slow background velocity model (2500 m/s) (figure 2.13).

Two sections of the extended reflectivity volume are shown: one at $h = 0$ m corresponding to the physical reflectivity (figure 2.13, left), and a CIG extracted at $x = 810$ m (figure 2.13, middle). The most right panel displays the same CIG multiplied by the absolute value of the subsurface offset $|h|$, the actual input of the MVA objective function. After one iteration, two main events are visible with opposite curvatures, indicated by the dashed lines for the first iteration (figure 2.13, top). Because of the too low velocity, the reflector is shifted toward the surface in the zero-offset section and spreads over non-zero offsets with a downward curve, as predicted by the theory (Mulder, 2014). The event curved upward above the reflector is a migration artefact due to the limited extension of sources and receivers in the acquisition. It is especially visible in the panels where $|h|\xi^{(n)}$ is displayed (figure 2.13, right). The data reconstructed with this reflectivity model does not match observed data (figures 2.12b and 2.12c), as only one iteration is performed.

The residuals are greatly reduced with iterations (figure 2.14) and the final reflectivity model obtained after 10 iterations perfectly explains observed data (figures 2.12h and 2.12i). Two improvements are visible on the final reflectivity image figure 2.13. Deconvolution of the wavelet source results in a more localised reflector, but the main effect of iterations is the attenuation of migration artefacts and the strengthening of the event corresponding to the true reflector.

We run the same experiment for the correct velocity model (3000 m/s) and for a too high velocity model (3500 m/s). This test is similar to those presented in Lameloise *et al.* (2014) and Hou and Symes (2016b). The CIGs obtained after a single iteration are all affected by migration artefacts with an upward curvature (figure 2.15, top), although in the case of a too high migration velocity, these artefacts look similar to the event corresponding to the true reflector and are not distinguishable. After 10 iterations all CIGs look cleaner, and defocusing of energy is only due to errors in the velocity model.

As a result, attenuation of migration artefacts in the CIGs improves the properties of the MVA objective function J_1. To illustrate this, we plot in figure 2.16 the value of J_1 calculated after several iterations of migration performed in homogeneous velocity models ranging from 2500 m/s to 3500 m/s every 100 m/s. To ease the comparison, we have plotted a normalised version of J_1:

$$\widetilde{\mathrm{J}}_1[c_0] = \frac{\left\| |h|\, \xi^{(\mathrm{N}+1)}[c_0] \right\|_{\mathscr{E}}^2}{\left\| \xi^{(\mathrm{N}+10)}[c_0] \right\|_{\mathscr{E}}^2}. \tag{2.43}$$

As already analysed by Lameloise *et al.* (2014), after one iteration the minimum of $\widetilde{\mathrm{J}_1}$ is not at the true velocity (3000 m/s) because of migration artefacts not focusing to zero-subsurface offset for the correct velocity. The minimum is reached for a too low velocity model as migration artefacts always have an upward curvature. The position of the minimum is a compromise between the defocusing of migration artefacts (they move away from the $h = 0$ axis as far as the velocity increases) and the defocusing of the event corresponding to the true reflector (decreasing when the velocity approaches the true velocity). As migration artefacts are attenuated when iterating, the minimum is progressively shifted to the true velocity.

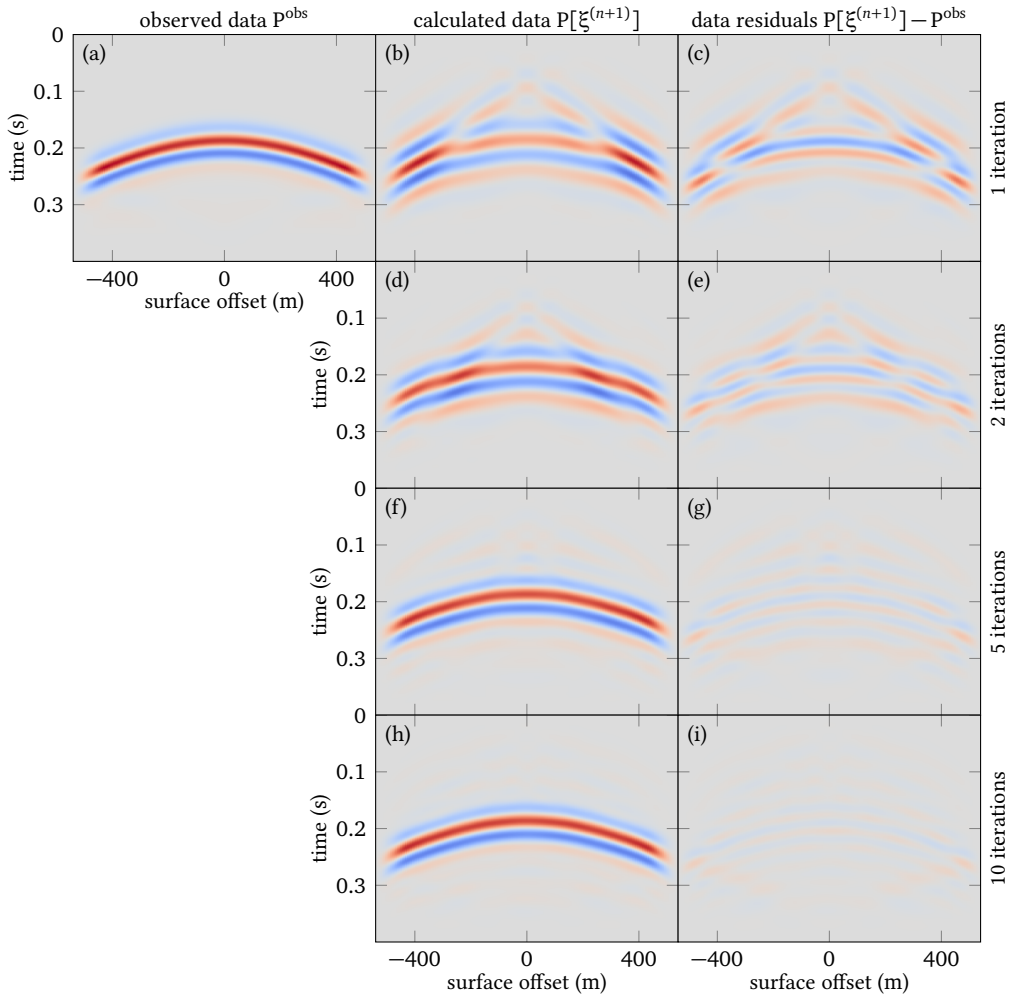


Figure 2.12. – Observed data obtained for a source at the middle point of the model (left). Corresponding calculated data (middle) and residuals (right) obtained after 1, 2, 5 and 10 iterations (from top to bottom), corresponding to the reflectivity models presented in figure 2.13.

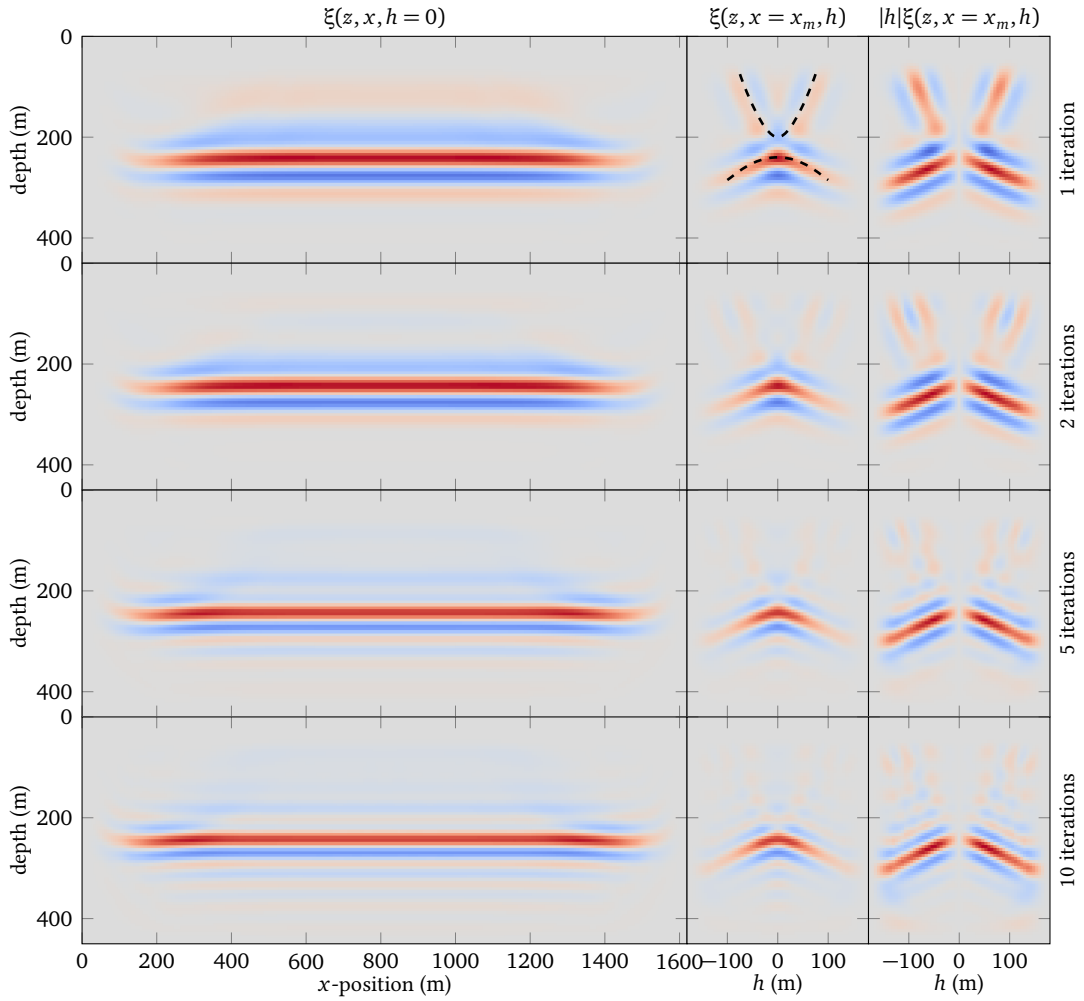


Figure 2.13. – Results of migration after 1, 2, 5 and 10 iterations (from top to bottom) when using a too slow velocity model (2500 m/s). We show a section at $h = 0$ (left), a CIG at $x = 1350$ m (middle), and the same CIG multiplied by $|h|$ (right). Blue, grey and red represent negative, null and positive values respectively. Each image is represented with its own colour scale. The two main events visible in the CIG at the first iteration are indicated by dashed lines: the downward curved event corresponds to the reflector, while the upward curved one is caused by the limited acquisition aperture.

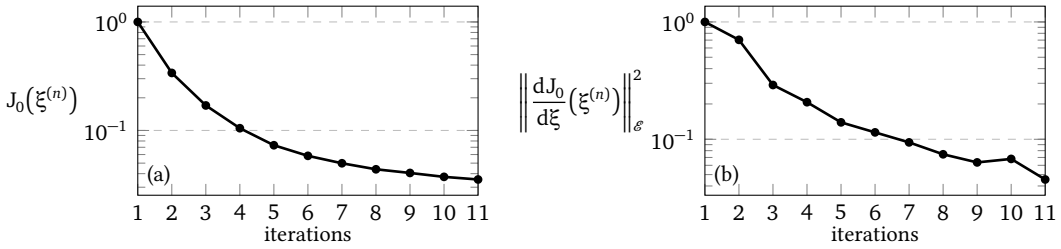


Figure 2.14. – Value of the objective function J_0 and norm of its gradient corresponding to the iterative migration results presented in figure 2.13.

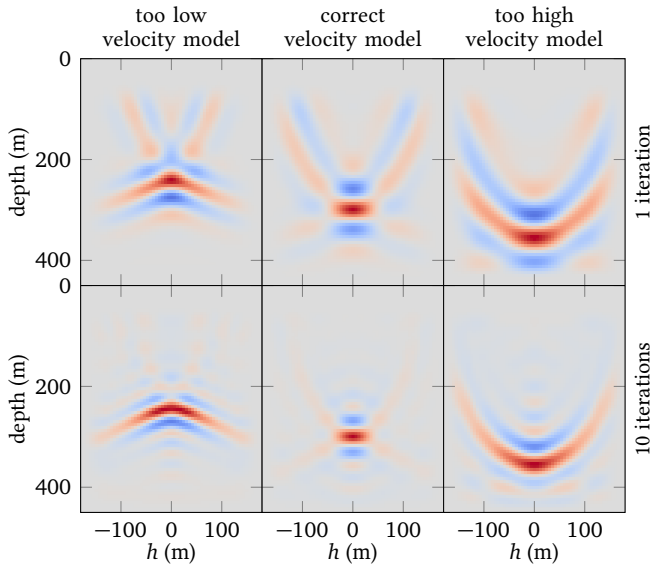


Figure 2.15. – CIGs at $x = 810\,\text{m}$ obtained for a too slow (2500 m/s, left), a correct (3000 m/s, middle) and a too high (3500 m/s, right) background velocity model after 1 (top) and 10 (bottom) iterations.

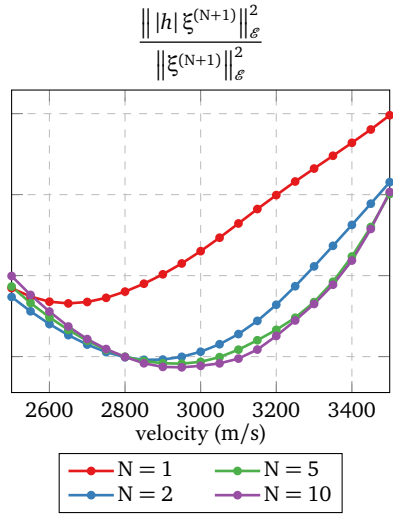


Figure 2.16. – Values of the objective function J_1 for different homogeneous velocity models. The correct velocity is 3000 m/s.

2.6.2. Primaries and multiples in a homogeneous medium

We now run the same example with first-order surface-related multiples. There are now two events in observed data (figure 2.17a). At the first iteration, both are interpreted as primary reflections, resulting in two events (figures 2.18a to 2.18c). The shallower one is the correct interpretation of the primary reflection and is similar to the primary-only case of the previous example. The misinterpretation of the surface multiple as a primary results in a cross-talk artefacts at twice the depth of the true event. As a consequence, reconstructed data do not match observed data. Iterations efficiently reduce the misfit (figures 2.17f, 2.17g and 2.19) and the final reflectivity section is free of migration artefacts and looks similar to the one obtained with primaries only (figure 2.18).

We also repeat the test of computing CIGs in three homogeneous velocity models (too low, correct and too high) (figure 2.20). After one iteration, all CIGs are altered with cross-talk artefacts (figure 2.20, top). Iterating allows to attenuate these artefacts and to obtain CIGs similar to those obtained in the previous example (figures 2.15 and 2.20, bottom). Note that in the special case of homogeneous background velocity models, cross-talk artefacts exhibit the same focusing behaviour as the true events, hence we do not present the equivalent of figure 2.16 in the multiple case. In chapter 4, we run a similar test in the case where velocity increases with depth and we show that cross-talk artefacts due to multiples favour lower velocities (see section 4.4.2 and figures 4.24b and 4.27 for more details).

2.7. Conclusion

We have described in this chapter the iterative migration scheme used to determine a model perturbation explaining observed data containing both primaries and first-order surface-related multiples. Data modelling is performed with a second-order Born approximation. Compared to a finite-difference acoustic propagation with a free surface condition, this approximation is accurate, provided that the velocity contrasts in the models are not too large (less than 1000 m/s), otherwise the kinematics of multiples is not correctly reproduced.

We have described an efficient way of computing the gradient of the migration objective function based on the adjoint state method. Synthetic examples show that a reflectivity section correctly explaining the data can be retrieved after a few iterations, even in an incorrect velocity model thanks to the extension of the model space with the subsurface offset. At the first iteration, migration artefacts due to the limited acquisition geometry appear on CIGs and the MVA objective function is not minimum for the correct velocity. Moreover spurious cross-talk events appear since multiple reflections are interpreted as primary events. Iterating allows attenuating both migration and cross-talk artefacts. As a consequence the minimum of the MVA objective function is at the correct velocity. It should also improve the shape of its gradient which defines a velocity updates. The derivation of this gradient and its analysis is detailed in the next chapter.

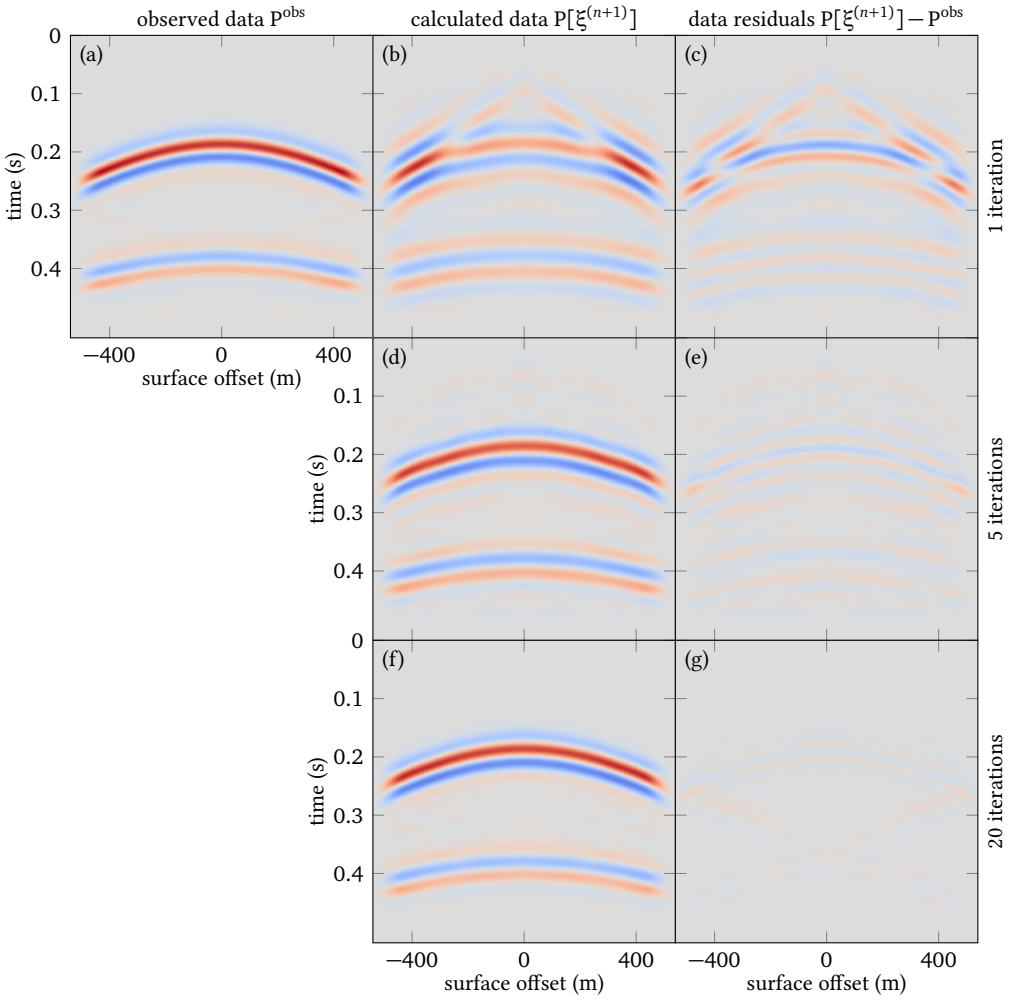


Figure 2.17. – Observed data obtained for a source at the middle point of the model (left). Corresponding calculated data (middle) and residuals (right) obtained after 1, 5 and 20 iterations (from top to bottom), corresponding to the reflectivity models presented in figure 2.18.

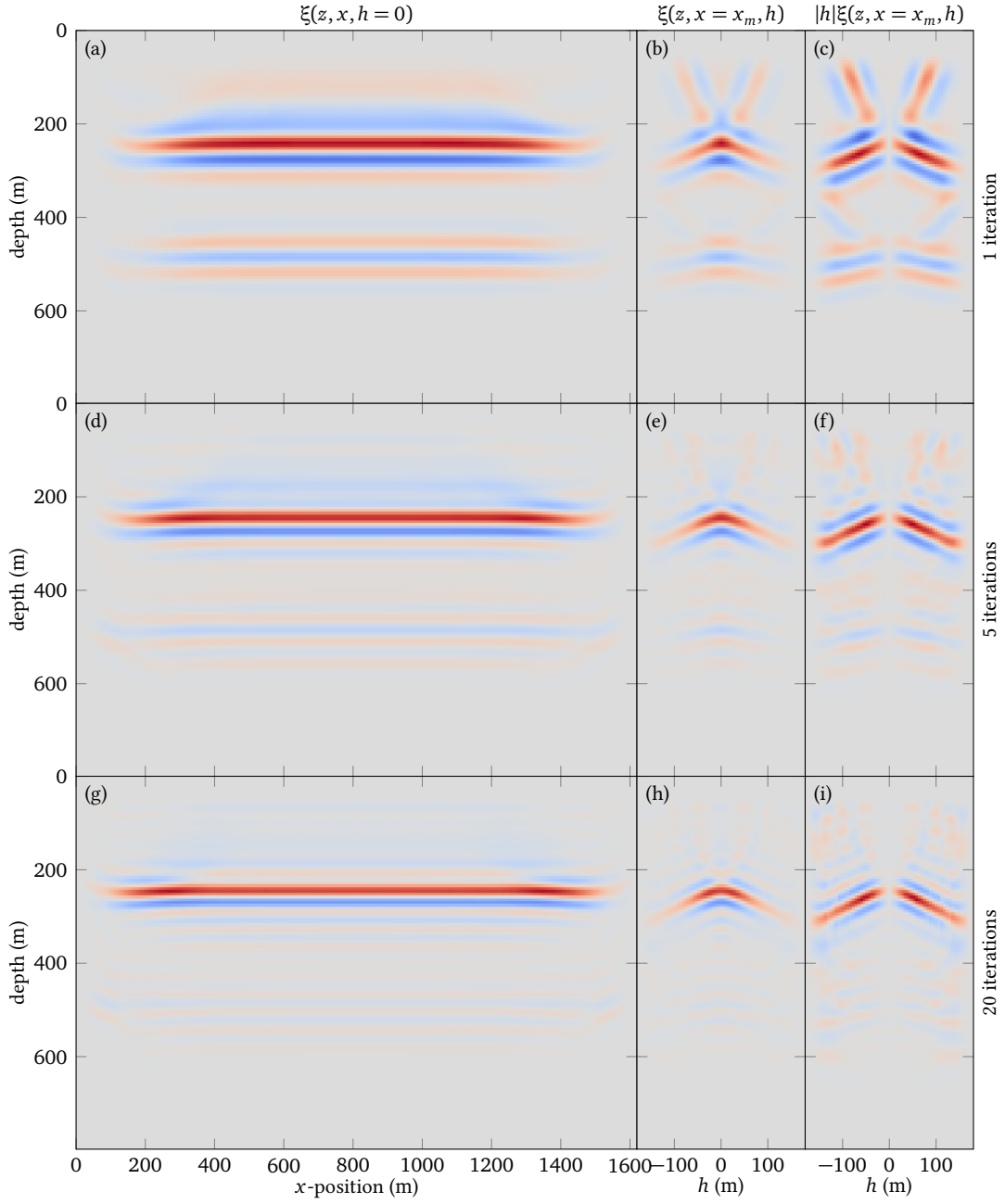


Figure 2.18. – Reflectivity model obtained after 1, 5 and 20 iterations (from top to bottom) in a too slow background velocity model $c_{\text{initial}} = 2500\,\text{m/s}$.

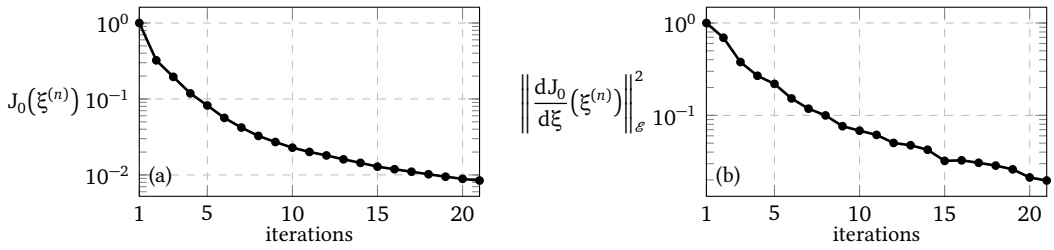


Figure 2.19. – Value of the objective function J_0 and norm of its gradient corresponding to the iterative migration results presented in figure 2.18.

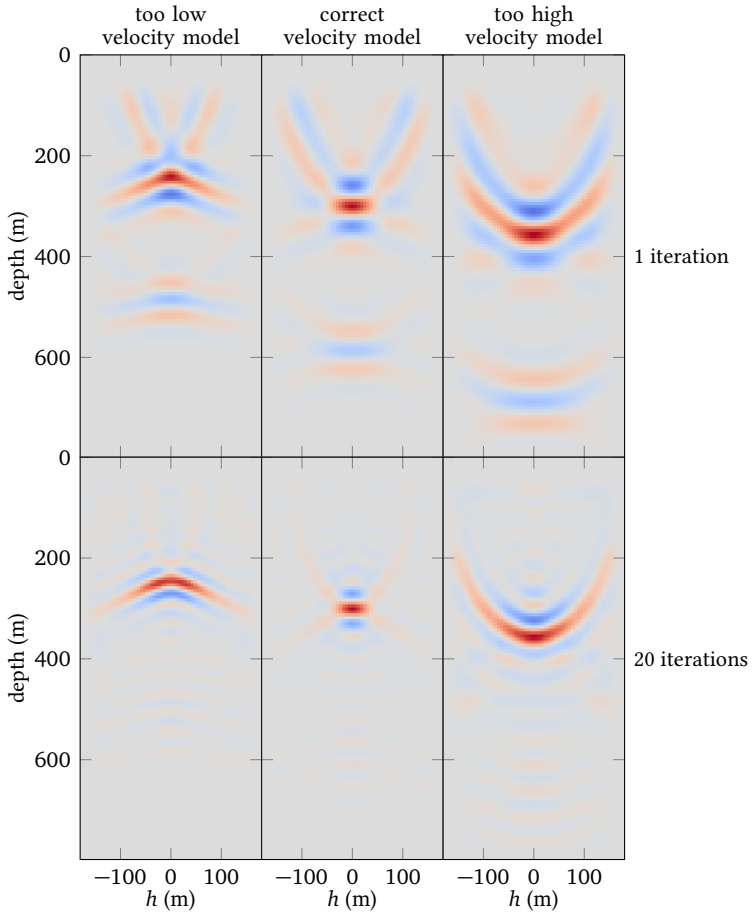


Figure 2.20. – CIGs at $x = 810$ m obtained for a too slow (2500 m/s, left), a correct (3000 m/s, middle) and a too high (3500 m/s, right) background velocity model after 1 (top) and 10 (bottom) iterations.

Chapter 3.

Iterative Migration Velocity Analysis

Contents

Résumé du chapitre 3

Ce chapitre s'intéresse à la remise à jour des grandes longueurs d'onde du modèle de vitesse de propagation des ondes P, ou « macro-modèle », par la méthode d'optimisation par semblance différentielle (DSO). La qualité du modèle de vitesse est évaluée sur le modèle de réflectivité étendu obtenu par l'algorithme de migration itérative présenté au chapitre 2. La fonction coût de la DSO pénalise l'énergie défocalisée dans ce modèle de réflectivité, et son gradient définit une remise à jour pour le macro-modèle. La construction du macro-modèle par minimisation de cette fonction sera illustrée dans le chapitre 6. Dans ce chapitre, je m'intéresse exclusivement au calcul du premier gradient.

L'algorithme d'*analyse de vitesse par migration itérative* étudié dans cette thèse se présente sous la forme d'un problème d'optimisation à deux niveaux (figure 1.22) (Colson *et al.*, 2007). Dans la boucle interne, le modèle de réflectivité est remis à jour pour une valeur fixe du macro-modèle. Dans la boucle externe, le macro-modèle est remis à jour en pénalisant l'énergie défocalisée dans le modèle de réflectivité solution du problème interne. Idéalement la fonction coût externe devrait être définie avec la solution optimale du problème interne. En pratique un nombre limité d'itérations est effectué dans la boucle interne, notamment pour limiter le coût numérique de la méthode. Ainsi une approximation de la fonction coût idéale et une approximation de son gradient sont utilisées en pratique. Il a été montré par Huang (2016) que le calcul de ce gradient était instable en pratique : de faibles différences entre deux modèles de réflectivité successifs, obtenus après un certain nombre d'itérations dans la boucle interne, ont un faible impact sur la valeur de la fonction coût interne, mais une grande influence sur la fonction coût externe et sur son gradient.

Afin d'illustrer et de mieux comprendre ce problème, je compare dans ce chapitre deux stratégies pour calculer une valeur approchée du gradient de la fonction coût externe vis-à-vis du macro-modèle. Toutes deux utilisent la méthode de l'état adjoint (Plessix, 2006). La méthode A évalue le gradient exact de la fonction coût approchée en tenant compte de la dépendance de chaque itération (interne) vis-à-vis du macro-modèle. Dans la méthode B, on suppose que le problème interne, appelé aussi problème direct, est résolu jusqu'à convergence, et on écrit les équations donnant le gradient exact de la fonction coût « idéale » définie avec cette solution optimale. Il faut alors résoudre un *problème adjoint*, linéaire même lorsque des multiples sont considérés, et le calcul du gradient fait intervenir les solutions optimales des problèmes direct et adjoint. En pratique, les calculs sont faits à partir des solutions approchées des problèmes direct et adjoint, si bien que la remise à jour du macro-modèle dans la méthode B est aussi un gradient approché.

Les deux méthodes sont testées dans le cas de primaires seuls sur un cas simple où un unique réflecteur horizontal est positionné dans un macro-modèle homogène. Le macro-modèle initial est homogène et trop lent, si bien que l'on s'attend à obtenir un gradient homogène et négatif au-dessus du réflecteur et nul en dessous. Les résultats obtenus en pratique illustrent l'importance déterminante du poids donnée à la régularisation dans la boucle interne. Avec une valeur satisfaisante de ce poids, les deux méthodes convergent après quelques itérations vers le même gradient, homogène et d'un signe cohérent avec la valeur du macro-modèle initial. En revanche si le poids de la régularisation est trop faible, les gradients données par les deux méthodes sont différents et ne convergent pas vers une valeur stable. Des oscillations subverticales apparaissent,

en particulier pour le gradient calculé avec la méthode A. Une analyse détaillée de cet exemple permet de mieux comprendre l'origine de ces instabilités : si la fonction coût associée à la migration itérative semble converger rapidement, la valeur de la fonction coût externe calculé pour les modèles de réflectivité successifs ne se stabilise pas et augmente graduellement. Des évènements sont progressivement créés aux grandes valeurs de l'offset en profondeur, dans le noyau de l'opérateur de modélisation mais avec une influence déterminante sur la fonction coût externe. Donner un poids suffisant à la régularisation contraint mieux la réflectivité aux grands offsets en profondeur et stabilise le calcul du gradient qui fournit alors une remise à jour cohérente du macro-modèle. En revanche un poids trop fort empêche une atténuation correcte des artefacts de migration ; de plus de fortes oscillations dans le gradient sont observées en pratique autour du réflecteur.

Ainsi l'atténuation des artefacts de migration au cours de la migration itérative conduit à des remises à jour du macro-modèle cohérentes, pourvu qu'une régularisation adéquate soit implémentée. Par la suite, seule la méthode B, plus simple à implémenter et moins sensible à l'apparition d'artefacts, sera considérée. En revanche, la migration itérative et la résolution du problème adjoint sont coûteuses numériquement et le choix d'un poids adéquat pour la régularisation reste difficile. Celui-ci peut être obtenu en traçant des courbes de Pareto, mais pour un coût de calcul prohibitif. Des solutions à ces difficultés sont proposées dans les chapitres suivants.

Finalement j'étudie le même exemple dans le cas de des réflexions multiples. Après une itération de migration itérative, les artefacts dus aux multiples dans le modèle de réflectivité conduisent à des valeurs non nulles du gradient en dessous du réflecteur. En itérant l'étape de migration, les artefacts dus aux multiples sont atténués et on obtient un gradient similaire à celui obtenu dans le cas de primaires seuls, pourvu qu'un poids adéquat soit donné à la régularisation.

3.1. Introduction

The aim of seismic imaging is the recovery of model's parameter such as the pressure wave velocity, allowing to numerically reproduce surface measurements. Determining the velocity model by minimising the least-squares misfit between observed data and calculated data may lead to an inaccurate estimation because the full waveform objective function has many local minima. The alternative method studied here consists of decomposing the velocity model into a smooth background model controlling the kinematics of wave propagation and a high-wavenumber part representing the reflectivity of the subsurface. Then for a given background velocity model, primary reflections are assumed to depend linearly on the reflectivity and multiple reflections quadratically. We have detailed in the preceding chapter the iterative determination of an extended reflectivity model free of cross-talk and migration artefacts minimising data misfit. The extension of the model allows to use the redundancy of observed data and to capture information about possible inaccuracies in the estimated background velocity model. Iterative migration corresponds to the inner loop of the nested optimisation strategy illustrated in figure 1.22. In this chapter, we are interested in the update of the background velocity model in the outer loop minimising defocused energy in the final reflectivity.

In this chapter we do not loop over the macro-model but only consider the derivation of the first update. Ideally, this update should be the exact gradient of the MVA objective function which measures defocused energy in a reflectivity model defined as the optimal solution of the migration inverse problem. This is not achievable in practice as we perform a limited number of iterations to determine the final reflectivity, as described in chapter 2.

The purpose of this chapter is the determination of an efficient way to compute an approximate gradient of the ideal MVA objective function. We would like the method to be robust against the gradient instability issue presented in section 1.3 (Huang, 2016). This issue will be illustrated in this chapter: we describe two different strategies to compute approximations of the ideal MVA gradient and show that they do not provide similar results. Furthermore, small differences on the reflectivity model appearing by further iterating in the inner loop result in large modifications to the gradient. We underline in this chapter the importance of a sufficiently strong regularisation coefficient to mitigate this effect. An alternative solution will be introduced in chapter 5.

In this chapter, we describe and compare two approximations of the ideal gradient. Both derivations use the adjoint-state method. In the first approach, we compute the exact gradient of an approximation of the ideal MVA objective function by differentiating all the iterations leading to the final reflectivity model. In the second approach, we assume that iterative migration has reached convergence and compute an approximate gradient of the ideal objective function. This optimality criterion is used to derive the equations of the second approach. These two methods are compared in terms of implementation and stability of the gradient computed for successive values of ξ.

In the first part of the chapter, the two methods are presented and compared on their computational merits. Then we present results of both methods on a simple synthetic example, first in the linear case of primary reflections only, modelled under the first-order Born linear approximation. Then the same example is run with first-order surface-related multiples using a second-order quadratic Born approximation. These examples illustrate the importance of regularising the migration objective function.

3.2. Computing the gradient of J_1

We investigate the computation of two different approximations of the gradient with respect to the background velocity model c_0 of the ideal MVA objective function

$$J_1^\infty[c_0] = \frac{1}{2}\left\|A\xi^\infty[c_0]\right\|_{\mathscr{E}}^2, \tag{3.1}$$

where ξ^∞ obeys the first order optimality condition of $J_0[c_0, \xi]$,

$$\frac{\partial J_0}{\partial \xi}[c_0, \xi^\infty] = 0. \tag{3.2}$$

Operator $A : \mathscr{E} \mapsto \mathscr{E}$ is called *annihilator*. For more detail on the meaning of this operator, we refer to section 1.3. In the numerical applications, the annihilator will consist of a multiplication by the absolute value of the subsurface offset: $A\xi(\mathbf{x}, h) = |h|\xi(\mathbf{x}, h)$. The subsurface offset is the extra parameter on which the focusing criterion is tested. For a correct model, $\xi(\mathbf{x}, h)$ should be null for $h \neq 0$. The annihilator penalises defocused energy arising in an incorrect background velocity model. We keep the notation A in this chapter for the sake of generality.

In the following, we first compute the exact gradient of an approximation $J_1^{(N)}$ of J_1^∞ obtained when N iterations are performed to solve the migration problem. We call this approach method A. In the second approach, called method B, we assume that iterative migration reaches convergence after N iterations and use the optimality criterion (3.2) to determine an approximate gradient of J_1^∞.

3.2.1. Method A: Exact gradient of an approximate objective function

In method A, we compute the exact gradient of the following approximation of J_1^∞, defined as

$$J_1^{(N)}[c_0] = \frac{1}{2}\left\|A\xi^{(N+1)}[c_0]\right\|_{\mathscr{E}}^2 \tag{3.3}$$

where $\xi^{(N+1)}$ is obtained after N iterations of gradient-based minimisation of the migration objective function (equation 2.21) described in the preceding chapter.

A general way of performing this computation is the use of automatic differentiation algorithms (Bell and Burke, 2008). Here we employ the adjoint state method (Plessix, 2006) to obtain an explicit expression allowing the analysis of each contribution to the total gradient. To keep simple expressions, we consider the case where iterative migration is performed with steepest descent and a linesearch procedure f_α. Then the state equations for the computations are given by

$$\xi^{(1)} = 0 \tag{3.4a}$$

$$\forall n \in [1, N] \quad \begin{cases} g^{(n)} = \dfrac{\partial J_0}{\partial \xi}\left[c_0, \xi^{(n)}\right] & \text{(3.4b)} \\ d^{(n)} = -g^{(n)} & \text{(3.4c)} \\ \alpha^{(n)} = f_\alpha\left(\xi^{(n)}, d^{(n)}\right) & \text{(3.4d)} \\ \xi^{(n+1)} = \xi^{(n)} + \alpha^{(n)} d^{(n)}, & \text{(3.4e)} \end{cases}$$

where $g^{(n)}$ is the gradient of the objective function, $d^{(n)}$ is the descent direction and $\alpha^{(n)}$ the step length at iteration n. We associate an adjoint variable to each equation of the iterative process (equations 3.4b to 3.4e). These Lagrange multipliers can be interpreted as constraints on the state equations. The associated Lagrangian reads:

$$\begin{aligned}\overline{J_1^A} = J_1^{(N)}[c_0] &- \sum_{n=1}^{N} \left\langle \eta^{(n)} \,\middle|\, \xi^{(n+1)} - \xi^{(n)} - \alpha^{(n)} d^{(n)} \right\rangle_{\mathscr{E}} \\ &- \sum_{n=1}^{N} \beta^{(n)} \left(\alpha^{(n)} - f_\alpha(\xi^{(n)}, d^{(n)}) \right) \\ &- \sum_{n=1}^{N} \left\langle \delta^{(n)} \,\middle|\, d^{(n)} + g^{(n)} \right\rangle_{\mathscr{E}} \\ &- \sum_{n=1}^{N} \left\langle \gamma^{(n)} \,\middle|\, g^{(n)} - \frac{\partial J_0}{\partial \xi}[c_0, \xi^{(n)}] \right\rangle_{\mathscr{E}}\end{aligned} \tag{3.5}$$

where the scalars $\beta^{(n)}$ and the vectors of $\mathscr{E}$ $\eta^{(n)}$, $\delta^{(n)}$ and $\gamma^{(n)}$ are the adjoint variables associated to the state variables $\alpha^{(n)}$, $\xi^{(n)}$, $d^{(n)}$ and $\gamma^{(n)}$ respectively . Similar to section 2.5.1 for the derivation of the gradient of J_0, the value of the adjoint variables is obtained by zeroing the derivatives of the Lagrangian with respect to the state variables,

$$\eta^{(N)} = A^T A \xi^{(N+1)} \tag{3.6a}$$

$$\forall n \in [N, 1] \begin{cases} \beta^{(n)} = \left\langle \eta^{(n)} \,\middle|\, d^{(n)} \right\rangle_{\mathscr{E}} & \text{(3.6b)} \\ \delta^{(n)} = \alpha^{(n)} \eta^{(n)} + \beta^{(n)} \dfrac{\partial f_\alpha}{\partial d}\left(\xi^{(n)}, d^{(n)}\right) & \text{(3.6c)} \\ \gamma^{(n)} = -\delta^{(n)} & \text{(3.6d)} \\ \eta^{(n-1)} = \eta^{(n)} + \dfrac{\partial^2 J_0}{\partial \xi^2}(c_0, \xi^{(n)})\gamma^{(n)} + \beta^{(n)} \dfrac{\partial f_\alpha}{\partial \xi}\left(\xi^{(n)}, d^{(n)}\right). & \text{(3.6e)} \end{cases}$$

The gradient of J_1 with respect to c_0 then equals to

$$G_A^{(N)} = \sum_{n=1}^{N} \left[\frac{\partial^2 J_0}{\partial \xi \partial c_0}[c_0, \xi^{(n)}] \right] \gamma^{(n)} + \sum_{n=1}^{N} \beta^{(n)} \frac{\partial f_\alpha}{\partial c_0}\left(\xi^{(n)}, d^{(n)}\right). \tag{3.7}$$

Thus method A is an iterative process initiated at iteration N with the adjoint source term $\eta^{(N)} = A^T A \xi^{(N+1)}$ named *image residual*. Then, equations (3.6b) to (3.6e) have to be solved from $n = N$ to $n = 1$. Method A requires the ability of computing the product of second-order derivatives of J_0 with the successive values of γ. An efficient technique to perform this calculation will be presented in section 3.2.3. Note that the linesearch function may depend on c_0, hence the second term in the gradient. If we consider a descent method with a fixed step, that is $\alpha^{(n)} = \alpha_0$, we have $\beta^{(n)} = 0$.

3.2.2. Method B: approximate gradient of the ideal objective function

In method B, we derive a simpler expression with the assumption that iterative migration is performed until convergence (Chauris *et al.*, 2015). That is, we suppose that the final reflectivity section obeys the optimality condition (3.2) and we consider the ideal MVA objective function (3.1), built with ξ^{∞} instead of $\xi^{(N+1)}$. The gradient can again be evaluated via the adjoint state method. The associated Lagrangian is

$$\overline{J_1^B}[c_0, \xi^{\infty}, \sigma] = J_1^{\infty}[c_0] - \left\langle \sigma \,\middle|\, \frac{\partial J_0}{\partial \xi}[c_0, \xi^{\infty}] \right\rangle_{\mathscr{E}}. \tag{3.8}$$

with σ the adjoint state that satisfies;

$$\left[\frac{\partial^2 J_0}{\partial \xi^2}[c_0, \xi^{\infty}]\right]\sigma = A^T A \xi^{\infty}. \tag{3.9}$$

The ideal solution σ^{∞} of this linear problem is the deconvolved version of the image residual $A^T A \xi^{\infty}$. The gradient of J_1 with respect to c_0 then reads

$$G_B^{\infty} = -\left[\frac{\partial^2 J_0}{\partial \xi \partial c_0}\left[c_0, \xi^{\infty}\right]\right]\sigma^{\infty}. \tag{3.10}$$

This is the optimal gradient obtained when both equations (3.2) and (3.9) are verified. In practice we approximate ξ^{∞} by $\xi^{(N+1)}$ and we solve iteratively the linear system

$$\left[\frac{\partial^2 J_0}{\partial \xi^2}[c_0, \xi^{(N+1)}]\right]\sigma = A^T A \xi^{(N+1)} \tag{3.11}$$

with M iterations. This leads to the approximate gradient of J_1

$$G_B^{(N,M)} = -\left[\frac{\partial^2 J_0}{\partial \xi \partial c_0}\left[c_0, \xi^{(N+1)}\right]\right]\sigma^{(N+1,M+1)}. \tag{3.12}$$

Even if the migration process is non-linear, the adjoint equation (3.9) is a linear problem of the type $H\sigma = \theta$ where H is the Hessian of J_0 and θ is the image residual. If the convergence assumption is satisfied, the residuals are small and the Hessian operator is positive semi-definite. Hence equation (3.9) is easily solved with a linear conjugate-gradient algorithm. Compared to method A, the adjoint equations of method B are solved independently of the optimisation strategy used for iterative migration. The adjoint problem may actually be solved with a different number of iterations, as the velocity update only involves the last value of $\xi^{(n)}$ and $\sigma^{(m)}$ regardless of how they are computed. Figures 3.1 and 3.2 illustrate how adjoint variables are computed in both methods and how they are combined with state variables to compute the velocity update.

We detail the advantages of both methods regarding implementation and memory requirements in section 3.2.5. First we elaborate on the computation of second-order derivatives of J_0.

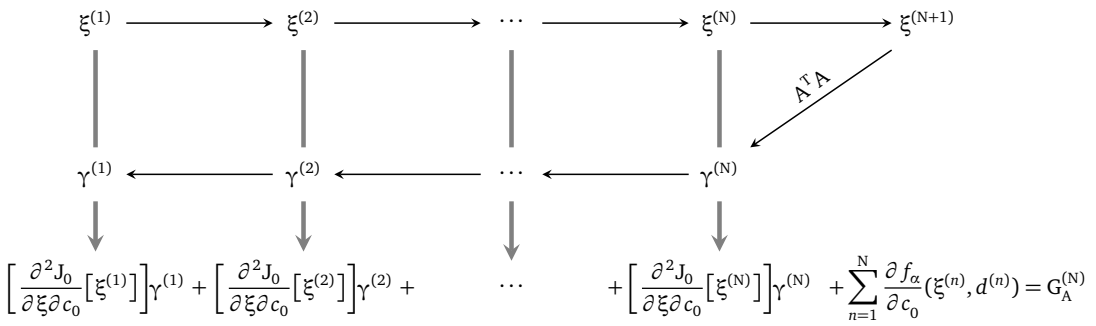


Figure 3.1. – Sketch of method A. Thin black arrows show the order of computation of the variables. Thick grey arrows show how they are combined to compute the velocity update

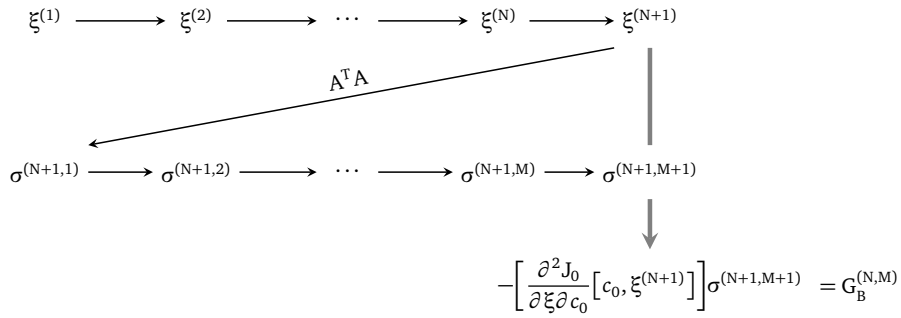


Figure 3.2. – Same as figure 3.1 for method B. Note that the numbers N and M of iterations performed to solve the direct and adjoint problems are not necessarily the same.

3.2.3. Computation of the second-order derivatives of J_0

Both methods involve the second-order derivatives of J_0. In method A, the product of $\partial^2 J_0/\partial\xi^2$ and $\partial^2 J_0/\partial\xi\partial c_0$ with the adjoint variable γ is required N times. In method B, the resolution of the adjoint system (equation 3.11) with the linear conjugate gradient requires M products of the Hessian $\partial^2 J_0/\partial\xi^2$ with a vector of $\mathscr{E}$ and the final computation of the gradient (equation 3.12) requires a single product of $\partial^2 J_0/\partial\xi\partial c_0$ with the final value $\sigma^{(N,M)}$. The contribution of the regularisation is straightforward, hence we focus on the second derivatives of J_{Migr}. Following Métivier *et al.* (2013), an efficient procedure to compute their product with a vector of $\mathscr{E}$ is defined with a second-order adjoint-state technique.

We consider a new scalar function Γ and a vector χ of $\mathscr{E}$

$$\Gamma[c_0,\xi,\chi] = \left\langle \frac{\partial J_{\text{Migr}}}{\partial\xi}[c_0,\xi] \,\middle|\, \chi \right\rangle_{\mathscr{E}} \tag{3.13}$$

such that the gradients of Γ with respect to ξ and c_0 equal the desired matrix-vector products

$$\begin{cases} \dfrac{\partial\Gamma}{\partial\xi} = \left[\dfrac{\partial^2 J_{\text{Migr}}}{\partial\xi^2}[c_0,\xi]\right]\chi, & \text{(3.14a)} \\ \dfrac{\partial\Gamma}{\partial c_0} = \left[\dfrac{\partial^2 J_{\text{Migr}}}{\partial\xi\partial c_0}(c_0,\xi)\right]\chi. & \text{(3.14b)} \end{cases}$$

We now apply the classic adjoint-state method to compute the gradients of Γ. We define a Lagrangian with the adjoint variables μ_i and ν_i associated to the state variables P_i and λ_i, respectively. The state equations are given by equations (2.20) and (2.34). The Lagrangian reads

$$\begin{aligned} \overline{\Gamma}[c_0,\xi,\chi,P_i,\lambda_i,\mu_i,\nu_i] = &\left\langle Q[P_0,\lambda_1] + Q[P_2,\lambda_3] \,\middle|\, \chi \right\rangle_{\mathscr{E}} \\ &- \left\langle \mu_0 \,\middle|\, \mathscr{L}_0 P_0 - S \right\rangle_{\mathscr{D}} \\ &- \left\langle \mu_1 \,\middle|\, \mathscr{L}_0 P_1 - K^-[P_0,\xi] \right\rangle_{\mathscr{D}} \\ &- \left\langle \mu_2 \,\middle|\, \mathscr{L}_0 P_2 - M_s P_1 \right\rangle_{\mathscr{D}} \\ &- \left\langle \mu_3 \,\middle|\, \mathscr{L}_0 P_3 - K^-[P_2,\xi] \right\rangle_{\mathscr{D}} \\ &- \left\langle \nu_3 \,\middle|\, \mathscr{L}_0^* \lambda_3 - M^{\mathrm{T}}\big(M(P_1+P_3) - P^{\text{obs}}\big) \right\rangle_{\mathscr{D}} \\ &- \left\langle \nu_2 \,\middle|\, \mathscr{L}_0^* \lambda_2 - K^+[\lambda_3,\xi] \right\rangle_{\mathscr{D}} \\ &- \left\langle \nu_1 \,\middle|\, \mathscr{L}_0^* \lambda_1 - M^{\mathrm{T}}\big(M(P_1+P_3) - P^{\text{obs}}\big) - M_s^* \lambda_2 \right\rangle_{\mathscr{D}}. \end{aligned} \tag{3.15}$$

We derive $\overline{\Gamma}$ with respect to the state variables P_i and λ_i to find the adjoint equations:

$$\begin{cases} \mathscr{L}_0 \nu_1 = K^-[P_0, \chi] & (3.16a) \\ \mathscr{L}_0 \nu_2 = M_s \nu_1 & (3.16b) \\ \mathscr{L}_0 \nu_3 = K^-[P_2, \chi] + K^-[\nu_2, \xi] & (3.16c) \\ \mathscr{L}_0^* \mu_3 = M^T M(\nu_1 + \nu_3) & (3.16d) \\ \mathscr{L}_0^* \mu_2 = K^+[\mu_3, \xi] + K^+[\lambda_3, \chi] & (3.16e) \\ \mathscr{L}_0^* \mu_1 = M^T M(\nu_1 + \nu_3) + M_s^* \mu_2 & (3.16f) \\ \mathscr{L}_0^* \mu_0 = K^+[\mu_1, \xi] + K^+[\lambda_1, \chi] & (3.16g) \end{cases}$$

The desired matrix-vector products are obtained by derivating $\overline{\Gamma}$ with respect to ξ and c_0

$$\frac{\partial \Gamma}{\partial \xi}[c_0, \xi, \chi] = \left[\frac{\partial^2 J_{\text{Migr}}}{\partial \xi^2}[c_0, \xi] \right] \chi = Q[P_0, \mu_1] + Q[P_2, \mu_3] + Q[\lambda_3, \nu_2] \tag{3.17}$$

$$\frac{\partial \Gamma}{\partial c_0}[c_0, \xi, \chi] = \left[\frac{\partial^2 J_{\text{Migr}}}{\partial \xi \partial c_0}[c_0, \xi] \right] \chi = \sum_{i=0}^{3} C[P_i, \mu_i] + \sum_{i=1}^{3} C[\lambda_i, \nu_i], \tag{3.18}$$

where we defined the classical normalised cross-correlation operator $C[u, v] : \mathscr{D} \times \mathscr{D} \mapsto \mathscr{M}$ for two vectors (u, v) of $\mathscr{D}$ as

$$C[u, v](\mathbf{x}) = \frac{2}{c_0^3(\mathbf{x})} \int_s \int_\omega (i\omega)^2 u(s, \mathbf{x}, \omega) v(s, \mathbf{x}, \omega) \, d\omega \, d\mathbf{s} \tag{3.19}$$

Note that both products can be computed simultaneously. Their calculations are twice as expensive as the computation of the gradient $\partial J_{\text{Migr}} / \partial \xi$ (table 3.1). Note that if we calculate only $\partial^2 J_0 / \partial \xi^2$, some direct and adjoint variables are not needed, leading to less computations. In particular, in the case of primaries only, this operator does not depend on ξ and is computed in the same way as the gradient of J_0.

3.2.4. Stability of the gradient

Ideally the MVA objective function should be evaluated at ξ^∞ and its gradient G^∞ used to update the background velocity model. In practice we perform a limite number of iterations and use $G_A^{(N)}$ or $G_B^{(N,M)}$. One may want to estimate the decrease in convergence speed made by considering approximate gradients for the minimisation of J_1^∞ (Friedlander and Schmidt, 2012) and bound the error defined as the difference between the macro-model recovered after a few outer iterations c_0^f and the exact macro-model c_0^*. Ideally it should be bound by the error made in the resolution of the inner problem by performing only N iterations. This analysis is not trivial as the gradient is not linear in ξ ($\partial J_1 / \partial c_0 = [\partial \xi / \partial c_0]^T A^T A \xi$). Some simple results can be stated, though. We have shown in chapter 2 that after an insufficient number of inner-iterations, the approximate objective function of method A is not minimum for the correct macro-model on one hand, and the gradient obtained with method B is very unlikely to lead to an accurate macro-model estimation on the other hand. An other issue is that the null space of the MVA

	primaries only	primaries and multiples
$J_0[c_0, \xi]$	2	4
$\frac{\partial J_0}{\partial \xi}[c_0, \xi]$	3	7
$\left[\frac{\partial^2 J_0}{\partial \xi^2}[c_0, \xi]\right]\chi$	3	11
$\left[\frac{\partial^2 J_0}{\partial \xi \partial c_0}[c_0, \xi]\right]\chi$	6	14

Table 3.1. – Number of wave-equations to be solved to evaluate the value, the gradient of the objective function at a point ξ, and the product of its Hessian with a vector χ. The number indicated here should be multiplied by the number of source positions considered in the acquisition.

objective function J_1 is not empty, meaning that different background velocity models can lead to relatively well-focused energy in CIGs. As an illustration, we present in figure 3.3 some results of iterations on the background velocity models. These results and the approach used to obtain them will be detailed in chapters 4 and 6. We display the exact macro-model used to compute observed data and two results obtained after twenty iterations on the macro-model starting with two different initial guesses. These three different macro-models result in CIGs with very similar focusing property. They differ in particular on the edges of the model where the acquisition setting does not constrain the model well. Larger surface-offset may better constrain these areas. This illustrates the practical difficulty to bound the error between the exact background model and the one recovered after several outer iterations.

Instead, we propose here to study the stability of the MVA gradient computed for a fixed background velocity model c_0 and successive values of the reflectivity model computed during iterative migration (Huang, 2016). Ideally we would like to be able to bound the error between the ideal gradient G^∞ and its approximations $G_A^{(N)}$ and $G_B^{(N,M)}$ by the error on the reflectivity model $\xi^{(N+1)} - \xi^\infty$. We begin by presenting three preliminary results for which we do not have formal proofs, but which allow to bound the error in the gradient by the error in the resolution of the inner problem. Then we use the numerical results presented in section 3.3.1 to determine if these hypotheses are reasonable.

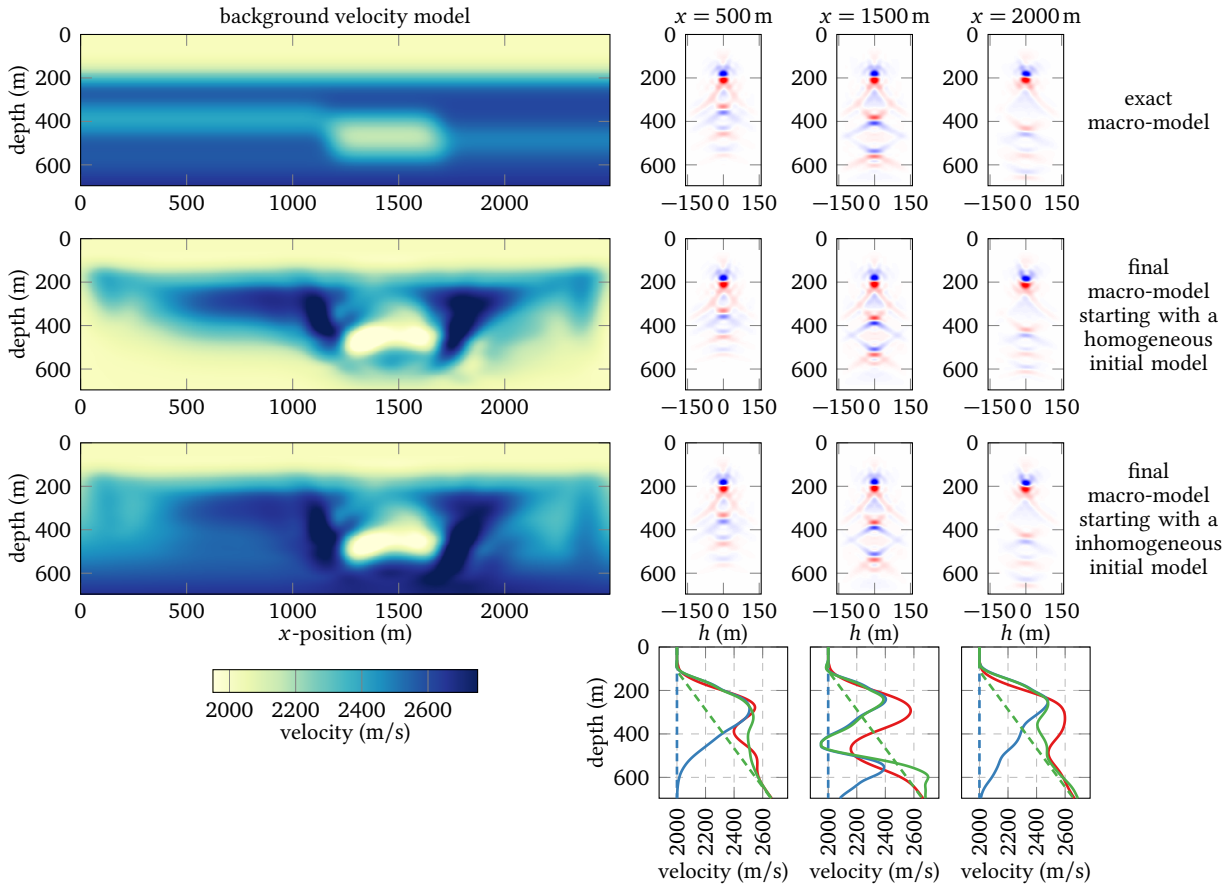


Figure 3.3. – Results of inversion (left, 2nd and 3rd row) obtained with observed data modelled in the exact model shown on the first row (left). These results are presented in more details in chapter 6. For each background velocity model, three CIGs are displayed at the lateral positions 500 m, 1500 m and 2000 m as well as three traces of the velocity models at the same locations. The red curve corresponds to the exact model. The blue and green curves correspond to the final models displayed in the 2nd and 3rd row, respectively. These results are obtained with two different initial models (dashed curves), a homogeneous macro-model and a model with values increasing with depth.

Lipschitz conditions

The stability of the MVA gradient across inner-iterations is based on the following Lipschitz conditions,

$$\left\| \frac{\partial}{\partial c_0}\left(\xi^{\infty} - \xi^{(N+1)}\right) \right\|_{\mathscr{E}\times\mathscr{M}} \leq \Xi \left\| \xi^{\infty} - \xi^{(N+1)} \right\|_{\mathscr{E}}, \tag{3.20a}$$

$$\left\| \frac{\partial^2 J_0}{\partial \xi \partial c_0}[c_0, \xi^{\infty}] - \frac{\partial^2 J_0}{\partial \xi \partial c_0}[c_0, \xi^{(N+1)}] \right\|_{\mathscr{E}\times\mathscr{M}} \leq \Xi_1 \left\| \xi^{\infty} - \xi^{(N+1)} \right\|_{\mathscr{E}}, \tag{3.20b}$$

$$\left\| \frac{\partial^2 J_0}{\partial \xi^2}[c_0, \xi^{\infty}] - \frac{\partial^2 J_0}{\partial \xi^2}[c_0, \xi^{(N+1)}] \right\|_{\mathscr{E}\times\mathscr{E}} \leq \Xi_2 \left\| \xi^{\infty} - \xi^{(N+1)} \right\|_{\mathscr{E}}, \tag{3.20c}$$

where Ξ, Ξ_1 and Ξ_2 are three positive constants.

We make two simplifying hypotheses, first a linear modelling $F : \mathscr{E} \mapsto \mathscr{D}_{\text{obs}}$ of the data $P[c_0, \xi] = F[c_0]\xi$, so that the derivatives of J_0 read

$$\frac{\partial J_0}{\partial \xi}[c_0, \xi] = H\xi - b, \tag{3.21a}$$

$$\frac{\partial^2 J_0}{\partial \xi^2}[c_0, \xi] = H, \tag{3.21b}$$

with $b = F^T P^{\text{obs}}$ and the Hessian $H = F^T F + a_\varphi I$. Second, we use a steepest descent direction and a constant step size α, then

$$H\xi^{(N+1)} = (I - C^N)b, \tag{3.22}$$

with the operator $C = I - \alpha H$ supposed to ensure convergence to ξ^{∞}, which satisfies

$$H\xi^{\infty} = b. \tag{3.23}$$

Then

$$\frac{\partial}{\partial c_0}(\xi^{(N+1)} - \xi^{\infty}) = -\frac{\partial}{\partial c_0}\left(H^{-1}C^N b\right) \tag{3.24a}$$

$$= -\left(\frac{\partial H^{-1}}{\partial c_0} C^N b + N H^{-1} \frac{\partial C}{\partial c_0} C^{N-1} b + H^{-1} C^N \frac{\partial b}{\partial c_0} \right) \tag{3.24b}$$

The convergence of this expression is determined by operator C^N and so is the convergence of the migration process (equation 3.22), which indicates that the Lipschitz condition (3.20a) may be acceptable.

In the linear case the Hessian $H = \partial^2 J_0 / \partial \xi^2[c_0, \xi]$ is independent of ξ, so that condition (3.20c) is trivial. Eventually we use equation (3.21a) to write

$$\frac{\partial^2 J_0}{\partial \xi \partial c_0}[c_0, \xi^{(N+1)}] - \frac{\partial^2 J_0}{\partial \xi \partial c_0}[c_0, \xi^{\infty}] = \frac{\partial H}{\partial c_0}(\xi^{(N+1)} - \xi^{\infty}) + H\frac{\partial}{\partial c_0}(\xi^{(N+1)} - \xi^{\infty}), \tag{3.25}$$

which guarantees condition (3.20b) provided the first Lipschitz condition (3.20a) is verified.

Stability of method A

In method A we consider the exact gradient of $J_1^{(N)}$, then the error can be expressed as

$$\Delta G_A^{(N)} = G_A^{(N)} - G^{\infty} = \left[\frac{\partial}{\partial c_0}\xi^{(N+1)}\right]^T A^T A\xi^{(N+1)} - \left[\frac{\partial}{\partial c_0}\xi^{\infty}\right]^T A^T A\xi^{\infty}. \tag{3.26}$$

We write $\xi^{\infty} = \xi^{\infty} - \xi^{(N+1)} + \xi^{(N+1)}$, so that

$$\Delta G_A^{(N)} = \left[\frac{\partial}{\partial c_0}\xi^{(N+1)}\right]^T A^T A\left(\xi^{(N+1)} - \xi^{\infty}\right) - \left[\frac{\partial}{\partial c_0}\left(\xi^{\infty} - \xi^{(N+1)}\right)\right]^T A^T A\xi^{\infty}. \tag{3.27}$$

Using the Lipschitz condition (3.20a), it leads to

$$\left\|\Delta G_A^{(N)}\right\|_{\mathscr{M}} \leq k_A \left\|\xi^{(N+1)} - \xi^{\infty}\right\|_{\mathscr{E}}, \tag{3.28}$$

with k_A a scalar constant. Hence, the stability of method A relies only on the hypothesis formulated in equation (3.20a).

Stability of method B

For method B, the optimal gradient is given by (3.10). In practice we use an approximate gradient: equation (3.2) is solved in N iterations to find $\xi^{(N+1)}$ and equation (3.11) is solved in M iterations to find $\sigma^{(N+1,M+1)}$. The error associated to these approximations reads

$$\Delta G_B^{(N,M)} = G_B^{(N,M)} - G^{\infty} = \left[\frac{\partial^2 J_0}{\partial\xi\partial c_0}\left[c_0, \xi^{\infty}\right]\right]\sigma^{(\infty,\infty)} - \left[\frac{\partial^2 J_0}{\partial\xi\partial c_0}\left[c_0, \xi^{(N+1)}\right]\right]\sigma^{(N+1,M+1)}. \tag{3.29}$$

Similarly to the previous case, we want to bound the error $\Delta G_B^{(N,M)}$ by the errors made in the resolution of the direct and adjoint problems, that is:

$$\left\|\Delta G_B^{(N,M)}\right\|_{\mathscr{M}} \leq k_\xi \left\|\xi^{(N+1)} - \xi^{\infty}\right\|_{\mathscr{E}} + k_\sigma \left\|\sigma^{(N+1,M+1)} - \sigma^{(N+1,\infty)}\right\|_{\mathscr{E}}, \tag{3.30}$$

with k_ξ and k_σ two constants. Therefore we decompose $\sigma^{(N+1,M+1)}$ into

$$\sigma^{(N+1,M+1)} = \sigma^{(N+1,M+1)} - \sigma^{(N+1,\infty)} + \sigma^{(N+1,\infty)} - \sigma^{(\infty,\infty)} + \sigma^{(\infty,\infty)}, \tag{3.31}$$

so that the error is the sum of three contributions

$$\begin{aligned}\Delta G_B^{(N,M)} = &\left[\frac{\partial^2 J_0}{\partial\xi\partial c_0}\left[c_0, \xi^{\infty}\right] - \frac{\partial^2 J_0}{\partial\xi\partial c_0}\left[c_0, \xi^{(N+1)}\right]\right]\sigma^{(\infty,\infty)} \\ &- \frac{\partial^2 J_0}{\partial\xi\partial c_0}\left[c_0, \xi^{(N+1)}\right]\left(\sigma^{(N+1,M+1)} - \sigma^{(N+1,\infty)}\right) \\ &- \frac{\partial^2 J_0}{\partial\xi\partial c_0}\left[c_0, \xi^{(N+1)}\right]\left(\sigma^{(N+1,\infty)} - \sigma^{(\infty,\infty)}\right).\end{aligned} \tag{3.32}$$

Using equation (3.20b), the first contribution in equation (3.32) may be bounded by $\|\xi^{\infty} - \xi^{(N+1)}\|$. The second contribution can be bounded by the error in the resolution of the adjoint

problem $\|\sigma^{(N+1,M+1)} - \sigma^{(N+1,\infty)}\|$. For the third term, we use the fact that $\sigma^{(N+1,\infty)}$ and $\sigma^{(\infty,\infty)}$ are the exact solutions of problems (3.11) and (3.9) respectively, so that

$$\left[\frac{\partial^2 J_0}{\partial \xi^2}\left[c_0, \xi^{(N+1)}\right]\right]\sigma^{(N+1,\infty)} - \left[\frac{\partial^2 J_0}{\partial \xi^2}\left[c_0, \xi^{\infty}\right]\right]\sigma^{(\infty,\infty)} = A^T A\left(\xi^{(N+1)} - \xi^{\infty}\right). \tag{3.33}$$

We write again $\sigma^{(N+1,\infty)} = \sigma^{(N+1,\infty)} - \sigma^{(\infty,\infty)} + \sigma^{(\infty,\infty)}$ to obtain

$$\begin{aligned}\left[\frac{\partial^2 J_0}{\partial \xi^2}\left[c_0, \xi^{(N+1)}\right]\right]\left(\sigma^{(N+1,\infty)} - \sigma^{(\infty,\infty)}\right) = &\left[\frac{\partial^2 J_0}{\partial \xi^2}\left[c_0, \xi^{\infty}\right] - \frac{\partial^2 J_0}{\partial \xi^2}\left[c_0, \xi^{(N+1)}\right]\right]\sigma^{(\infty,\infty)} \\ &+ A^T A\left(\xi^{(N+1)} - \xi^{\infty}\right).\end{aligned} \tag{3.34}$$

Using equations (3.20c) and (3.34), we can bound the third contribution in equation (3.32) by the error $\|\xi^{\infty} - \xi^{(N+1)}\|$. In the linear case, equation (3.20c) is automatically verified and the validity of equation (3.20b) is implied by equation (3.20a), as observed above. We conclude that for both methods A and B, the validitity of equation (3.20a) is essential to the stability of the MVA gradient across inner-iterations. The numerical examples of section 3.3.1 will help us to determine under which conditions this hypothesis is verified.

3.2.5. Comparison in terms of implementation

Methods A and B differ regarding implementation of the adjoint equations. In method A, the adjoint equations depend on the optimisation method chosen for iterative migration, whereas in method B, the adjoint variables are solution of a linear system solved independently of the calculation made for the migration. Method B is thus easier to implement. In addition, the adjoint equations of method A (equations 3.6) involve the values of the state variables of migration $\xi^{(n)}$, $g^{(n)}$, and $d^{(n)}$ at each iteration. Hence method A requires storage of 3N $\mathscr{E}$-vectors. An alternative is to recompute their value when required during the computation of the adjoint variables. But they are needed in the reverse order to the one in which they are computed: we need first $\xi^{(N)}$, then $\xi^{(N-1)}$, etc. Recomputation is then very expensive. In comparison, the adjoint variables in method B are computed independently of the successive values of the state variables. Only the last value of $\xi^{(n)}$ is needed to initiate the resolution, then we only store the variables needed to use the conjugate-gradient algorithm.

In both methods the main computational step is the product of the Hessian with a $\mathscr{E}$-vector. We have presented in section 3.2.3 an efficient way to perform this calculation based on second-order adjoint-state approach (Métivier *et al.*, 2013). Note that the wavefields P_i and λ_i involved in the matrix-vector products have already been computed during the migration step, so that storing them would save a lot of computation time. However, these 4D arrays ($n_t \times n_z \times n_x \times n_s$) are too big to be stored and recomputation is necessary. We cannot process data source by source as for example the calculation of ξ implies a summation over all sources.

In method A, computing the derivative of the optimal step length with respect to the velocity model and the state variables adds extra computation of J_0 derivatives, especially in the non-linear case where the optimal step $\alpha^{(n)}$ is the result of a sophisticated procedure involving several trial steps; the formulas for this derivatives may not be straightforward as well. Hence, method A is roughly 1.5 times more expensive in computation time.

3.3. Synthetic examples

3.3.1. Examples with primaries only

We consider the linear case of primary reflections only and a model similar to the one studied in the numerical application of section 2.6. The model is 360 m deep and 2700 m large. The exact and initial macro-models are homogeneous (3000 m/s and 2500 m/s, respectively). A single horizontal reflector is located at 300 m depth. Sources are located at each grid point between $x = 540$ m and $x = 2160$ m. Receivers are located at each grid point within ±540 m around the source. As we are dealing with a very simple model, the ℓ_2-norm is chosen as regularisation function. In case of a more complex geology with dipping reflectors for example, we may need a regularisation term function of depth to account for the poorer illumination of deeper structures. The importance of this regularisation term for the velocity update will be discussed later (section 3.3.2). After 10 iterations of linear conjugate gradient with optimal step length, both the value of J_0 and J_1 reached convergence (figures 3.4a and 3.4c) and the gradient of J_0 is close to zero (figure 3.4b).

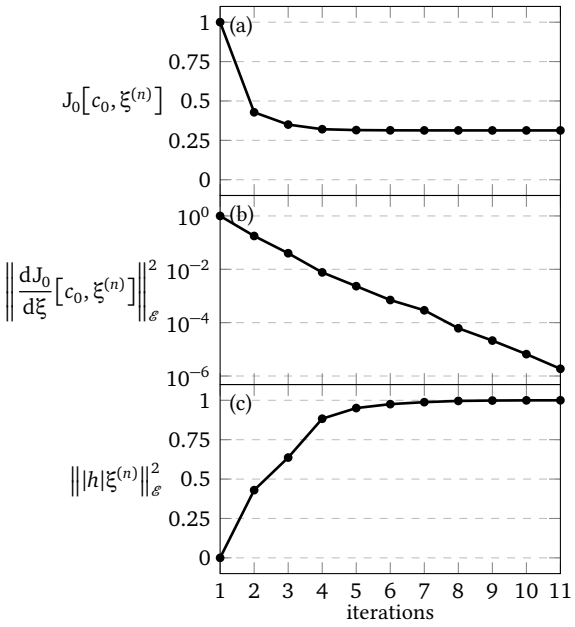


Figure 3.4. – (a) Value and (b) norm of the gradient of the objective function J_0 for ten iterations of conjugate gradients with sufficient regularisation. (c) Norm of penalised CIGs. All plots are normalised by their maximum value. Note that $c_0(\mathbf{x})$ is fixed.

The gradient of J_1 is computed with both methods after each iteration (figure 3.5). For method B, we choose in this example to use the same number of iterations to solve equations (3.2) and (3.11). The convergence speed of the resolution of the adjoint problem will be studied in more details and compared to the one of the direct problem in chapters 4 and 5 with the introduction of preconditioning. Here we set $M = N$ for simplicity. At the first iteration, positive values appear above 150 m depth for both methods. They are related to migration

artefacts appearing in CIGs after a single iteration (figure 3.6). With iterations these artefacts are progressively attenuated. After 10 iterations, the gradients obtained with both methods are similar and have the expected constant negative value. The central part is homogeneous, while the effect of the acquisition geometry alters the edges of the gradient. Note also that the imprint of the reflector is still visible on the final velocity update (around $z = 250\,\mathrm{m}$).

Both methods converge with approximately the same speed, although the gradient obtained after 5 iterations for method B is already very close to the final result.

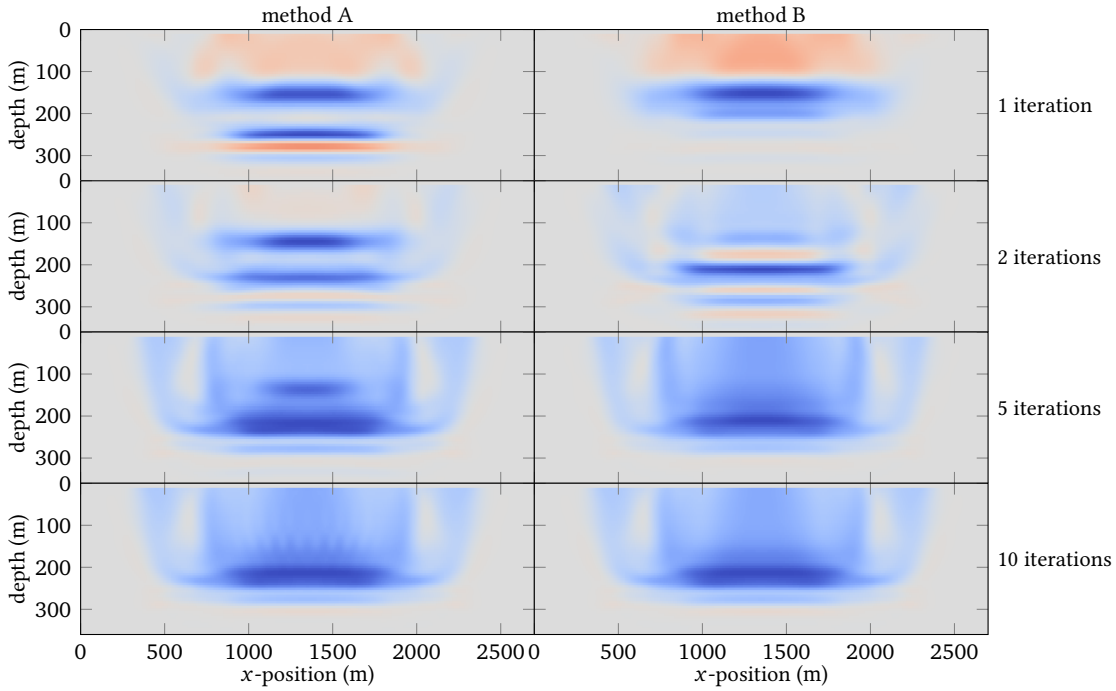


Figure 3.5. – Gradients of J_1 obtained with method A (left) and method B (right) after (from top to bottom) 1, 2, 5 and 10 iterations of iterative migration with sufficient regularisation. Blue, grey and red represent negative, null and positive values respectively. Each velocity update is plotted with its own colour scale.

3.3.2. Importance of regularisation

To illustrate the importance of the regularisation term in the migration objective function, we run the same example with a 50 times smaller value of a_φ. The corresponding gradients of J_1 (figure 3.7) have shapes similar to the previous case (figure 3.5) for the first iterations but after 10 iterations they are altered with sub-vertical spurious oscillations, emphasised on figure 3.7 with dashed lines for method A. Moreover the gradients obtained for both methods do not look similar, although the value of J_0 seems to reach convergence (figure 3.8a).

Actually, the norm of the penalised CIGs still increases steadily after 10 iterations (figure 3.8c). This norm is not used in the inner loop iterative process, but we expect it to converge to a

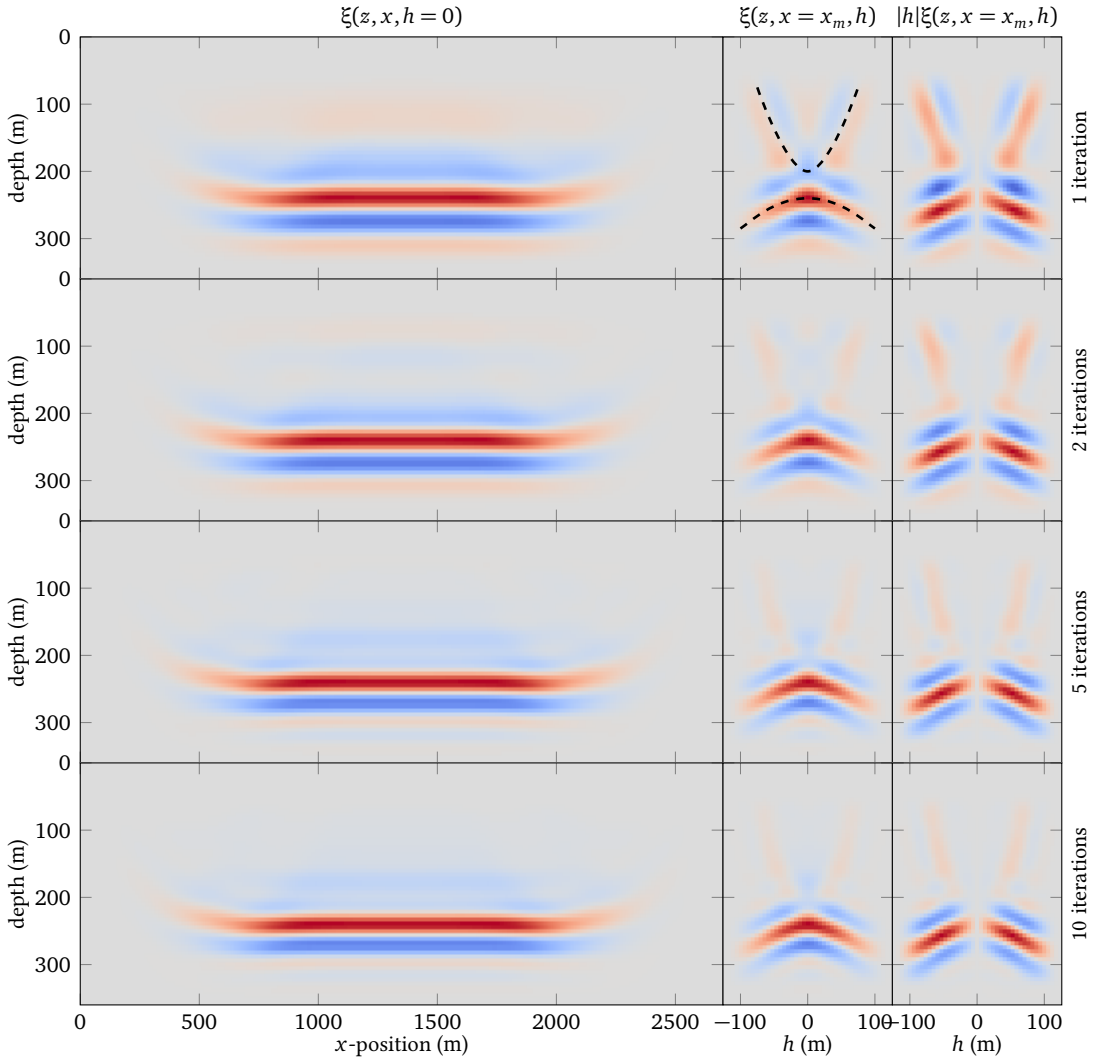


Figure 3.6. – Results of migration after 1, 2, 5 and 10 iterations (from top to bottom) when using a too slow velocity model (2500 m/s). We show a section at $h = 0$ (left), a CIG at $x = 1350$ m (middle), and the same CIG multiplied by $|h|$ (right). Blue, grey and red represent negative, null and positive values respectively. Each image is represented with its own colour scale. The two main events are indicated by dashed lines: the downward curved event corresponds to the reflector, while the upward curved one is caused by the limited acquisition aperture. This figure is very similar to figure 2.13, except that a different acquisition has been used here.

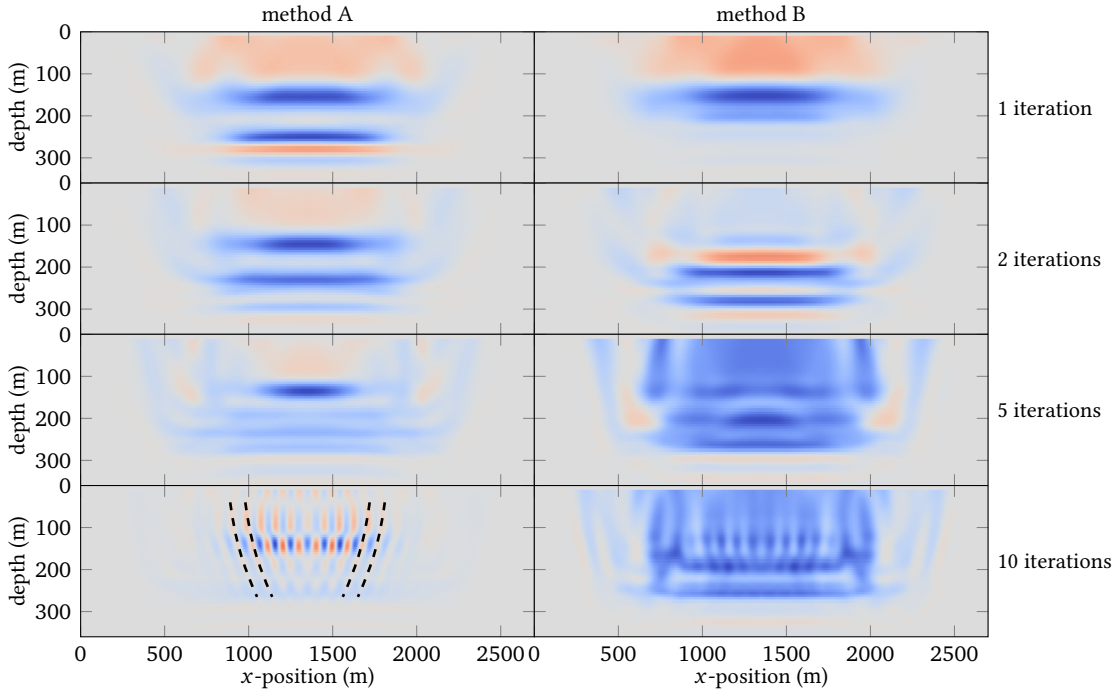


Figure 3.7. – Same as figure 3.5 with a smaller regularisation weight a_φ.

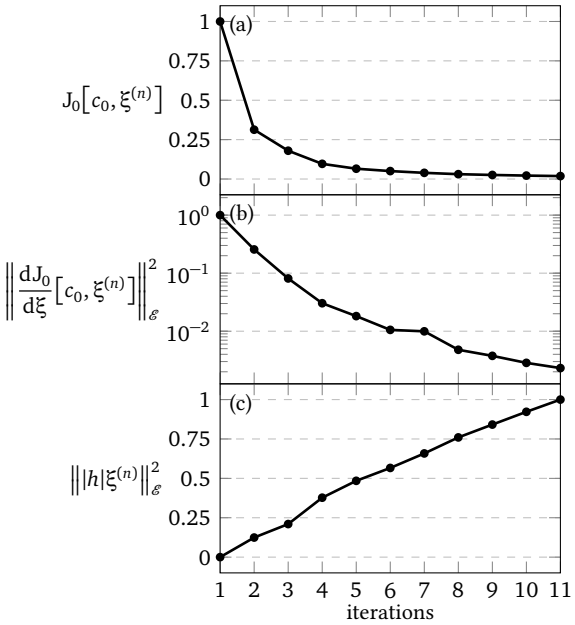


Figure 3.8. – Same as figure 3.4 with a smaller regularisation weight a_φ.

stable value as does ξ. This result is not satisfactory and indicates that residual energy at large values of h in the CIGs (figure 3.9, right) has a weak impact on the objective function J_0, but is amplified by the multiplication by h and deeply influences the value of J_1 and the velocity update. A sufficiently strong regularisation term allows to better constrain the reflectivity model

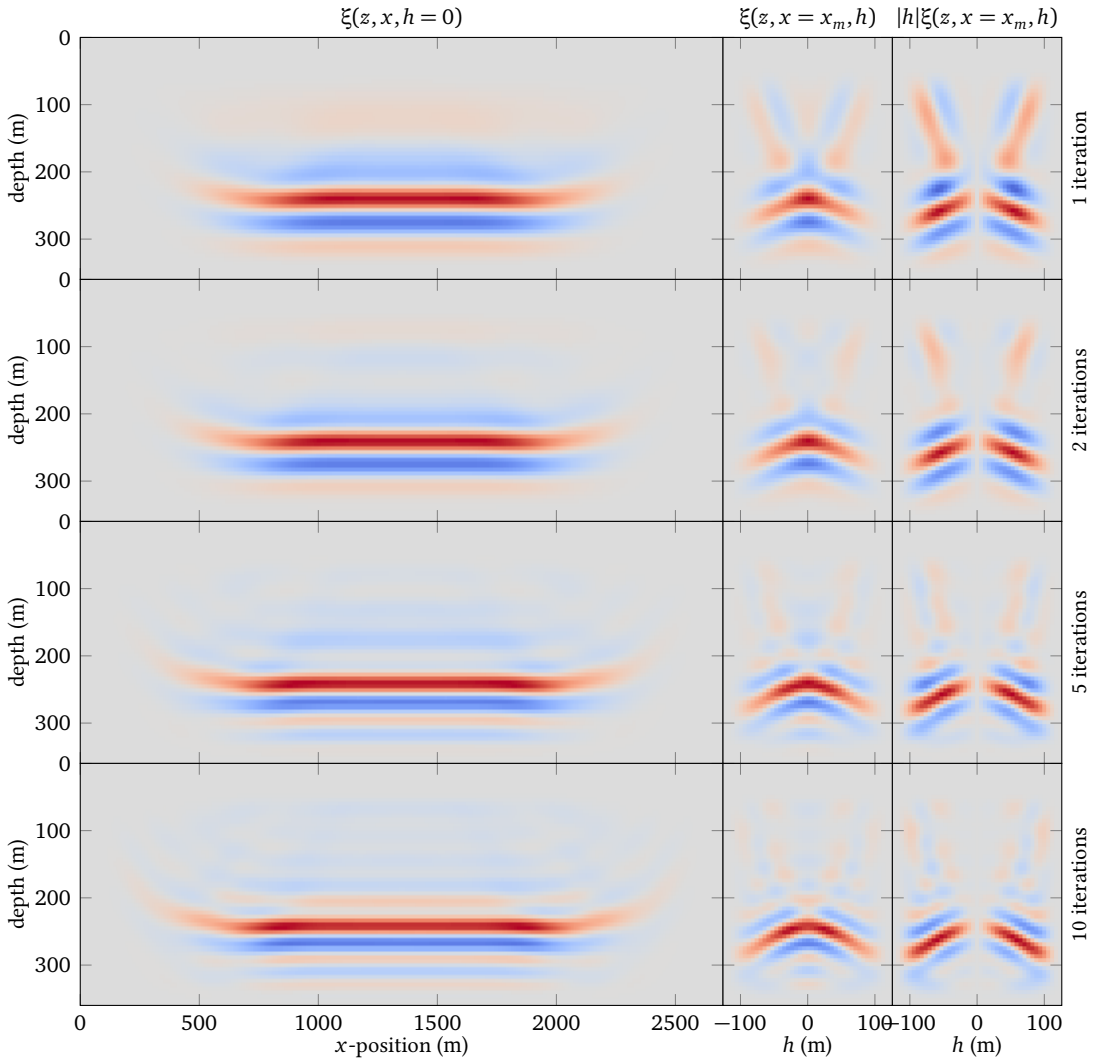


Figure 3.9. – Same as figure 3.6 with a smaller regularisation weight a_φ.

at large values of h and ensures the convergence of the value of the MVA objective function with iterations, as illustrated by the former example. Finally, we would like to put these results in perspective with the theoretical study of the gradient stability led in section 3.2.4. We have shown that the stability of the gradient obtained with both methods in the linear case was ensured by the validity of the Lipschitz condition (3.20a). The results shown in figures 3.5 and 3.7 suggest that equation (3.20a) is not valid unless sufficient regularisation is introduced during iterative migration. This observation is complementary to the observations of Huang

(2016), that no theoretical relation between the error on the gradient and the error in the inner iterations can be found. More precisely, referring to section 1.3, figure 1.11d suggests that the error increases exponentially.

We now identify the terms in the calculation of the gradient with method A that are responsible for the sub-vertical spurious oscillations. We use an example with smaller dimensions but with the same reflectivity and velocity value. Here, after 8 iterations of migration with steepest descent and optimal step size, we obtain with method A the gradient presented in figure 3.10a, which is altered with the same kind of artefacts. We control the correct computation of the gradient by also deriving the gradient with a finite difference approach, requiring only to evaluate the objective function. As this is very expensive (proportional to the number of model parameters), we restrict the computation to a single line at depth $z = 96\,\mathrm{m}$ and obtain a similar result (figure 3.10b). These spurious oscillations are already visible in the adjoint variables (see for example $\gamma^{(8)}$ in figure 3.11a) and come from small oscillations located above the reflector in the final migrated image, especially at large offsets. Their energy in $\xi^{(N+1)}$ is weak compared to the reflector, but they are strengthened in the computation of successive adjoint variables by two mechanisms. The first is the iterative application of the Hessian $\partial^2 \mathrm{J}_{\mathrm{Migr}}/\partial \xi^2[c_0, \xi^{(n)}]$ in the construction of the adjoint variables. In the linear case, it does not depend on $\xi^{(n)}$ and its application to a vector χ of $\mathscr{E}$ is a sequence of a data modelling step using χ followed by a migration of this data back to the $\mathscr{E}$-space. Both steps involve two Green's functions from the source and the receiver position to a point of the subsurface. In a smooth model, the asymptotic amplitude term of the Green's function decays with the square-root of the distance, therefore the Hessian has a dynamic effect of strengthening the shallow events and attenuating deeper ones. As the background velocity used for modelling and migration is the same, the diagonal term of the Hessian has no kinematic effect and the position of events is thus not modified. We show the effect of the Hessian applied to the adjoint variable $\gamma^{(8)}$ (figure 3.11a), which contains both horizontal events (emphasised with blue dashed lines) and vertical events (red dashed lines) with the same shape as the artefacts of the velocity update. By nature, these artefacts are always above the interfaces and thus are amplified after application of the Hessian (figure 3.11b), while the deeper horizontal events are attenuated. The regularisation term mitigates the attenuation of deeper events and prevents the apparition of artefacts.

The second origin of artefacts comes from the linesearch contributions $\partial f_\alpha/\partial \xi(\xi^{(n)}, d^{(n)})$ and $\partial f_\alpha/\partial d(\xi^{(n)}, d^{(n)})$ (figure 3.12a) to the construction of adjoint variables in method A (equation 3.6). They depend on the corresponding gradient of migration $\partial \mathrm{J}_0/\partial \xi[c_0, \xi^{(n)}]$, which has non-negligible energy above the reflector in the last iterations (figure 3.12b). The contributions of the linesearch also involve a sequence of modelling and migration steps applied to this gradient, hence, similarly to the application of the Hessian, the spurious events above the reflector are strengthened.

3.3.3. Example with multiples

To investigate the ability of iterative migration velocity analysis to handle multiple reflections, we run the same example with first-order surface-related multiples added to observed and calculated data. To take cross-talk events into account, the model is extended to 600 m depth. Two events are present in observed data: the primary reflection and its associated surface

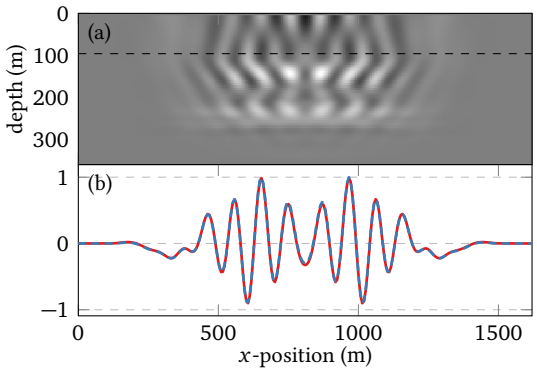


Figure 3.10. – (a) Gradient of J_1 computed using method A after 8 iterations. (b) Section of this image at $z = 96\,\mathrm{m}$ (red, solid) compared to the gradient obtained with finite differences (blue, dashed). Both are normalised with the same constant.

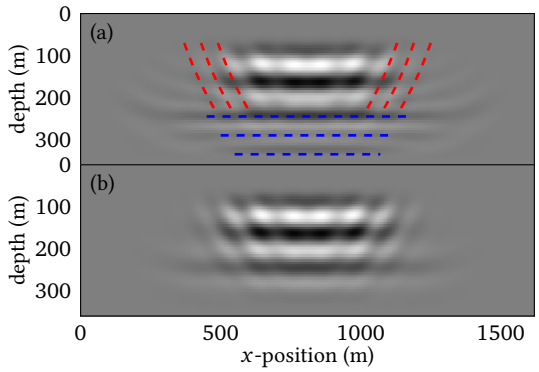


Figure 3.11. – Illustration of the effect of migration/demigration on $\gamma^{(8)}$. (a) Value of variable γ at iteration 8 (section at $h = 0$). (b) result of the application of the Hessian on this image (section at $h = 0$).

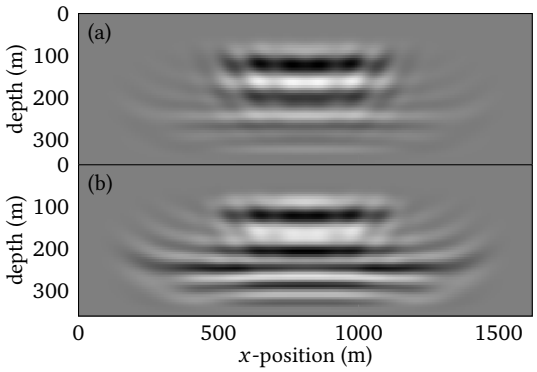


Figure 3.12. – (a) Contribution of the line search $\frac{\partial f_\alpha}{\partial d}(\xi^{(8)}, d^{(8)})$ to the equation of $\delta^{(8)}$. (b) Gradient of J_0 at the end of the iterations $g^{(8)} = \frac{\partial J_0}{\partial \xi}[c_0, \xi^{(8)}]$ (section at $h = 0$).

multiple, recorded at approximately twice the time of the primary. To ensure the multiple event has enough influence, the residuals between observed data and calculated data are multiplied by the recording time t in the definition of J_0. The data weight is taken into account in the derivation of the gradient of J_1 with respect to the background model.

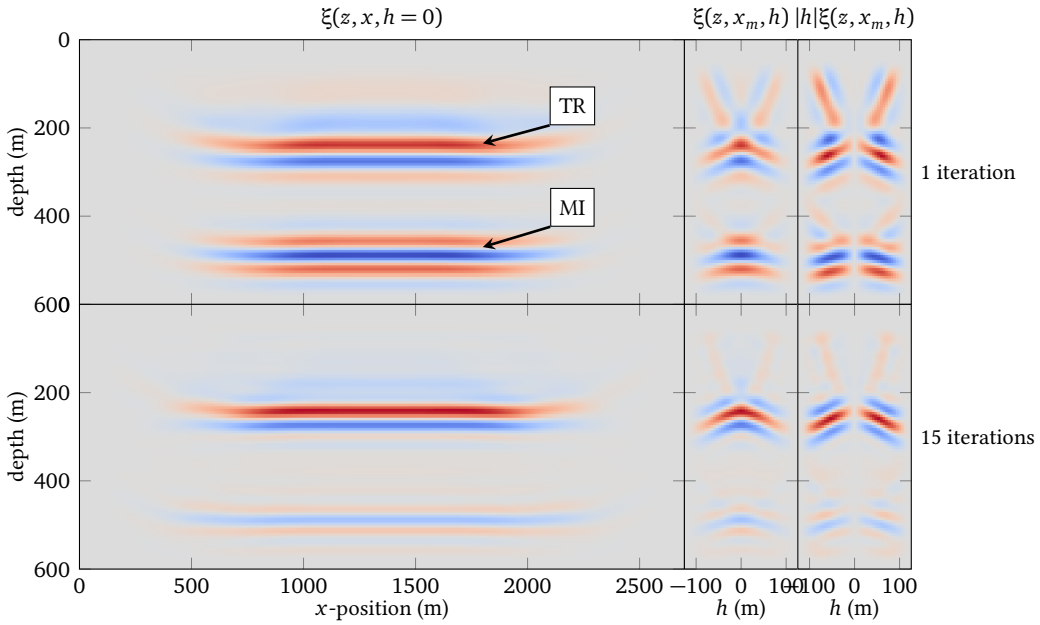


Figure 3.13. – Results of migration after 1 (top) and 15 (bottom) iterations of conjugate gradient when we consider primaries and first-order surface-related multiples. Section at $h = 0$ (left), CIG at $x = 1350\,\text{m}$ (middle), and the same CIG penalised by $|h|$ (right). Two event are visible: the true reflector (TR) and the multiple imprint caused by cross-talk (MI)

In the first iteration, starting from $\xi = 0$, both events in observed data are interpreted as primary reflections. The primary reflection produces the same event as in the primaries-only case, while the multiple event adds an imprint at twice the depth of the true reflector. This artefact creates a new primary in the modelled data at the next iteration, which in turn will produce an update with a negative sign in the new reflectivity update. This way cross-talk artefacts are progressively attenuated.

We now compute the gradient of the MVA objective function after one and fifteen iterations of migration. Compared to the primary only case, a more complex linesearch procedure f_α is used during migration. Thus the derivation of the associated contributions to the exact gradient in method A become much more complex too. Moreover, method B seems less sensitive to vertical spurious oscillations illustrated in figure 3.7. Therefore we use method B for the computation of the gradient in this multiple example figure 3.14. At the first iteration, the cross-talk artefact adds non-physical energy below the true reflector. After 15 iterations of migration, this artefact is largely attenuated and the gradient looks similar to the one obtained in the case where only primaries were considered.

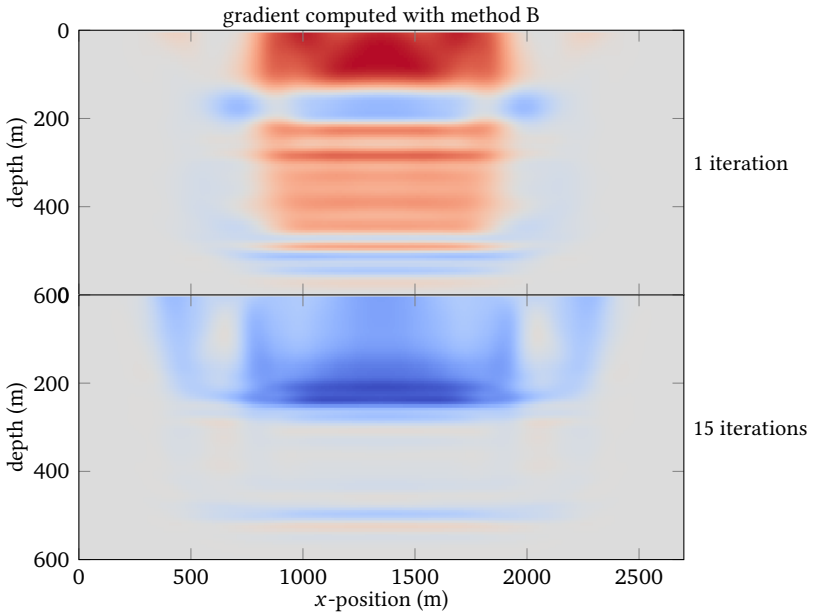


Figure 3.14. – Velocity updates obtained with method B after 1 (top) and 15 (bottom) iterative migration of data containing both primaries and first-order surface-related multiples. They correspond to the migrated images of figure 3.13. Blue, grey and red represent negative, null and positive values respectively. Each velocity update is plotted with its own colour scale.

Choice of the regularisation coefficient

Similar to the case of primaries only, a sufficiently strong regularisation coefficient is required to avoid the apparition of spurious vertical oscillations. However if too much weight is given to the regularisation term, the attenuation of cross-talk artefacts is not as efficient and their imprint is still visible both on the final reflectivity model and the MVA gradient. As an illustration, we compute the gradient obtained with method B and 15 iterations of iterative migration in three cases corresponding to three values of the regularisation coefficient (figure 3.15). The choice of the regularisation is thus a trade-off between the smoothness of the gradient and the attenuation of artefacts and should in principle be the result of a Pareto curve analysis (Hansen, 1999). However such an analysis is quite expensive, and a_φ is chosen empirically in practice.

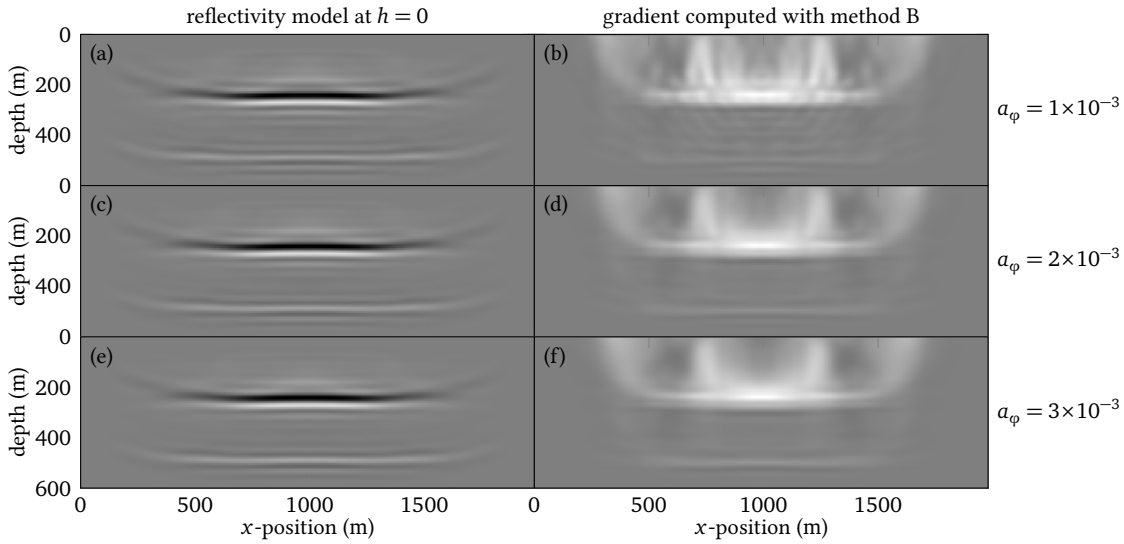


Figure 3.15. – Reflectivity sections at $h = 0$ m (left) and velocity updates (right) obtained for three increasing (from top to bottom) values of the regularisation parameter $a_\varphi = 1\times10^{-3}$, 2×10^{-3} and 3×10^{-3}. In this example, positive and negative values correspond to black and white colours, respectively.

3.4. Conclusion

Reflectivity images produced after a single step of migration are perturbed with migration artefacts and cross-talk imprint of multiples. Minimising the migration misfit function iteratively attenuates these undesirable effects and improves the properties of the MVA objective function. In this chapter, we have introduced two approximations of the gradient of the ideal MVA objective function measuring defocused energy in a reflectivity model defined as the optimal solution of the migration inverse problem. In the first case, we compute the exact gradient of an approximate objective function defined with the reflectivity model obtained after N iterations. In the second one, we derive a simpler expression by assuming that the iterative process converged and compute an approximate gradient of the ideal objective function. Provided that

sufficient regularisation is applied, both methods yield consistent velocity updates, but the method assuming convergence is much easier to implement and requires less computational effort and memory. We have provided a numerical example showing the instabilities of the gradient in a nested optimisation problem (Huang, 2016) and have analysed some reasons for them. Regularisation is essential to produce a coherent velocity update; the ℓ_2-norm is a simple but efficient choice.

Beyond the stability issues, the disadvantage of this iterative MVA process is its high computational cost: two problems have to be iteratively resolved to determine the gradient of the outer objective function updating the background velocity model. In chapter 4, the derivation of a preconditioner to accelerate the resolution of the direct and adjoint problems is investigated. The other difficulty is the empirical choice of the regularisation weight as a trade off between the smoothness of the gradient and the attenuation of migration artefacts. In particular the fact that J_1 does not converge to a stable value in the inner iterations with small regularisation contrary to J_0 (see figure 3.8) is not satisfactory. This issue will be addressed and further analysed in chapter 5.

Chapter 4.

Weighted Iterative Migration Velocity Analysis

Contents

Résumé du chapitre 4

Ce chapitre décrit l'introduction d'un préconditionneur dans l'algorithme de migration itérative. L'objectif est d'accélérer la convergence de la migration itérative et de limiter le nombre d'itérations dans la boucle interne. Le préconditionneur utilisé ici est un pseudo-inverse de l'opérateur de modélisation de Born dans le domaine étendu, défini dans le cas de primaires seuls. Cet opérateur de modélisation est linéaire et noté F. Le pseudo-inverse est noté $F^\dagger$ et est défini uniquement dans le cas de primaires seuls. L'obtention des CIGs par inversion « directe » à l'aide de cet opérateur est une alternative économique à la migration itérative pour l'atténuation des artefacts de migration. La fonction coût de DSO associée à l'inversion directe et son gradient sont étudiés et comparés à la stratégie itérative sur un exemple simple.

Le pseudo-inverse est présenté dans la section 4.2 qui est une transcription d'un article publié dans la revue *Geophysics* (Chauris et Cocher, 2017). La formule présentée ici est inspirée de celle proposée par Hou et Symes (2015). Dans les deux cas, le pseudo-inverse est défini au sens asymptotique : la formule d'inversion est obtenue en exprimant l'opérateur de modélisation à l'aide de la théorie des rais, puis en définissant des poids tels que $FF^\dagger$ soit proche de l'identité. Ces poids peuvent être interprétés à l'aide de la théorie des rais, mais la formule finale ne fait pas intervenir de quantités liées au rais. La formulation du pseudo-inverse est proche de celle de l'adjoint de l'opérateur de modélisation utilisé dans le cas de la migration classique. Des poids sont incorporés dans la formule de l'adjoint pour prendre en compte la décroissance géométrique et l'illumination inhomogène du sous-sol. En pratique les trois modifications essentielles apportées à l'implémentation de l'opérateur adjoint sont : (1) l'application aux termes sources des équations définissant les champs sources et récepteurs de dérivées par rapport à la profondeur des sources et récepteurs ; (2) l'utilisation d'une version déconvoluée de l'ondelette de source pour la construction du champ source ; (3) l'application d'une intégration temporelle d'ordre un à la place d'une dérivée seconde avant corrélation des champs sources et récepteurs, et l'application d'une dérivée verticale dans le domaine image au résultat de la corrélation de ces deux champs.

Les CIGs calculés en appliquant le pseudo-inverse aux données observées ne présentent pas d'artefacts de migration. De plus, le gradient de la fonction coût de DSO associée est plus homogène que celui obtenu dans le cas de la migration classique avec l'opérateur adjoint. Enfin une légère modification de la fonction coût de DSO consistant en une multiplication de la réflectivité par une puissance spécifique du macro-modèle permet d'atténuer les artefacts du gradient localisés à la position des réflecteurs. Ainsi remplacer l'opérateur adjoint par un pseudo-inverse pour le calcul des CIGs a des effets bénéfiques sur l'analyse de vitesse pour un coût numérique semblable. En effet le temps de calcul supplémentaire demandé par l'application des poids listés ci-dessus est négligeable devant le coût de la propagation des champs d'onde. La stratégie d'inversion « directe » est donc une alternative efficace à la migration itérative dans le cas de primaires seuls et d'une acquisition dense, mais elle n'est pas aisément transposable au cas de multiples, où il faudrait définir un nouveau pseudo-inverse dont l'existence n'est pas garantie.

La section 4.3 détaille l'introduction du pseudo-inverse comme préconditionneur dans la migration itérative et dans la résolution du problème adjoint (cas de la méthode B définie au chapitre 3). Le préconditionnement peut être vu comme un changement de variable et ne

demande que peu de modifications de l'implémentation. La section 4.4 illustre sur un exemple simple l'accélération de la convergence de la migration itérative, dans le cas de primaires seuls, mais aussi dans le cas de multiples, bien que le pseudo-inverse utilisé comme préconditionneur ne soit défini que pour les réflexions primaires. Le préconditionnement est aussi testé dans le cas de primaires seuls pour le problème adjoint. S'il permet également une résolution accélérée, celle-ci demeure plus lente que pour le problème direct, même avec régularisation. Par ailleurs le gradient calculé dans le cas itératif avec la méthode B ne converge pas vers une valeur stable similaire au gradient obtenu par l'inversion « directe ». En particulier, introduire une puissance spécifique du macro-modèle dans la définition de la fonction coût de la DSO ne permet pas d'atténuer les oscillations résiduelles localisées au niveau du réflecteur dans le gradient, contrairement au cas de l'inversion directe. Ces difficultés sont analysées dans le chapitre 5.

4.1. Introduction

We propose in this chapter a modification of the iterative migration process described in chapter 2. It consists of introducing a preconditioner in the minimisation of the migration objective function. Preconditioning can be interpreted as a change of variables allowing faster convergence. The preconditioner considered here is an approximate inverse of the extended Born modelling operator, defined for primary reflections only. Its formulation is close to the usual adjoint's one, the difference being the introduction of migration weights compensating for uneven illumination of the subsurface and for geometrical spreading. Preconditioning should result in faster attenuation of migration artefacts in CIGs. We investigate if this is also the case in the presence of multiples.

The derivation of the approximate inverse operator is presented in section 4.2, which is the transcription of an article published in Geophysics (Chauris and Cocher, 2017). Only primary reflections are considered. The inverse operator considered here is close to the one presented in Hou and Symes (2015). Both derivations determine an approximate inverse $\mathbf{F}^\dagger$ of the extended Born modelling operator $\mathbf{F}$ such that their composition $\mathbf{F}\mathbf{F}^\dagger$ is close to the identity operator. However Chauris and Cocher (2017) use a linearisation of the phase of $\mathbf{F}\mathbf{F}^\dagger$ whereas Hou and Symes (2015) directly apply the stationary phase approximation, leading to different final formulations for $\mathbf{F}^\dagger$.

In terms of implementation, the formula of Chauris and Cocher (2017) is close to the standard migration algorithm, with three main modifications: (1) vertical derivative with respect to source and receiver positions are applied to the source term in the equations defining the source and receiver wavefields; (2) the source wavefield is constructed with the deconvolved version of the source wavelet; (3) a first-order integration in time is applied before cross-correlation instead of a second-order derivative and a vertical derivative is applied to the result of the cross-correlation. These modifications amount to applying the following weights, respectively: (1) cosines of take-off angles at the sources and receivers positions; (2) deconvolution of the source wavelet; (3) cosines of the half-opening angle at the image point. A larger weight is given to small scattering angles and short surface offsets. The approach shares similarity with the quantitative extended migration presented in Lameloise *et al.* (2014). However no ray quantities are required in the final expression, which makes this strategy amenable to wave-equation based MVA techniques.

This direct inversion formula is actually an alternative to iterative migration for the case of primaries. CIGs obtained with the inverse formula are free of migration artefacts and a new MVA objective function can be constructed with these CIGs. The derivation of the associated gradient is performed with the adjoint-state technique. Numerical examples show that it is smooth and homogeneous above the reflector. Moreover, a simple modification of the MVA objective function allows to remove the oscillations around the reflector described in section 3.3.1. The inversion strategy is then an alternative to the horizontal contraction (Fei and Williamson, 2010; Shen and Symes, 2015). The advantage of the direct inversion approach is that the velocity update is the gradient of an objective function, contrary to the horizontal contraction technique.

Inversion is an efficient alternative to iterative migration in the case of primaries. Computing the reflectivity image and the associated background velocity update with inversion is as costly as with standard migration. Iterative migration is much more expensive but keeps the ability

of dealing with more complex cases such as multiples for which an inverse formula does not necessarily exists. In the second part of this chapter, we introduce migration weights in the iterative migration algorithm described in chapter 2 to improve its convergence rate. A strategy to properly include them in the optimisation process has to be defined. The inversion formula of Hou and Symes (2015) can be written as the composition of a weighting operator in the data space, the usual adjoint and a weighting operator in the model space. The iterative extension is thus natural and consists of changing the definition of the norms in the data and extended-model space (Hou and Symes, 2016a). The inversion formula used here cannot be decomposed in this way. Instead, we choose to use the approximate inverse operator as a right preconditioner. In terms of implementation, this amounts to a change of variables in the definition of the objective function of migration. Although the inverse formula does not take multiple reflections into account, it is expected to accelerate the convergence of iterative migration in the case of multiples as well as in the case of primaries only.

In the case of primaries only, direct and iterative inversions are expected to provide similar results for the reflectivity image, but not necessarily similar MVA gradients. We compare the corresponding background velocity updates obtained with both methods in the case of primaries only. Note that using method B presented in the preceding chapter, preconditioning does not change the way the gradient is computed, it only provides a more efficient way to find solutions to the direct and adjoint problems.

4.2. Definition of migration weights (Chauris and Cocher, 2017)

This section is a transcription of the following article published in Geophysics,

> H. Chauris and E. Cocher, 2017. From Migration to Inversion Velocity Analysis. Geophysics, **82**(3), S207–S223.

4.2.1. Abstract

Migration Velocity Analysis is a technique defined in the image domain to determine the background velocity model controlling the kinematics of wave propagation. In the presence of discontinuous interfaces, the velocity gradient used to iteratively update the velocity model exhibits spurious oscillations. For more stable results, we propose to replace the migration part by an inversion scheme. By definition, migration is the adjoint of the Born modelling operator, whereas inversion is its asymptotic inverse. We propose new expressions in 1D and 2D cases, based on two-way wave-equation operators. The objective function measures the quality of images obtained by inversion in the extended domain depending on the subsurface offset. In terms of implementation, the new approach is very similar to classical migration velocity analysis. A 1D analysis shows that oscillatory terms around the interface positions can be removed by multiplying the inversion result with the velocity at a specific power before evaluating the objective function. Several 2D synthetic data sets are discussed through the computation of the gradient needed to update the model parameters. Even for discontinuous reflectivity models, the new approach provides results without artificial oscillations. The model update corresponds to a gradient of an existing objective function, which was not the case for

the horizontal contraction approach proposed as an alternative to deal with gradient artefacts. It also correctly handles low velocity anomalies, contrary to the horizontal contraction approach. Inversion Velocity Analysis offers new perspectives for the applicability of image domain velocity analysis.

4.2.2. Introduction

Migration Velocity Analysis (MVA) is a technique to determine the Earth's properties from seismic surface measurements. It is based on the assumption that the model parameters can be split into a background model containing the large scale structure of the model and a model perturbation characterising the details. The background model controls the kinematics of wave propagation. Under the Born approximation, the reflected data linearly depend on the model perturbations supposed to be small compared to the background model. In a given model, the reflectivity part is obtained through migration defined as the crosscorrelation between the incident wavefield and the back-propagated residual wavefields (Claerbout, 1971). Migration is by definition the adjoint of the Born modelling operator and maps the data residuals to some model perturbations in the image domain. Once data have been migrated, the quality of the background model is evaluated by measuring the focusing of such images. Except for semblance measurements in the stacking power approach, the traditional principle consists of splitting the data into subsets, such as common shot or common offset gathers and to migrate them independently in the same background model (Symes, 2008). If all images are consistent, then the model is said to be optimal for migration. Moveout residuals observed in Common Image Gathers (CIGs) indicate that the background model is not sufficient and should be updated. The final image indeed should not depend on the acquisition geometry, at least the part correctly illuminated. Differential Semblance Optimisation (DSO) is a way to evaluate the quality of the model used for migration (Symes and Carazzone, 1991; Symes, 2008). For surface-oriented shot or offset, residual energy is computed by comparing adjacent images. Such a measure is known to have a convex behaviour, at least for simple models.

More recently, an extension has been proposed. Instead of splitting the input data into different subsets, the full data set is migrated but a spatial or temporal shift is introduced in the imaging condition (Faye and Jeannot, 1986; de Bruin *et al.*, 1990; Sava and Fomel, 2006; Symes, 2008). For a spatial shift, typically horizontal, we refer to the subsurface offset. We only consider this approach here. With the additional subsurface offset, the image domain becomes extended (Symes, 2008). The main reason for this approach is the following: the dimension of the model space should be the same as the dimension of the data space (table 4.1). The investigated model is correct if energy focuses around the zero subsurface offset. As the model and data spaces have the same dimension, it is potentially possible to reconstruct the observed data from the reflectivity, even in an inaccurate model (Symes, 2008).

MVA in the extended domain still faces a number of challenges (Lameloise *et al.*, 2014; Lameloise, 2015). (1) Due to limited acquisition, migration smiles are visible in CIGs. They always have an upward curvature (Mulder, 2014; Li *et al.*, 2016). For simple models, a downward (upward) curvature indicates a too low (high) velocity. Due to the presence of migration smiles, the optimal velocity is underestimated; in that model, there is a compromise between downward residuals and upward artefacts. (2) The objective function may also exhibit spurious oscillations

Dimension	Data domain	Space domain	Extended domain
1D	t	z	z
2D	(s,r,t)	(x,z)	(x,z,h)
3D	(s_x,s_y,r_x,r_y,t)	(x,y,z)	(x,y,z,h_x,h_y)

Table 4.1. – Dimensions of the data and space domains, with (s,r) being the source and receiver coordinates, t the time, (x,y,z) the spatial coordinates, and h the subsurface offset.

leading to local minima (Lameloise *et al.*, 2014). (3) As revealed by Vyas and Tang (2010) and Fei and Williamson (2010), the gradient of the DSO objective function with respect to the background model contains unwanted oscillations around the reflector positions, especially when the reflectivity is discontinuous. It means that MVA does not fully behave in a tomographic algorithm for which we would expect a homogeneous update at least for simple models (Sava and Biondi, 2004; Symes, 2008; Alkhalifah and Wu, 2016a). The straightforward solution is to spatially smooth the gradient, but in practice, a discontinuous reflectivity has an impact on a much larger zone than the zone containing the discontinuities. Moreover, the artefacts can have the same sign, especially around continuous interfaces: a smoothing approach would not destroy them. As a partial solution, Fei and Williamson (2010) have introduced a modification in the computation of the gradient such that spurious oscillations are removed. This spectacular effect is not fully understood yet. It is however now clear that the modified expression is not the gradient of an objective function (Shen and Symes, 2015). The main objective of this paper is to discuss possible alternatives to overcome the artefacts. Another limitation of DSO is its sensitivity to coherent noise, such as linear noise or multiples (Chauris and Noble, 2001; Mulder and ten Kroode, 2002). In practice, the input data set for MVA should only contain reflected data. Finally, operational CPU cost and memory requirement are real difficulties for the 3D implementation (Duveneck, 2013; van Leeuwen *et al.*, 2015). The main reason is the increased size of the model space, in principle up to the size of the data space (table 4.1).

In practice, only few applications on real data have been published. Among others, we can cite Symes and Carazzone (1991), Chauris and Noble (2001), Mulder and ten Kroode (2002), Alkhalifah (2005), Shen and Symes (2008), Weibull and Arntsen (2013) and Mulder (2014). Recently, approaches have been proposed to overcome some of the difficulties encountered by MVA. In particular, MVA can potentially be extended to transmitted waves (Chauris *et al.*, 2013; Shen, 2013; Biondi and Almomin, 2014; Lameloise *et al.*, 2015; Lameloise and Chauris, 2016). By considering higher orders in the Born approximation, it is also possible to take into account multiples (Staal and Verschuur, 2012; Cocher *et al.*, 2015). In these approaches, spurious oscillations related to truncated interfaces are not addressed.

To deal with this issue, we propose here to replace migration by inversion (ten Kroode, 2012; Hou and Symes, 2015). Migration is the adjoint of the Born modelling operator. It is defined as the first gradient of the least-squares misfit between modelled and observed data. An alternative is to fully solve the quadratic minimisation. The first attempts coupled to velocity analysis were obtained by iteratively determining the model perturbation (Liu *et al.*, 2014b; Chauris and Lameloise, 2014; Cocher *et al.*, 2015). The shape of the objective function appears to be modified and more quadratic; some artefacts in the gradient were also attenuated. This motivates further

investigations on iterative MVA. The main drawback with the iterative approach is the CPU-cost, proportional to the number of iterations. If we suppose that iterative migration has converged, the implementation simplifies but it is still very expensive (Cocher *et al.*, 2017a), except if iterative migration can be accelerated with suitable preconditioners (Plessix, 2009). We propose here to investigate Inversion Velocity Analysis (IVA) by coupling inversion to velocity analysis (Liu *et al.*, 2014b; Chauris *et al.*, 2015).

Inversion is a direct approach as an alternative to the iterative migration. It is remarkable that such an inversion formula exists (ten Kroode, 2012; Hou and Symes, 2015). The derivation is performed under the high frequency approximation and depends on the absence of multiple ray paths. It appears that the result is accurate, even in the presence of triplicated wavefields, as the normal operator is elliptic, i.e. asymptotically invertible, whether caustics are present or not (Stolk *et al.*, 2009; ten Kroode, 2012). The strongest limitation is the absence of turning waves, as the sub-surface offset has only a horizontal component (Biondi and Symes, 2004). The final formula only consists of a combination of wave-equation based operators without ray quantities.

The main differences between inversion and adjoint versions are additional derivative operators applied to the Green's functions or observed data at the source and receiver positions and to the reflectivity image. These operators compensate for geometrical spreading and for uneven illumination. For that, a change of variable is needed between the surface coordinates and the subsurface coordinates. It is thus essential to work in the extended image domain where precisely the model size equals the data size. The main objective of this work is to replace the adjoint result by the inverted result in MVA and to study the benefit of this modification. We first review the derivation of the Born inversion formula in 2D. It is not exactly the same as the one proposed by Hou and Symes (2015). We provide the equivalent result in 1D. Then we explain how to compute the gradient of the new IVA objective function with respect to the background velocity model. With the 1D formulation, it is possible to analyse the impact of inversion on the shape of the gradient. We propose a slightly modified inversion (multiplication of the inversion result by the velocity at a specific power) such that, at least in 1D, the gradient does not contain spurious oscillations. We then apply the same strategy in 2D, with applications on a series of synthetic data, in particular with discontinuous reflectivity functions for which artefacts have been observed (Vyas and Tang, 2010; Fei and Williamson, 2010). We focus on the shape of the gradient as it is the main ingredient to update the background model.

4.2.3. Migration versus Inversion

In Migration Velocity Analysis, the model $m(\mathbf{x}) = m_0(\mathbf{x}) + \xi(\mathbf{x})$ is split into two parts, where m_0 denotes the background model containing the large scale component of the velocity structure and where ξ is a model perturbation providing the detailed part of the model (Wu and Alkhalifah, 2015). m_0 controls the kinematics of wave propagation, whereas ξ generates reflections and diffractions. As a choice, we consider m_0 being the squared slowness model. The two components (m_0, ξ) are inverted in a nested loop approach: in a given model m_0, the inner loop updates ξ in a migration process, typically starting from $\xi = 0$. The final ξ depends on m_0. The outer loop updates m_0 according to some focusing criteria. Under the Born approximation, the modelled data d linearly depends on ξ, the model perturbation, supposed to be small in

front of m_0. The first objective function to determine the optimal ξ parameter is defined as the least-squares differences between observed data d^{obs} and computed data d in ξ

$$J_0(\xi) = \frac{1}{2}\|d(\xi) - d^{\text{obs}}\|^2. \tag{4.1}$$

Both observed and modelled data are evaluated at source $\mathbf{s}$, receiver $\mathbf{r}$ and time t or alternatively angular frequency ω. We define the Born operator $\mathscr{B}_0$ in the extended domain (Symes, 2008). With the choice of the squared slowness model, it reads

$$\mathscr{B}_0(\xi)(\mathbf{s},\mathbf{r},\omega) = -(i\omega)^2\Omega(\omega)\int d\mathbf{x}\, d\mathbf{h}\, G_0(\mathbf{s},\mathbf{x}-\mathbf{h},\omega)\xi(\mathbf{x},\mathbf{h})G_0(\mathbf{x}+\mathbf{h},\mathbf{r},\omega), \tag{4.2}$$

where Ω is the input source wavelet, typically a Ricker function. The two Green's function G_0 satisfy the wave equation, here the constant density acoustic wave equation in the model m_0. The associated source term is a Dirac distribution in space and time. Extended domain means that the model perturbation ξ depends on the spatial coordinates $\mathbf{x} = (x, z)$ in 2D and on an extra parameter, here a spatial horizontal shift $\mathbf{h} = (h, 0)$ in 2D (Sava and Fomel, 2006; Symes, 2008; Sava and Vasconcelos, 2011). Compared to classical Born modelling, an extra loop on $\mathbf{h}$ allows to consider non-physical models in the sense that the two Green's functions are evaluated at two distinct positions $\mathbf{x}-\mathbf{h}$ and $\mathbf{x}+\mathbf{h}$.

The first strategy to determine ξ consists of defining it as the first gradient of J_0 with respect to ξ and for $\xi = 0$. This is the classical migration approach, here formulated in the extended domain, yielding

$$(\mathscr{B}_{\text{mig}} d^{\text{obs}})(\mathbf{x},\mathbf{h}) = -\int d\mathbf{s}\, d\mathbf{r}\, d\omega\, (i\omega)^2\Omega^*(\omega)G_0^*(\mathbf{s},\mathbf{x}-\mathbf{h},\omega)d^{\text{obs}}(\mathbf{s},\mathbf{r},\omega)G_0^*(\mathbf{x}+\mathbf{h},\mathbf{r},\omega). \tag{4.3}$$

This adjoint formulation implies the complex conjugate version G_0^* of the Green's functions and an integration over the acquisition parameters $(\mathbf{s},\mathbf{r},\omega)$. The second strategy consists of minimising J_0 and to define ξ through the inverse operator $\mathscr{B}_{\text{inv}}$ (ten Kroode, 2012; Hou and Symes, 2015). Two aspects are essential in the derivation of such operator: first, the extended domain plays a crucial role as the size of the data domain is the same as the size of the model domain. During the derivation, a change of variables is applied. The second element is the microlocal analysis, meaning that $\mathscr{B}_{\text{inv}}$ is the inverse of $\mathscr{B}_0$ under high frequency approximation and some additional approximations, mainly that there is no triplicated wavefield. The inverse operator should compensate for geometrical spreading and uneven illumination. A general form inspired from (Hou and Symes, 2015) is

$$\mathscr{B}_{\text{inv}}(\delta d)(\mathbf{x},\mathbf{h}) = k(\mathbf{x},\mathbf{h})\mathscr{D}_p \int d\mathbf{s}\, d\mathbf{r}\, d\omega (i\omega)^\nu \widetilde{\Omega}(\omega)\mathscr{D}_{s_z} G_0^*(\mathbf{s},\mathbf{x}-\mathbf{h},\omega)\delta d(\mathbf{s},\mathbf{r},\omega)$$
$$\mathscr{D}_{r_z} G_0^*(\mathbf{x}+\mathbf{h},\mathbf{r},\omega), \tag{4.4}$$

where $\mathscr{D}_p$ is the partial derivative with respect to variable p. For example, $\mathscr{D}_{s_z}$ indicates the vertical derivative with respect to the source position. The values of p in $\mathscr{D}_p$, ν in $(i\omega)^\nu$ and $k(\mathbf{x},\mathbf{h})$ in front of the integral need to be specified such that the composition of the two operators

$\mathscr{B}_{\text{inv}} \cdot \mathscr{B}_0 = \mathscr{I}$, where $\mathscr{I}$ is the identity operator. Let $\mathscr{F}\xi$ be the integrand of the product of the two operators. It is linear in ξ and satisfies

$$(\mathscr{B}_{\text{inv}} \cdot \mathscr{B}_0)\xi(\mathbf{y}) = \int \mathrm{d}\mathbf{y}' \mathscr{F}(\mathbf{y}, \mathbf{y}')\xi(\mathbf{y}'), \tag{4.5}$$

with $\mathbf{y} = (\mathbf{x}, \mathbf{h})$. $\mathscr{B}_{\text{inv}}$ is indeed an inverse if $\mathscr{F}(\mathbf{y}, \mathbf{y}') = \delta(\mathbf{y} - \mathbf{y}')$. After derivations reported in appendix 4.2.10, the final expression for the 2D case reads

$$(\mathscr{B}_{\text{inv}}\delta d)(\mathbf{x}, \mathbf{h}) = 32\sqrt{m_{0-}m_{0+}} \\ \mathscr{D}_p \left\{ \int \frac{\widetilde{\Omega}(\omega)}{\mathrm{i}\omega} \mathscr{D}_{s_z} \mathrm{G}_0^*(\mathbf{s}, \mathbf{x} - \mathbf{h}, \omega) \delta d(\mathbf{s}, \mathbf{r}, \omega) \mathscr{D}_{r_z} \mathrm{G}_0^*(\mathbf{x} + \mathbf{h}, \mathbf{r}, \omega) \, \mathrm{d}\mathbf{s} \, \mathrm{d}\mathbf{r} \, \mathrm{d}\omega \right\}, \tag{4.6}$$

with

$$\mathscr{D}_p = \frac{1}{2}\left(\sqrt{\frac{m_{0+}}{m_{0-}}} + 1 \right) \mathscr{D}_z + \frac{1}{2}\left(\sqrt{\frac{m_{0-}}{m_{0+}}} - 1 \right) \mathscr{D}_{hz}, \tag{4.7}$$

where $\widetilde{\Omega}$ is the inverse of Ω, $m_{0+} = m_0(\mathbf{x} - \mathbf{h})$, $m_{0-} = m_0(\mathbf{x} + \mathbf{h})$. $\mathscr{D}_z$ is the vertical derivative at the image point, whereas $\mathscr{D}_{hz}$ is the derivative with respect to the vertical subsurface offset. It is important to note that the final expression does not contain ray quantities. In an invariant model, $m_{0-} = m_{0+} = m_0$ and $\mathscr{D}_p = \mathscr{D}_z$. This result is not exactly the same as the one published in Hou and Symes (2015), as explained in the Discussion section. The main differences between the adjoint and inverse versions are the following: (1) instead of a second derivative of the seismic source, a first-order integration in time has to be applied to the inverse of the source function; (2) a normalisation factor $32\sqrt{m_{0-}m_{0+}}$ has to be applied on the reflectivity to retrieve absolute amplitudes; (3) the vertical derivatives with respect to the source and receiver positions are applied to the Green's function. The main dynamic effect is a multiplication by the cosine weight at the source and receiver positions; (4) similarly, for laterally invariant model, an additional vertical derivative has to be applied at the image points. It acts as if the kernel would be multiplied by $\cos\theta \cos\phi$, where θ and ϕ are the half-opening and dip angles at the image point, measured from the normal to the dip and the horizontal, respectively (see figure C.1 p. 265 for the definition of these angles). Such weights emphasise short offsets, small dips and vertical rays at source and receiver positions. Only wave-equation based operators are present in the definition of the inverse operator $\mathscr{B}_{\text{inv}}$.

Following the same strategy, we propose equivalent and new expressions for the 1D case. This case will be useful in the next section to analyse the shape of the gradients with respect to m_0.

$$(\mathscr{B}_0\xi)(\omega) = -(\mathrm{i}\omega)^2 \Omega(\omega) \int_z \mathrm{G}_0^2(z, \omega)\xi(z) \, \mathrm{d}z, \tag{4.8a}$$

$$(\mathscr{B}_{\text{mig}}\delta d)(z) = -\int_\omega (\mathrm{i}\omega)^2 \Omega^*(\omega) \mathrm{G}_0^{2*}(z, \omega)\delta d(\omega) \, \mathrm{d}\omega, \tag{4.8b}$$

$$(\mathscr{B}_{\text{inv}}\delta d)(z) = 16 m(z) \mathscr{D}_z \left\{ \int_\omega \frac{\widetilde{\Omega}(\omega)}{(\mathrm{i}\omega)} [\mathscr{D}_{s_z} \mathrm{G}_0^*]^2(z, \omega)\delta d(\omega) \, \mathrm{d}\omega \right\}. \tag{4.8c}$$

In 1D, there is no need to extend the model. We consider a single source and a single receiver, both at the surface. The source and receiver Green's functions are the same. We recognise the similar terms. The main difference with the 2D case is the factor $16m(z)$ in front of the vertical derivative in equation (4.8c). In the case of a homogeneous model m_0 and an original reflectivity function defined as $R_e\delta(z-z_e)$, for a single reflector at depth z_e, the expressions can be further simplified using analytic formulations for the Green's functions, yielding in 1D

$$\xi_{\text{mig}}(z) = R_e\left(\frac{v_0}{2}\right)^2\left(\frac{v_e}{2}\right)^2 g_{\text{mig}}\left[2\left(\frac{z_e}{v_e}-\frac{z}{v_0}\right)\right], \tag{4.9a}$$

$$\xi_{\text{inv}}(z) = 2\frac{R_e}{v_0}\left(\frac{v_e}{v_0}\right)^2 g_{\text{inv}}\left[2\left(\frac{z_e}{v_e}-\frac{z}{v_0}\right)\right], \tag{4.9b}$$

where $g_{\text{mig}}(t)$ and $g_{\text{inv}}(t)$ are obtained by crosscorrelating the seismic wavelet with itself or with its inverse, namely $g_{\text{mig}}(t) = \Omega(t)\otimes\Omega(t)$ and $g_{\text{inv}}(t) = \Omega(t)\otimes\widetilde{\Omega}(t)$. The homogeneous velocities values are v_e and v_0, respectively in the exact and tested cases. In 1D homogeneous models, the adjoint and inverse expressions are very similar. For the inverse case, a deconvolution is applied; the same time shift appears in the two expressions. A different normalisation term appears. Note that there is no geometrical spreading and no need to compensate for uneven illumination.

We now have the expressions in 1D and 2D for the adjoint and inverse approaches. We introduce them in a velocity analysis approach.

4.2.4. MVA versus IVA

Two nested loops are defined in image domain velocity analysis. The inner loop determines the reflectivity model ξ_{mig} or ξ_{inv} through the J_0 objective function (equation 4.1). The second objective function for the outer loop indicates how focused the energy is in the extended domain. For the adjoint and inverse formulation, we define

$$J^{\alpha}_{\text{mig}}(m_0) = \frac{1}{2}\left\| m_0^{\alpha}|\mathbf{h}|\xi_{\text{mig}}\right\|^2, \tag{4.10a}$$

$$J^{\beta}_{\text{inv}}(m_0) = \frac{1}{2}\left\| m_0^{\beta}|\mathbf{h}|\xi_{\text{inv}}\right\|^2. \tag{4.10b}$$

The multiplication by the annihilator $|\mathbf{h}|$ aims at evaluating how energy is concentrated around $\mathbf{h}=0$ (Sava and Fomel, 2006). For the correct model m_0, energy is expected to be localised in the physical domain. Extra terms m_0^{α} and m_0^{β} have been introduced inside the ℓ_2 norm. In the classical formulation, $\alpha=\beta=0$. They should be interpreted as an additional weight for more flexibility in the shape of the gradient of J^{α}_{mig} and J^{β}_{inv} with respect to m_0. As explained later, the additional term in the gradient for α or β if not equal to zero is localised around the reflectivity: this will be used to attenuate oscillations in the final gradient. We use the adjoint state technique (Plessix, 2006) as an efficient tool for the gradient derivation. More details are provided in appendix 4.2.11. The adjoint state technique can be seen as a minimisation process under constraints. Instead of only considering m_0 as unknown, we enlarge the model space by incorporating other variables. We add Lagrangian constrains to force the wavefield to

satisfy the wave equation. For an efficient implementation, we define the following variables S_0, R_0, μ_s, μ_r and λ (appendix 4.2.11): S_0 and R_0 are the forward and backward wavefields, whereas μ_s and μ_r are their adjoint variables, respectively. The final gradient for $\alpha = 0$ and $\beta = 0$ reads

$$\frac{\partial J^0_{\text{mig}}}{\partial m_0} = -\int (i\omega)^2 S_0(\mathbf{s},\mathbf{x},\omega)\mu_s^*(\mathbf{s},\mathbf{x},\omega)\,d\mathbf{s}\,d\omega - \int (i\omega)^2 R_0(\mathbf{s},\mathbf{x},\omega)\mu_r^*(\mathbf{s},\mathbf{x},\omega)\,d\mathbf{s}\,d\omega, \tag{4.11a}$$

$$\begin{aligned}\frac{\partial J^0_{\text{inv}}}{\partial m_0} = &-\int (i\omega)^2 S_0(\mathbf{s},\mathbf{x},\omega)\mu_s^*(\mathbf{s},\mathbf{x},\omega)\,d\mathbf{s}\,d\omega - \int (i\omega)^2 R_0(\mathbf{s},\mathbf{x},\omega)\mu_r^*(\mathbf{s},\mathbf{x},\omega)\,d\mathbf{s}\,d\omega \\ &+ \int \lambda(\mathbf{x},\mathbf{h})\frac{\partial \mathscr{D}_p}{\partial m_0}\xi_0(\mathbf{x},\mathbf{h})\,d\mathbf{h},\end{aligned} \tag{4.11b}$$

with $\xi_{\text{inv}}(\mathbf{x},\mathbf{h}) = \mathscr{D}_p\xi_0(\mathbf{x},\mathbf{h})$. $\mathscr{D}_p$ is defined in equation (4.7). In laterally invariant models, $\mathscr{D}_p$ does not depend on m_0. The adjoint variable $\lambda_0 = a^2\xi_{\text{mig}}$ in the adjoint case and $\lambda_0(\mathbf{x},\mathbf{h}) = \mathscr{D}_p^* a^2 \mathscr{D}_p \xi_0$ in the inversion case (appendix 4.2.11). μ_s and μ_r are solution of the wave equation for a source term excited by the product of λ_0 by the back-propagated or forward wavefields (figure 4.1). In the gradient expression (equations 4.11a and 4.11b), S_0 and μ_s contribute in a tomographic mode as they correlate from any interface to the surface. The same holds for R_0 and μ_r.

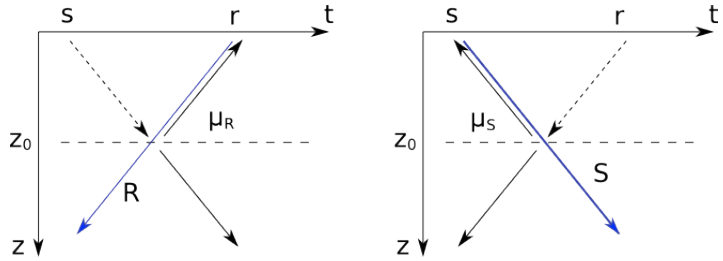


Figure 4.1. – Schematic view of the wavefields (left: from the source, right: from the receiver) contributing to the gradient in the presence of a single reflector at depth z_0. Both S and μ_s and R and μ_r correlate between the surface and depth z_0.

For non zero α and β values, the gradient expressions are simply obtained as a combination of the gradient for α or $\beta = 0$ and a positive term located around the position of the reflectors

$$\frac{\partial J^\alpha_{\text{mig}}}{\partial m_0}(\mathbf{x}) = m_0^{2\alpha}(\mathbf{x})\frac{\partial J^0_{\text{mig}}}{\partial m_0}(\mathbf{x}) + 2\alpha m_0^{2\alpha-1}(\mathbf{x})\int d\mathbf{h}\,|\mathbf{h}|^2\xi^2_{\text{mig}}(\mathbf{x},\mathbf{h}), \tag{4.12a}$$

$$\frac{\partial J^\beta_{\text{inv}}}{\partial m_0}(\mathbf{x}) = m_0^{2\beta}(\mathbf{x})\frac{\partial J^0_{\text{inv}}}{\partial m_0}(\mathbf{x}) + 2\beta m_0^{2\beta-1}(\mathbf{x})\int d\mathbf{h}\,|\mathbf{h}|^2\xi^2_{\text{inv}}(\mathbf{x},\mathbf{h}). \tag{4.12b}$$

We have seen that the gradient of the second objective function can thus be efficiently derived with the adjoint state technique. The adjoint approach contains 3 terms whereas the inverse 4 contributions. The two strategies have very similar implementation and the inverse is not really more difficult. In the next section, we analyse the different contributions to understand

if they act in a tomographic mode and/or they have an oscillating contribution around the reflector position. This leads to the determination of an optimal β value. Then we present 2D applications with the same β.

4.2.5. Analysis of the gradient in 1D

The 1D case is a particular case. Instead of defining a spatial shift h introduced in the imaging condition, we propose to define the annihilator as $z - z_e$, where z_e is the exact depth of the reflector. This is only valid for a single reflector. However, the expressions are similar to the ones in 2D and help us understand the structure of the gradient. For homogeneous exact v_e and initial v_0 models, the expressions for the 1D adjoint and inverse gradients are

$$\begin{aligned} G^{\alpha}_{\text{mig}}(z) = &- \frac{R_e^2 v_e^4}{2^8 v_0^{2\alpha-5}} \int (y-z_e)^2 g_{\text{mig}}\left[2\left(\frac{z_e}{v_e}-\frac{y}{v_0}\right)\right] g'_{\text{mig}}\left[2\left(\frac{z_e}{v_e}-\frac{z+y+|z-y|}{2v_0}\right)\right] \mathrm{d}y \\ &+ \alpha \frac{R_e^2 v_e^4}{2^6 v_0^{2\alpha-4.5}} (z-z_e)^2 g^2_{\text{mig}}\left[2\left(\frac{z_e}{v_e}-\frac{z}{v_0}\right)\right] \end{aligned} \tag{4.13}$$

$$\begin{aligned} G^{\beta}_{\text{inv}}(z) = &\frac{4R_e^2}{v_0^{2\beta+1}}\left(\frac{v_e}{v_0}\right)^4 \int (y-z_e)^2 g_{\text{inv}}\left[2\left(\frac{z_e}{v_e}-\frac{y}{v_0}\right)\right] g'_{\text{inv}}\left[2\left(\frac{z_e}{v_e}-\frac{z+y+|z-y|}{2v_0}\right)\right] \mathrm{d}y \\ &+ 2(\beta+2)\frac{R_e^2}{v_0^{2\beta}}\left(\frac{v_e}{v_0}\right)^4 (z-z_e)^2 g^2_{\text{inv}}\left[2\left(\frac{z_e}{v_e}-\frac{z}{v_0}\right)\right] \\ &+ \frac{4R_e^2}{v_0^{2\beta}}\left(\frac{v_e}{v_0}\right)^4 \int (y-z_e) g_{\text{inv}}\left[2\left(\frac{z_e}{v_e}-\frac{y}{v_0}\right)\right] g_{\text{inv}}\left[2\left(\frac{z_e}{v_e}-\frac{z+y+|z-y|}{2v_0}\right)\right] \mathrm{d}y. \end{aligned} \tag{4.14}$$

The first term in G^{α}_{mig} has a non zero contribution above the position of the reflector after migration, while the second term is positive and localised around the reflector.

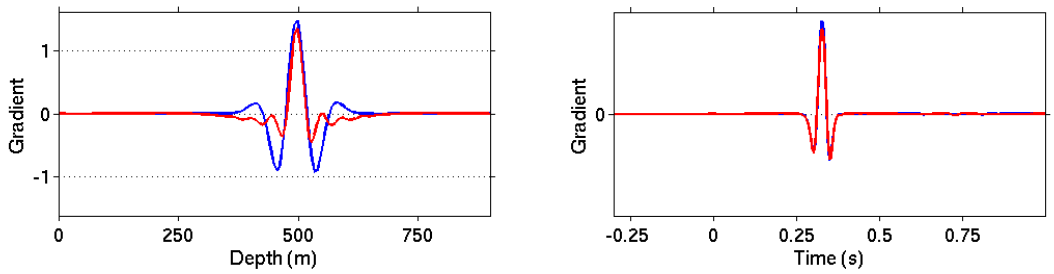


Figure 4.2. – Left: reflectivity sections ξ_{mig} (blue) and ξ_{inv} (red) for $h = 0$. Right: observed (blue) and computed data after inversion and modelling (red).

We use a classical finite-difference modelling tool to solve the constant density acoustic wave equation. Perfectly Matched Layers (PMLs) are implemented to deal with model boundaries, with no free-surface condition. The reflectivity model consists of a single interface at 500 m depth. After migration and inversion using the exact model $v_0 = v_e = 3000\,\text{m/s}$, the reflector is localised around the correct depth (figure 4.2, left). The deconvolution aspect in the inversion

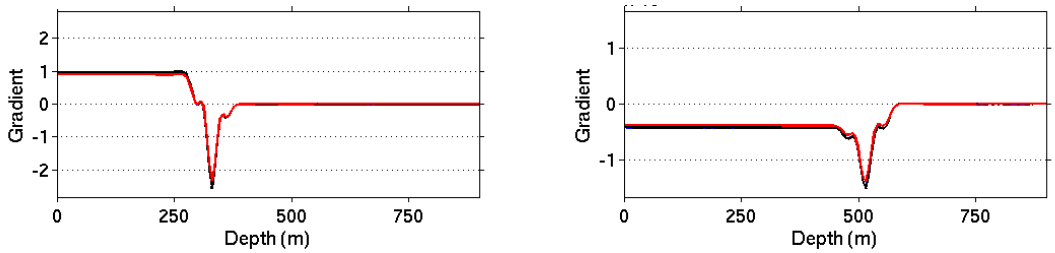


Figure 4.3. – Gradients for the migration approach. The blue, red and black curves (almost superimposed) are respectively computed with the adjoint, finite-difference and analytic approaches. In all cases, $\alpha = 0$ Left: $v_0 = 2500\,\mathrm{m/s}$ and $v_e = 3000\,\mathrm{m/s}$. Right: same for $v_0 = 3000\,\mathrm{m/s}$ and $v_e = 2500\,\mathrm{m/s}$.

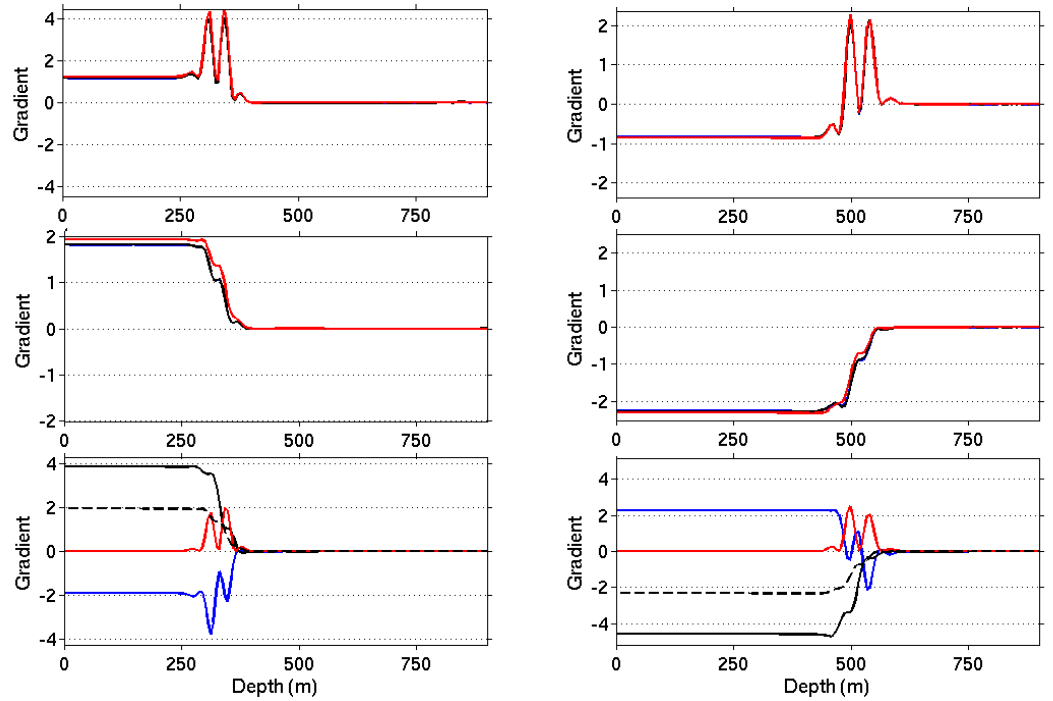


Figure 4.4. – Same as for Figure 4.3, but here for the inversion approach (top: $\beta = 0$ and middle: $\beta = -3/2$). Bottom: decomposition of the gradient G_{inv} (dashed line) for $\beta = -3/2$ into the three contributions G^1_{inv} (blue), G^2_{inv} (red) and G^3_{inv} (black).

approach makes the signal more localised in space. From ξ_{inv}, the modelled data nicely matches with the observed data (figure 4.2, right). Both phase and amplitude are correctly retrieved.

For the gradient, we compare three approaches: (1) the inverse approach obtained with the adjoint state method (equation 4.11b), (2) a finite-difference approach and (3) the analytic expression valid in homogeneous background models (equation 4.14). The finite-difference approach consists of evaluating the gradient by computing J^{β}_{inv} for slightly perturbed models around the reference model. This is a much more expensive approach than the one provided by the adjoint state method but this is affordable in 1D. We first consider $\beta = 0$. The three approaches give very similar results for the migration approach (figure 4.3), both if the initial model has a too low and too high value (2500 or 3000 m/s for a correct velocity at 3000 m/s or 2500 m/s, respectively). The gradient is homogeneous above the reflector, while oscillations are present around the reflector. In this case, the amplitudes of these oscillations are larger than the homogeneous part. We conclude that the DSO gradient in 1D has a tomographic mode together with a migration mode around the reflectivity. The additional contribution requires in practice to smooth the gradient before updating the model.

The same results hold for the inversion approach, still for $\beta = 0$ (figure 4.4, top). It is interesting to note that the oscillations around the reflectivity are always positive, even if the initial model is lower or larger than the exact model. The three approaches (adjoint, finite-difference and analytic) are also consistent. The α and β values introduced in the objective function offer additional possibilities. Appendix 4.2.12 indicates that there is an optimal $\beta = -3/2$ value for which the gradient becomes much smoother (figure 4.4, middle). An alternative would be to optimise the shape of the gradient as in Wu and Alkhalifah (2015). To further analyse this, we first display the three contributions to the gradient in 1D (figure 4.4, bottom). The red curve is always positive by definition as it contains only squared terms (second term in equation 4.12b). It appears that the oscillations in the first term are in phase, with an opposite sign, to the second contribution (blue line). The third contribution is smooth (black line), leading as a final result to a smooth gradient (dashed line).

This effect is confirmed by analysing more carefully the expression of the gradient in homogeneous models (equation 4.14). The integration over y is performed from 0 to the maximum depth z_{max}. We split this integration between 0 to z and then between z to z_{max}. We then integrate by parts. For $\beta = -3/2$, the oscillating contributions are exactly opposite. More details are provided in appendix 4.2.12. The same analysis with the gradient associated to migration does not lead to a smooth gradient. We conclude that in 1D, the gradient related to inversion can have a smooth behaviour for a particular normalisation ($\beta = -3/2$). We now investigate if these conclusions can be extended to 2D.

4.2.6. 2D applications

We study the shape of the gradient of J^{α}_{mig} and J^{β}_{inv} with respect to the background model and for $\alpha = 0$, $\beta = 0$ and $\beta = -3/2$. We investigate five 2D synthetic cases. The first one contains a single horizontal reflector to study (1) how the inversion impacts the shape of the gradient, and (2) if $\beta = -3/2$ is still a good candidate as in the 1D case. Then, we deal with discontinuous reflectivity sections, first considering a single diffraction point and then interfaces with abrupt truncations. They are typical cases for which the classical gradient exhibits spurious oscillations

(Fei and Williamson, 2010; Vyas and Tang, 2010). The fourth case contains a complex rough interface between two homogeneous structures. Here, the observed data are not modelled under the Born approximation. A pre-processing is needed to remove the direct arrival. Finally, the last case contains a low velocity anomaly in which the horizontal contraction solution does not provide a proper solution as explained in Shen and Symes (2015). In this example, we perform non-linear iterations to update the velocity model; in other cases, we display the first velocity update. In all cases, the data are generated with a finite-difference time domain constant density acoustic code under the Born approximation (Noble, 1992). Even if the velocity model is not invariant, we simplify $\mathscr{D}_p$ to $\mathscr{D}_z$.

Single horizontal interface

In the first model, we consider a single horizontal interface at depth $z = 240\,\text{m}$ in the exact model corresponding to a homogeneous model $v_e = 3000\,\text{m/s}$. Sources and receivers are located at the surface, with a maximum offset of $\pm 600\,\text{m}$. The maximum frequency content of the data is 40 Hz.

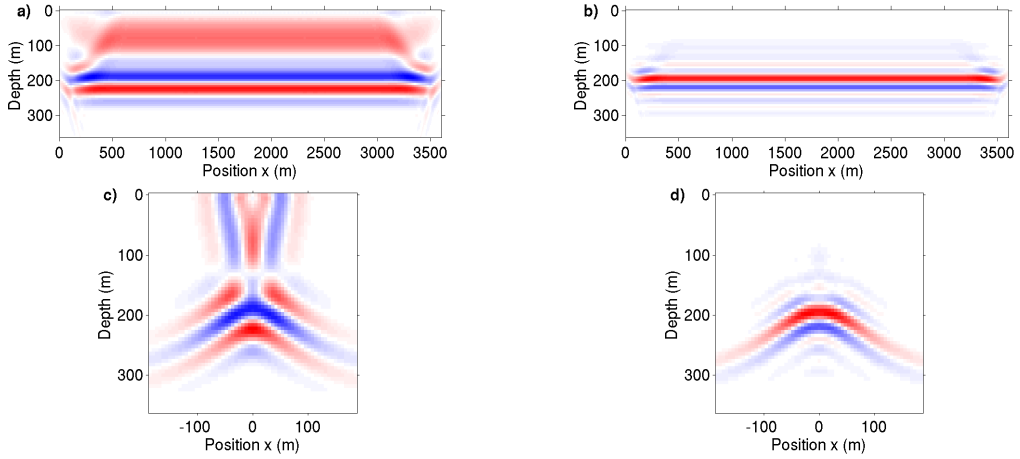


Figure 4.5. – Reflectivity sections ξ_{mig} (a) and ξ_{inv} (b) for $h = 0$ and for $v_e = 3000\,\text{m/s}$ and $v_0 = 2500\,\text{m/s}$, in the case of a single horizontal interface, and associated CIGs at position $x = 1800\,\text{m}$ for the migration (c) and inverse (d) cases.

The classical migrated section in a homogeneous model at $v_0 = 2500\,\text{m/s}$ exhibits a flat interface around $z = 290\,\text{m}$ and low frequency variations above it (figure 4.5a). This can be understood by looking at the CIGs (figure 4.5c). For this specific depth interface and maximum surface offsets, migration smiles are present off the zero-subsurface offset. The shape of such events are predicted by Mulder (2014). With the application of the inverse instead of the adjoint, events are more localised due to the deconvolution in the inverse formula, but more importantly migration smiles are removed: defocused energy corresponds to an incorrect velocity model (figures 4.5b and 4.5d).

The associated gradient is not homogeneous above the interface for the migration case (figure 4.6a) due to the contribution of migration smiles. As their curvature is always towards the surface, a positive (red) contribution is added to the gradient. Note that the gradients are

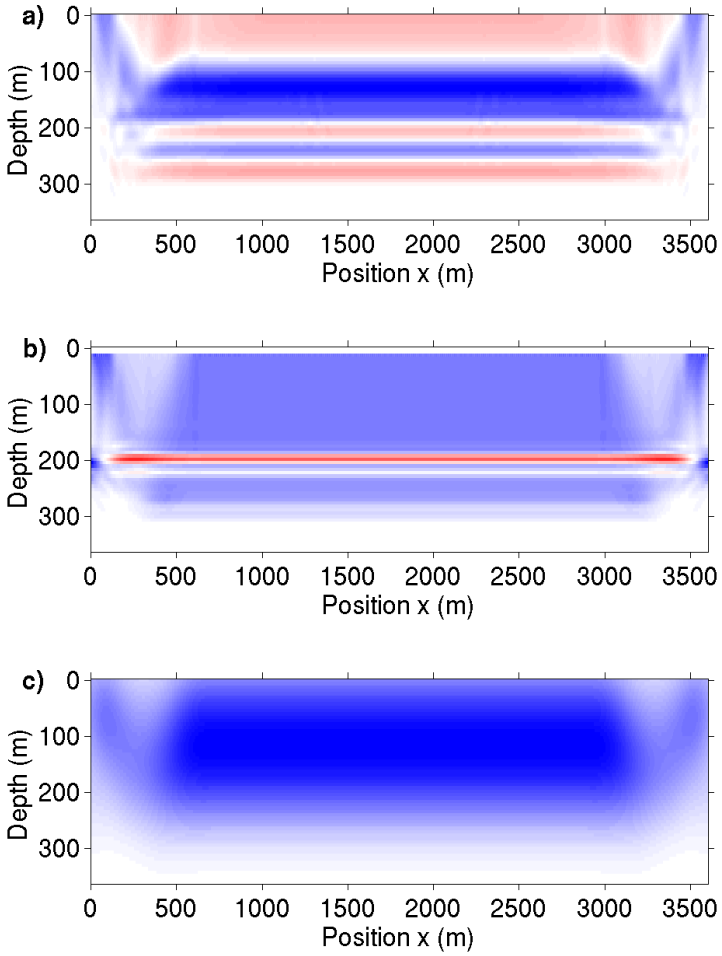


Figure 4.6. – Gradients for the migration (a) and inversion (b and c) approaches, for $\beta = 0$ (b) and $\beta = -3/2$ (c).

computed with respect to the squared slowness: we thus expect only negative (blue) values above the interface. This is partly the case for the inversion gradient and $\beta = 0$ (figure 4.6b), but the impact of the reflectivity around $z = 200\,\text{m}$ is visible and adds a positive (red) contribution to the gradient. This effect was already observed in the 1D case (figure 4.4). A closer look shows that the gradient has also a negative contribution below the interface: this could appear to be strange, but can be explained as follows: energy is visible in CIGs for depths below $z = 200\,\text{m}$ and non-zero subsurface offsets as the velocity model is too slow (figure 4.5d). The contribution of the gradient below that depth is due to the third term in equation (4.11b). This was not visible in 1D for which there is no need to rely on an extended model. With the introduction of $\beta = -3/2$, oscillations around the interface and the smooth update below disappear. As for the 1D case, the second term in equation (4.12b) cancels unwanted oscillations for $\beta = -3/2$, even if there is no formal proof in 2D.

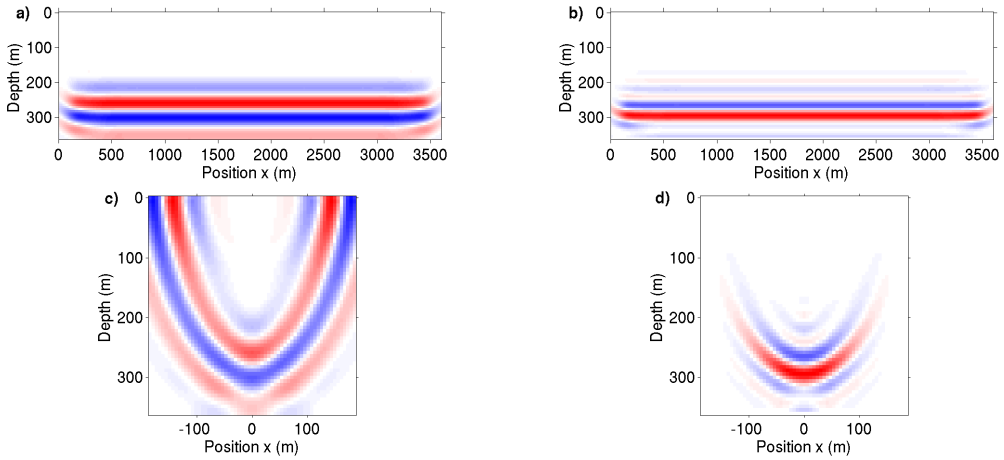


Figure 4.7. – Same as for figure 4.5, but here with $v_e = 2500\,\text{m/s}$ and $v_0 = 3000\,\text{m/s}$.

When the velocity v_0 is higher than the exact velocity v_e, the same conclusions hold (figures 4.7 and 4.8). CIGs could appear to be cleaner in the migration case but energy due to an incorrect model is partly superimposed with migration smiles (figure 4.7a): the associated gradient indeed oscillates with positive and negative values above the interface (figure 4.7c). With the inversion approach, CIGs and gradients are cleaner (figures 4.7d and 4.8c). For $\beta = 0$, the gradient has a positive (red) contribution around the reflector as in 1D and as for the previous case. For $\beta = -3/2$, the gradient is homogeneous above the interface. Note that in figures 4.6c and 4.8c, edge effects for $x = 0$ and $x = 3500\,\text{m}$, are visible with lower energy due to limited acquisition.

We conclude from this first test that the inversion scheme with $\beta = -3/2$ provides a smooth gradient as in a tomographic mode, at least for continuous interfaces.

Diffraction model

The extreme discontinuous case is a model with a single diffraction, here embedded in a homogeneous model $v_e = 3000\,\text{m/s}$. The maximum surface offset is $\pm 600\,\text{m}$. The exact diffraction anomaly is at $(x, z) = (1800\,\text{m}, 240\,\text{m})$. For a too slow velocity model $v_0 = 2500\,\text{m/s}$, energy for $h = 0$ follows a diffraction curve, but energy is visible above it in the migration case

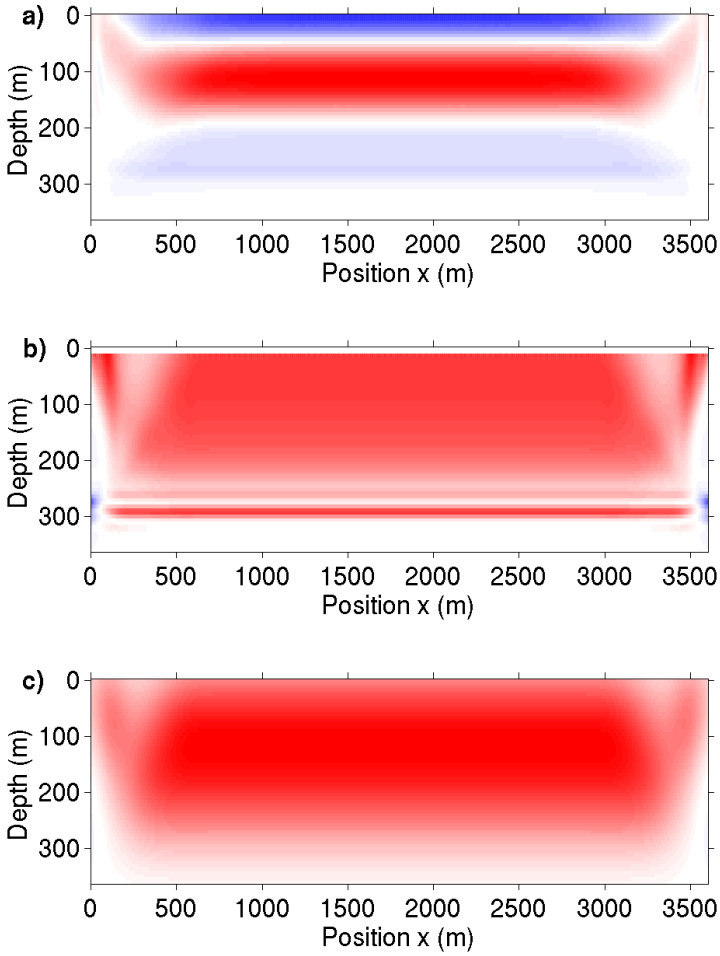


Figure 4.8. – Same as for Figure 4.6, but here with $v_e = 2500\,\mathrm{m/s}$ and $v_0 = 3000\,\mathrm{m/s}$.

(figure 4.9a).

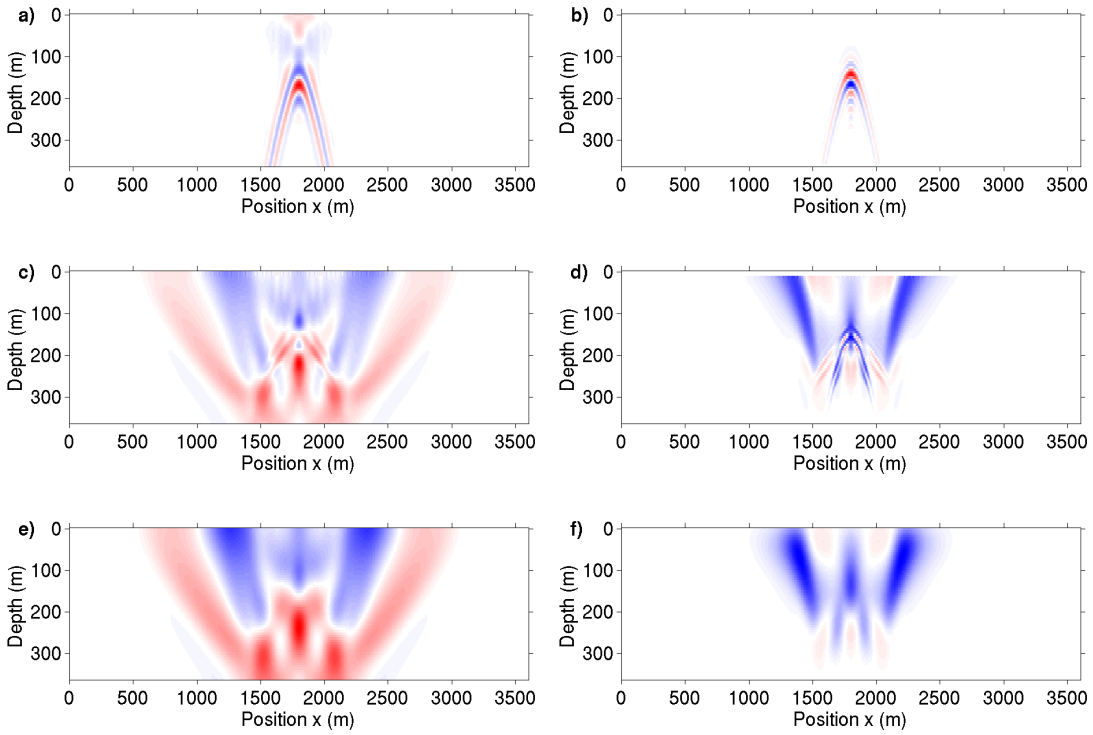


Figure 4.9. – Single diffraction case, with the migration (left) and inversion (right) approaches. Top: reflectivity section, middle: velocity gradient, bottom: smoothed version of the velocity gradient.

Both for migration and inversion, the gradient is oscillating (figures 4.9c and 4.9d). We propose to smooth it as a regularisation before iterating over the background model. The Gaussian smoothing filter is expressed as

$$g_\sigma(r_0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{r_0^2}{2\sigma^2}}, \tag{4.15}$$

where r_0 is the distance. The smoothing parameter is chosen as $\sigma = \lambda_{\text{data}}/2$, with λ_{data} being the mean wavelength of the data. Only negative values are obtained for the inversion case, but not for the migration case (figures 4.9e and 4.9f).

Discontinuous interfaces

The third example consists of 4 truncated interfaces inspired from Shen and Symes (2015). The maximum surface offset and frequency content of the data are the same as for the diffraction case. The smoothing parameter σ equals the half-wavelength. Once more, the migrated reflectivity contains energy above the interface positions, leading to oscillatory gradients, especially around the termination of interfaces (figures 4.10a, 4.10c and 4.10e). Inversion results

are much sharper (figure 4.10b). The gradient is negative (blue) everywhere, even without smoothing (figures 4.10d and 4.10f). Note that the reflectivity sections (figure 4.10) can be reconstructed from the diffraction case (figure 4.9) as there is a linear relationship between ξ_{mig} and the exact reflectivity. This is not true for the gradient as it is not linear in ξ_{mig}. The same remark holds for the inversion case: figure 4.10f cannot be reconstructed as a linear combination of gradients associated to a single diffraction point as in figure 4.9f.

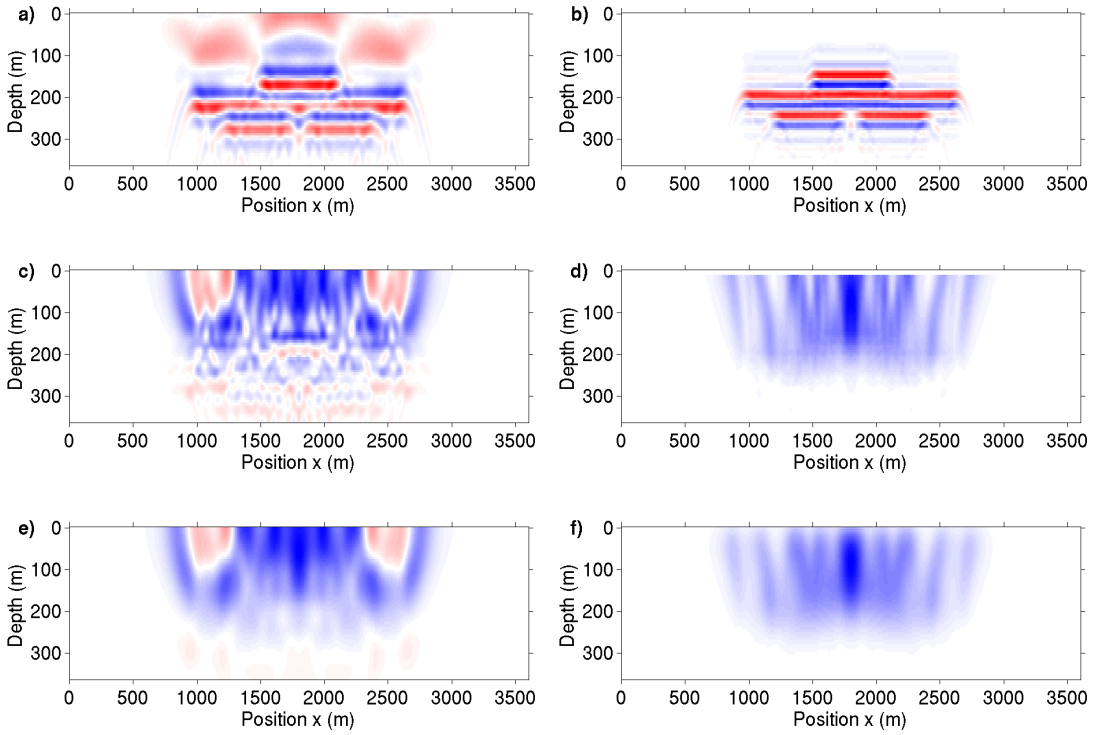


Figure 4.10. – Multi-layer case, with the migration (left) and inversion (right) approaches. Top: reflectivity section, middle: velocity gradient, bottom: smooth version of the velocity gradient.

Rough interface

In the next example, the model is extracted from the BP salt dome model (Billette and Brandsberg-Dahl, 2005). Here the velocity above the rough interface is set to 3000 m/s and below to 3800 m/s (figure 4.11a). As before, the observed data is computed with a finite-difference modelling code, but here without relying on the Born approximation. Absorbing boundaries are implemented via Perfect Matched Layers (Bérenger, 1994; Komatitsch and Martin, 2007), including for the top interface. As a pre-processing, the direct arrival is removed. The maximum surface offset is ±600 m and the maximum frequency content of the data is 40 Hz.

We compare the result obtained with the migration and inverse approaches. For the migration case obtained for $v_0 = 2500$ m/s, the interface is globally too shallow but also defocused and contains low frequency energy above the interface. As for the simple case (figure 4.6a), the

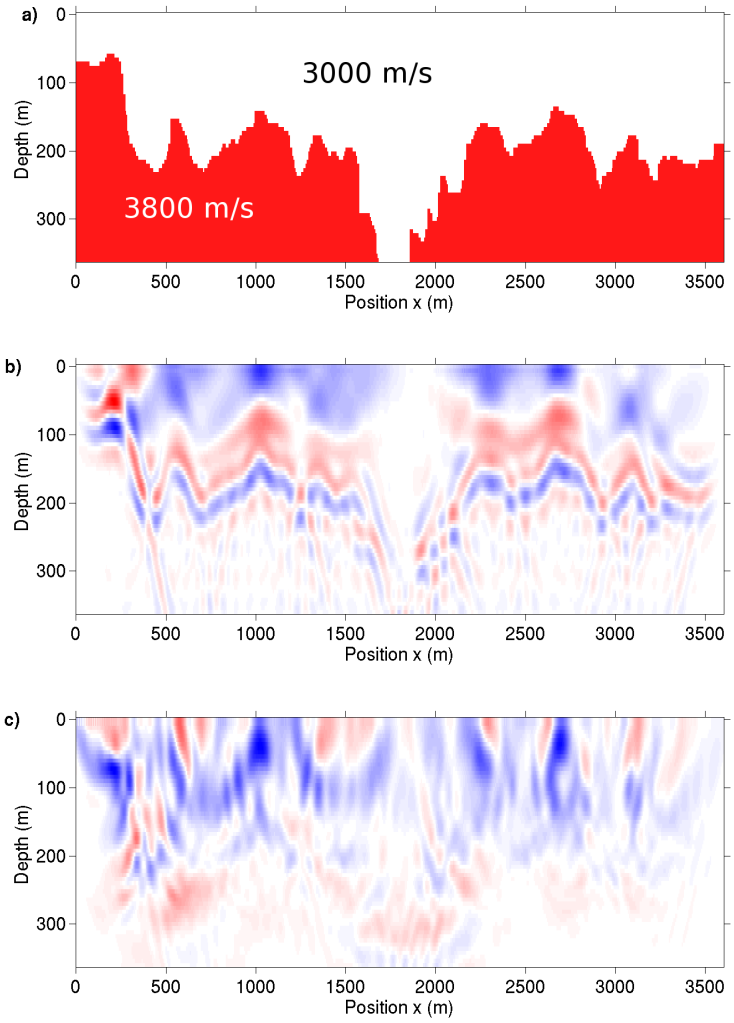


Figure 4.11. – Exact velocity model (a), migrated section in $v_0 = 2500\,\mathrm{m/s}$ and for $h = 0\,\mathrm{m}$ (b) and associated gradient (c).

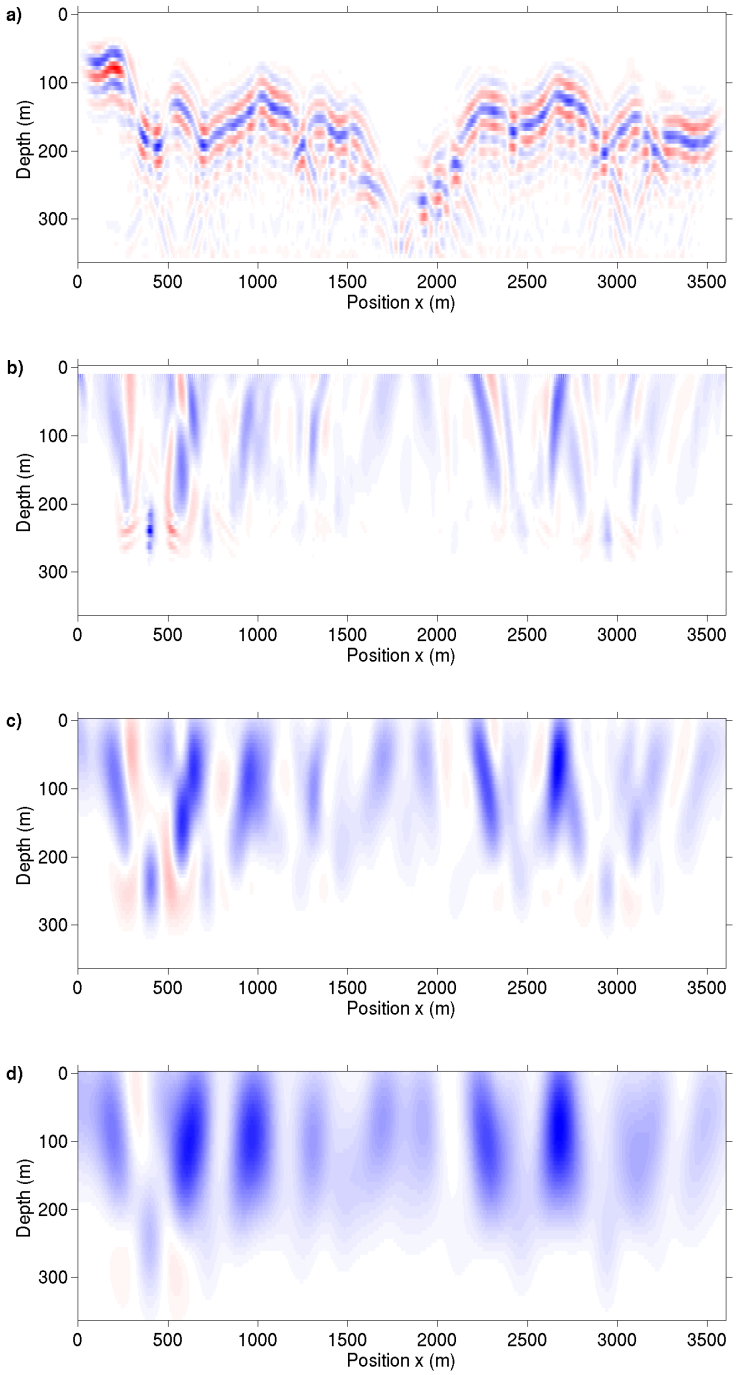


Figure 4.12. – Inversion results for $h = 0\,\mathrm{m}$ (a) and associated gradient for $\beta = -3/2$, before (b) and after smoothing with (c) $\sigma = \lambda_{\mathrm{data}}/2$ and (d) $\sigma = \lambda_{\mathrm{data}}$.

gradient is mainly positive at very shallow depths and then becomes negative as expected (figure 4.11c). With the inversion approach, the reflectivity is sharper (figure 4.12a) and the gradient has a mean negative (blue) value (figure 4.12b). After smoothing with $\sigma = \lambda_{data}/2$, the gradient is mainly negative, except in the left part where some oscillations are visible (figure 4.12c). A stronger smoothing with $\sigma = \lambda_{data}$ would remove the positive values (figure 4.12d).

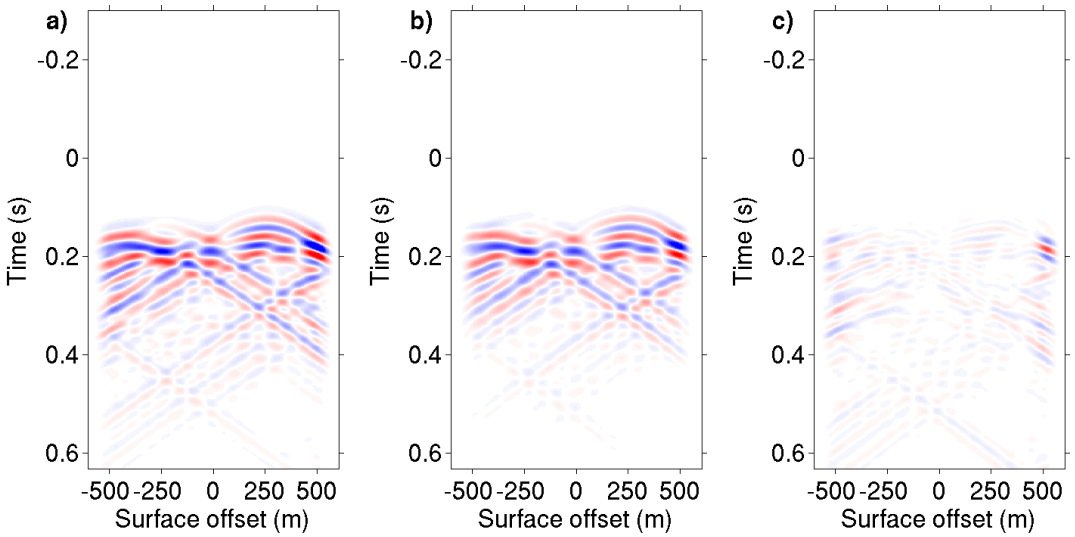


Figure 4.13. – Observed (a) and computed shot gather at $s = 1750\,\text{m}$, after inversion and modelling in the incorrect model $v_0 = 2500\,\text{m/s}$ (b) and differences at the same scale (c).

It is interesting to compare the observed data and the data after inversion and modelling (figure 4.13). For the derivation of the gradient, we used $\beta = -3/2$ to remove the imprint of the interfaces in the gradient. Here we use $\beta = 0$ to check the quality of inversion. The first events correspond to reflections, whereas later events are associated to diffractions. Events are correctly retrieved. The largest residuals are visible at large offsets. Traces extracted for offsets $-300\,\text{m}$, $0\,\text{m}$ and $+300\,\text{m}$ show that both phases and amplitudes are correctly retrieved, meaning that inversion provides indeed an inverse, even in an incorrect model, here a homogeneous model at $v_0 = 2500\,\text{m/s}$ (figure 4.14). Note that the way to generate the data (without Born approximation) is not the same as the one used to define the modelling (with Born approximation). We would expect small differences at least in the case of small velocity contrasts: it seems to be robust as the velocity contrast is here $+800\,\text{m/s}$.

Low velocity anomaly

Finally, we investigate the case of a velocity model containing a low velocity zone of $-500\,\text{m/s}$ (figure 4.15a), leading to triplicated wavefields. A similar example is discussed in Shen and Symes (2015) as this is a model for which the horizontal contraction (Fei and Williamson, 2010) does not provide a correct answer (Shen and Symes, 2015).

After migration in $v_0 = 3000\,\text{m/s}$, the reflectivity gathers show deflected energy, supposed to be horizontal in the exact model (figure 4.15b). The associated gradient is expected to display

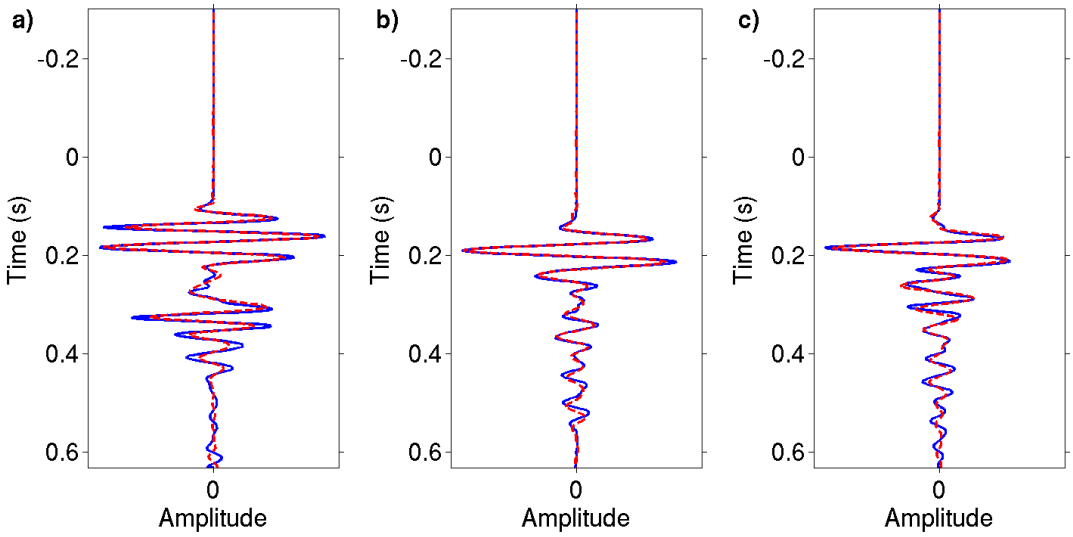


Figure 4.14. – Traces extracted from figure 4.13, for surface offset −300 (a), 0 (b) and +300 m (c). The blue colour corresponds to the observed shot and the red to the shot after inversion and modelling in the incorrect model $v_0 = 2500\,\mathrm{m/s}$.

a red (positive) anomaly. This is not the case (figure 4.15c). The zones in the dotted ellipses are positive (red) due to migration smiles in the CIGs. The structure of the central part (solid ellipse), with a negative (light blue) value below and between positive values, is similar to the one presented in Shen and Symes (2015), figure 2c in their paper. As explained in Shen and Symes (2015), the horizontal contraction in the case of triplicated wavefields does not provide a proper solution. With the inversion approach, the reflectivity section is similar (figure 4.16a) but the gradient displays vertical stripes (figure 4.16b). This is only the first step in model building. After 10 iterations with a non-linear conjugate gradient, the inverted background model contains a main velocity anomaly (figure 4.16c). The colorbar scale is the same as the one figure 4.15a. As the model only contains three reflectors, the shape of the anomaly is not perfectly constrained, but is localised around the correct position (dotted circle). Finally, the reflectivity in the inverted model displays relatively horizontal shapes. The focusing is largely improved after updating the background velocity model as indicated by the CIGs (figure 4.17).

We conclude from these different investigations that the inversion formulas proposed for the 1D and 2D cases (equations 4.6 and 4.8c) are indeed inverse and not adjoint, as the data after inversion followed by modelling are nicely retrieved, even in an incorrect velocity model. The combination of inversion and velocity analysis provides a gradient that does not suffer from artefacts as described by (Fei and Williamson, 2010; Vyas and Tang, 2010). We propose to incorporate m_0^β in the definition of the second objective function to assess the quality of focusing in CIGs. With $\beta = -3/2$, the imprint of reflector is not visible in the gradient. Finally, the gradient may be smoothed with a characteristic length proportional to the wavelength of the data. In the last example, triplicated wavefields are visible. Even if the derivation of the

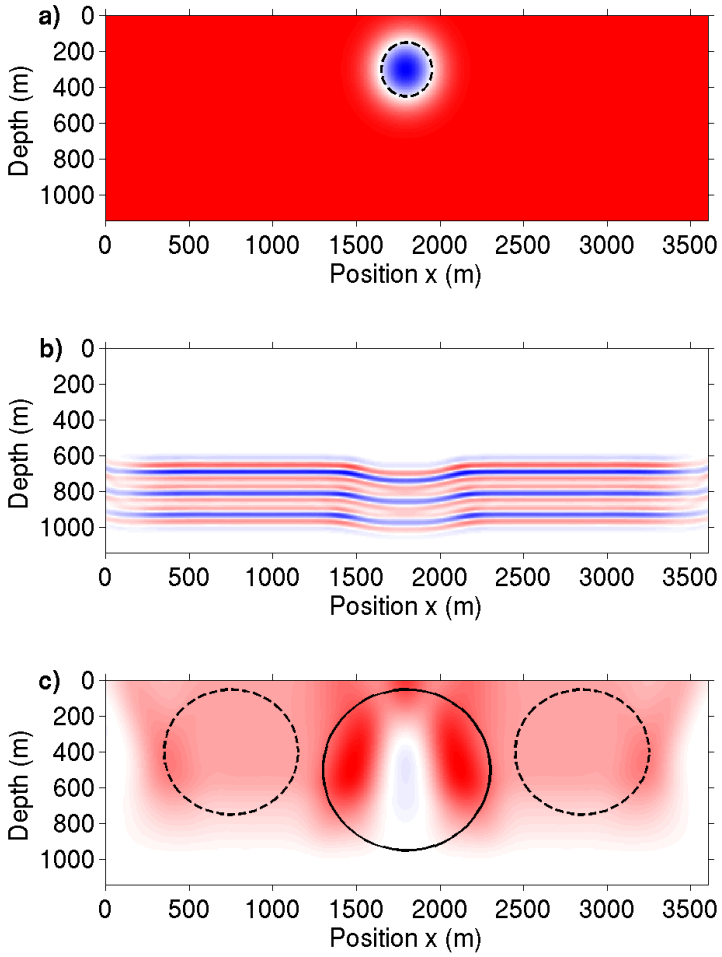


Figure 4.15. – Exact velocity model (a), migrated section in $v_0 = 3000\,\mathrm{m/s}$ and for $h = 0\,\mathrm{m}$ (b) and associated gradient (c). The colorbar ranges from 2500 (blue) to 3000 m/s (red).

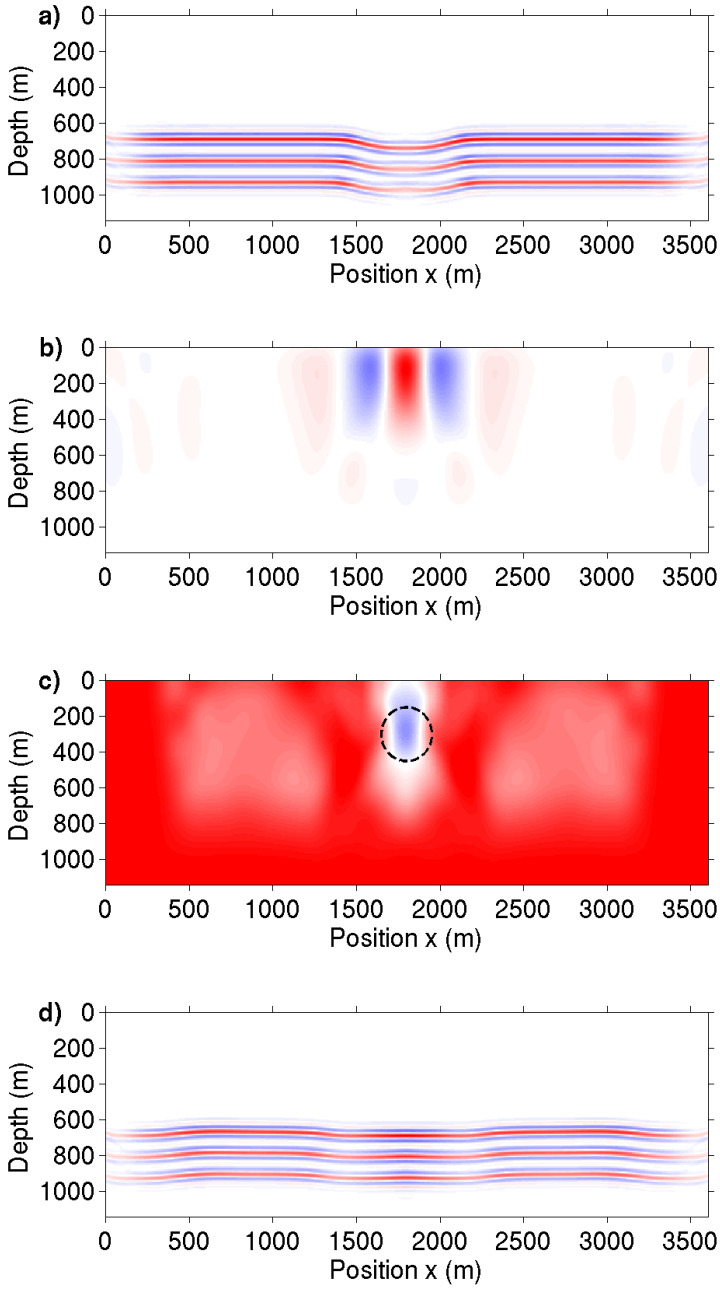


Figure 4.16. – Inversion section in $v_0 = 3000\,\mathrm{m/s}$ and for $h = 0\,\mathrm{m}$ (a), first gradient (b), final inverted model after 10 iterations (c), and associated inverted reflectivity (d). The colorbar ranges from 2500 (blue) to 3000 m/s (red).

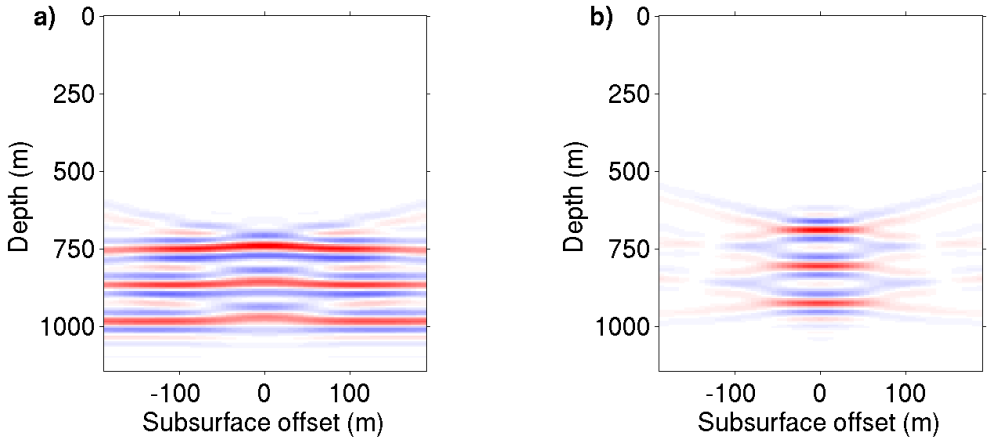


Figure 4.17. – CIGs for $x = 1750$ m computed in (a) the initial homogeneous model and (b) the final model (figure 4.16c).

inverse formula depends on absence of multiple ray paths, the final result may incorporate them as indicated by Stolk *et al.* (2009), ten Kroode (2012) and Hou and Symes (2015).

4.2.7. Discussion

The inversion formula 4.6 is not the same as the one provided by Hou and Symes (2015). In the derivation proposed here, we have supposed that the main contribution in equation (4.5) is obtained when $\mathbf{y}'$ is close to $\mathbf{y}$ (Beylkin, 1985): we thus linearised the phase of the $\mathscr{F}$ operator, whereas Hou and Symes (2015) directly applied the stationary phase approximation. More importantly, the main advantage of their formulation is that they express the inverse operator as a combination of a model preconditioner $\mathrm{W}_{\mathrm{model}}$, the classical adjoint operator and a data-domain preconditioner $\mathrm{W}_{\mathrm{data}}$. Such an expression allows to easily adapt the iterative conjugate gradient by simply modifying the norms in the data and model domains (Hou and Symes, 2016a). In our case, the inversion formula (equation 4.6) cannot be recast as a combination of different preconditioners, but the vertical source derivatives are applied to the Green's function and not to the observed data. Note that the integration over sources is performed along the horizontal component of the source position and not the vertical one.

In the case of laterally invariant model, the inversion expressions simplify in the two approaches. In Hou and Symes (2015), the $\mathrm{W}_{\mathrm{model}}$ operator is obtained through two Lapacians in the (x,z) and (h,z) domains $\mathscr{L} = \sqrt{-\nabla^2_{x,z}}\sqrt{-\nabla^2_{h,z}}$ (equation 30 in Hou and Symes, 2015). In our approach, only a vertical derivative $\mathscr{D}_z$ is needed. The number of time integration also differs. Hou and Symes (2015) considered a linearisation of the velocity model, whereas we used a squared slowness perturbation. The weighting factor difference $2/v_0^3$ appears twice in the modelling and inverse operators. It means that the differences between Hou and Symes (2015) and the approach proposed here is a factor $\mathscr{R}$

$$\mathscr{R} = \frac{2}{v_0}\frac{\omega k_z}{\sqrt{k_z^2+k_x^2}\sqrt{k_z^2+k_h^2}} \simeq \frac{2}{v_0}\frac{\omega}{k_z}, \tag{4.16}$$

where (k_z, k_x, k_h) are the wavenumbers associated to (z, x, h). The second equality is only valid for small k_x and k_h values. Operator $\mathscr{R}$ is adimensional and it is similar to a quantity arising in two-way reflection travel time. For the optimal β, a multiplication by $m_0^{-3/2}$ in front of the reflectivity function in equation (4.6) modifies the $32m_0$ factor to $32m_0 \cdot m_0^{-3/2} = 32v_0$. In 1D, it would lead to $16v_0$.

In terms of implementation, Hou and Symes (2015) need to apply the $\mathrm{W}_{\mathrm{data}}$ operator to the data residuals. The derivative with respect to the vertical component of the source and receiver can be obtained with the free surface (ten Kroode, 2012; Hou and Symes, 2015). The $\mathscr{L}$ operator can be obtained with two 2D Fourier transform, in (x, z) and (h, z) domains. In our approach, the implementation is very similar to the classical adjoint implementation. The vertical derivatives at the source and at the receiver positions are applied to the Green's functions and can be obtained by introducing a dipole instead of point sources. For the computation of the gradient with respect to the model perturbation, only two extra variables ξ and λ are needed, compared to classical MVA. ξ and λ are function of ξ_0 and λ_0 (appendix-4.2.11).

For the numerical illustrations, we assume that the background model was laterally invariant. In that case, the $\mathscr{D}_p$ operator (equation 4.7) simplifies to a vertical derivative $\mathscr{D}_z$. The study of the correct operator is beyond the scope of this article. It would imply the derivative with respect to h_z, taken at $h_z = 0$. More research is needed to know how to compute such quantity, as CIGs depend on h_x and are computed for $h_z = 0$ only.

The additional weights introduced in the inversion formula are mainly cosines of take-off angles at the source and receiver positions, and cosine of opening and dip angles at the image point. The main effects are visible for shallow depths and large offset contributions. As illustrated in figures 4.5 and 4.7, the migration smiles are largely attenuated. Weibull and Arntsen (2013) proposed to introduce a vertical derivative $\mathscr{D}_z$ to improve the robustness of MVA. Here, we give a justification for the introduction of this weight that is explicitly defined in the inversion formula. It is interesting to note that the same derivative in z also appears in the inverse one-way Born formula (Joncour *et al.*, 2011). Here, we rely on two-way wave-equation operators. The vertical derivative is a high-pass filter, attenuating the most dipping events in CIGs, especially vertical events associated to lower wavenumbers. An important aspect is the introduction of the determinant (equation 4.23) due to a change of variables between the surface acquisition variables and the subsurface image variables. An alternative is to compensate for uneven illumination through ad-hoc weights usually derived in simple models (Wu and Chen, 2006; Yang *et al.*, 2013; Tang *et al.*, 2013). The inversion formula provides an automatic and correct strategy.

Instead of considering inversion, one could solve the least-squares functional in an iterative manner (equation 4.1). The computational cost is proportional to the number of iterations. However, it may have a number of advantages. When the data do not linearly depend on the reflectivity, as in the case of surface multiples, inverse formula does not necessarily exist. For the derivation of the gradient with respect to the background model, one would need to take into account the iterations to derive the optimal reflectivity ξ. An alternative is to suppose that the convergence has been reached, leading to a simple linear problem to be solved (Cocher *et al.*, 2017a). More investigations are needed to further compare MVA with iterative or direct inversion of the reflectivity.

Finally, the gradients only provide the first iteration needed to update the background velocity model. For a complete minimisation, the estimation of the Hessian of the second objective function is essential (equation 4.10b). A first possibility is to estimate the Hessian in a homogeneous model. An interesting alternative is proposed by Shen and Symes (2015) for the estimation of the diagonal term. The derivation was in the case of the horizontal contraction. More research, beyond the scope of this article, is needed to determine an equivalent approach when inversion is coupled to velocity analysis.

4.2.8. Conclusions

We have replaced migration with the inverse of the modelling operator in image domain velocity analysis. This is an automatic way to compensate for uneven illumination and to remove migration smiles especially visible for shallow depth and large offsets. As a consequence, the gradient associated to the quality of focusing has a smoother behaviour in simple models and does not suffer from oscillations even in models with truncated reflectivities. Compared to an alternative solution based on horizontal contraction, the velocity update is the gradient of an objective function. To remove the imprint of the reflectivity, we have proposed to use the inverse of the modelling operator multiplied by the velocity at a specific power determined in 1D and applicable in 2D. The next step consists of applying non-linear iterations to fully determine the velocity model.

4.2.9. Acknowledgements

The authors are grateful to the editor, to the associated editor T. Alkhalifah, to the reviewers W. Symes and B. Sun, and to an anonymous reviewer, as they all have provided very detailed comments to improve the initial manuscript. The authors would like to thank W. Symes for inspiring this work. They are grateful to him and to R.-É. Plessix for fruitful discussions. The research described in this article was carried out as a part of the Paris Exploration Geophysics Group project (GPX) funded by the French National Research Agency (ANR), CGG, Total and Schlumberger.

4.2.10. Inversion (appendix-A)

We develop the main steps to derive the inversion formula (equation 4.6). Under the high frequency approximation, the Green's function in the frequency domain is given by

$$G_0(\mathbf{s},\mathbf{x},\omega) = K(\omega)A(\mathbf{s},\mathbf{x})e^{i\omega T(\mathbf{s},\mathbf{x})}, \tag{4.17}$$

where T and A are the travel times and geometrical amplitudes solutions of the Eikonal and transport equations. In 1D, $K(\omega) = 1/i\omega$ and in 2D, $K(\omega) = 1/\sqrt{i\omega}$, introducing an integration or half-integration in the time domain. Let define $\mathbf{k_x} = i\omega\nabla T(\mathbf{s},\mathbf{x}-\mathbf{h}) + i\omega\nabla T(\mathbf{r},\mathbf{x}+\mathbf{h})$. The $\mathscr{F}$ operator obtained as a combination of the forward and inverse operators in equation (4.5)

reads

$$\mathscr{F}(\mathbf{y},\mathbf{y}') = k(\mathbf{x},\mathbf{h}) \int \mathrm{d}\mathbf{k}\,(\mathrm{i}\omega)^{2+\nu+\beta} \left|\frac{\partial(\mathbf{s},\mathbf{r},\omega)}{\partial(\mathbf{k})}\right| \mathrm{A}^2(\mathbf{s},\mathbf{x}-\mathbf{h})\mathrm{A}^2(\mathbf{r},\mathbf{x}+\mathbf{h}) \\ \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{y}'-\mathbf{y})} m_{0s}^{1/2} m_{0r}^{1/2} \cos\beta_s \cos\beta_r, \tag{4.18}$$

where $\mathbf{y} = (\mathbf{x},\mathbf{h})$ is the vector in the extended space, and $\mathbf{k} = (\mathbf{k_x},\mathbf{k_h})$ the associated wavenumber vector. m_{0s} and m_{0r} are model values at the source and receiver positions, whereas β_s and β_r are take-offset angles for the source and receiver positions. To obtain equation (4.18), we have first used the following approximations

$$\mathscr{D}_{s_z}\mathrm{G}_0(\mathbf{s},\mathbf{x},\omega) \simeq (\mathrm{i}\omega)\cos\beta_s m_{0s}^{1/2}\mathrm{G}_0(\mathbf{s},\mathbf{x},\omega) \tag{4.19}$$

$$\mathscr{D}_{r_z}\mathrm{G}_0(\mathbf{r},\mathbf{x},\omega) \simeq (\mathrm{i}\omega)\cos\beta_r m_{0r}^{1/2}\mathrm{G}_0(\mathbf{r},\mathbf{x},\omega) \tag{4.20}$$

$$\mathscr{D}_z\mathrm{G}_0(\mathbf{s},\mathbf{x}-\mathbf{h},\omega)\mathrm{G}_0(\mathbf{r},\mathbf{x}+\mathbf{h},\omega) \simeq (\mathrm{i}\omega)(+m_{0-}^{1/2}\cos\theta_s + m_{0+}^{1/2}\cos\theta_r)\mathrm{G}_0(\mathbf{s},\mathbf{x}-\mathbf{h},\omega) \\ \mathrm{G}_0(\mathbf{r},\mathbf{x}+\mathbf{h},\omega) \tag{4.21}$$

$$\mathscr{D}_{hz}\mathrm{G}_0(\mathbf{s},\mathbf{x}-\mathbf{h},\omega)\mathrm{G}_0(\mathbf{r},\mathbf{x}+\mathbf{h},\omega) \simeq (\mathrm{i}\omega)(-m_{0-}^{1/2}\cos\theta_s + m_{0+}^{1/2}\cos\theta_r)\mathrm{G}_0(\mathbf{s},\mathbf{x}-\mathbf{h},\omega) \\ \mathrm{G}_0(\mathbf{r},\mathbf{x}+\mathbf{h},\omega), \tag{4.22}$$

where θ_s and θ_r are the angles at the image points $\mathbf{x}-\mathbf{h}$ and $\mathbf{x}+\mathbf{h}$ associated to the source and receiver positions. The model values m_{0-} and m_{0+} are also evaluated at $\mathbf{x}-\mathbf{h}$ and $\mathbf{x}+\mathbf{h}$. Then we apply a change of variable from $(\mathbf{s},\mathbf{r},\omega)$ to $\mathbf{k}$ and replace the Green's functions by their high frequency approximation. The squared amplitudes are due to the application of the forward and inverse operators.

The remaining quantities to be evaluated are the determinant and the amplitude terms. Here we use the fact that the kernel $\mathrm{e}^{\mathrm{i}\mathbf{k}(\mathbf{y}-\mathbf{y}')}$ is oscillating and that the amplitude terms have a smoother behaviour (ten Kroode, 2012): we thus evaluate the amplitudes only at $\mathbf{y}$. The 3×3 determinant in 2D can easily be obtained as $\mathbf{k_s}$ and $\mathbf{k_r}$ have explicit expressions, yielding

$$\left|\frac{\partial(\mathbf{k})}{\partial(\mathbf{s},\mathbf{r},\omega)}\right| = -(\mathrm{i}\omega)^2 m_{0+}^{1/2} m_{0-}^{1/2} \left|\frac{\partial\theta_s}{\partial s}\right| \left|\frac{\partial\theta_r}{\partial r}\right| \begin{vmatrix} \cos\theta_s & \cos\theta_r & m_{0-}^{1/2}\sin\theta_s + m_{0+}^{1/2}\sin\theta_r \\ -\sin\theta_s & -\sin\theta_r & m_{0-}^{1/2}\cos\theta_s + m_{0+}^{1/2}\cos\theta_r \\ -\cos\theta_s & \cos\theta_r & -m_{0-}^{1/2}\sin\theta_s + m_{0+}^{1/2}\sin\theta_r \end{vmatrix}, \\ = 2(\mathrm{i}\omega)^2 m_{0+}^{1/2} m_{0-}^{1/2} \left|\frac{\partial\theta_s}{\partial s}\right| \left|\frac{\partial\theta_r}{\partial r}\right| \left(m_{0+}^{1/2}\cos\theta_s + m_{0-}^{1/2}\cos\theta_r\right). \tag{4.23}$$

The determinant expression contains terms evaluated at $\mathbf{x}-\mathbf{h}$ and $\mathbf{x}+\mathbf{h}$, and partial derivatives of the angles θ_s and θ_r at the image positions $\mathbf{x}-\mathbf{h}$ and $\mathbf{x}+\mathbf{h}$, with respect to the source and receiver positions. Note that the determinant also has the term $m_{0+}^{1/2}\cos\theta_s + m_{0-}^{1/2}\cos\theta_r$. The application of $\mathscr{D}_z$ and $\mathscr{D}_{hz}$ on the Green's function provides $m_{0-}^{1/2}\cos\theta_s + m_{0+}^{1/2}\cos\theta_r$ and $m_{0-}^{1/2}\cos\theta_s - m_{0+}^{1/2}\cos\theta_r$ (equations 4.21 and 4.22): in other words, the model value $m_{0-}^{1/2}$ in front of $\cos\theta_s$ is estimated at $\mathbf{x}-\mathbf{h}$ and not at $\mathbf{x}+\mathbf{h}$ as it should be in the determinant expression. The same holds for the receiver side. The new term $\mathscr{D}_p$ is defined as a linear combination of $\mathscr{D}_z$ and $\mathscr{D}_{hz}$ (equation 4.7) to fully remove $m_{0+}^{1/2}\cos\theta_s + m_{0-}^{1/2}\cos\theta_r$ in equation (4.23). Finally, the

expressions for the amplitudes are given in Zhang *et al.* (2005), ten Kroode (2012) and Hou and Symes (2015).

$$A^2(\mathbf{s},\mathbf{x}-\mathbf{h}) = \frac{1}{8\pi^2}\frac{1}{m_{0s}^{1/2}\cos\beta_s}\left|\frac{\partial\theta_s}{\partial s}\right|, \tag{4.24a}$$

$$A^2(\mathbf{r},\mathbf{x}+\mathbf{h}) = \frac{1}{8\pi^2}\frac{1}{m_{0r}^{1/2}\cos\beta_r}\left|\frac{\partial\theta_r}{\partial r}\right|. \tag{4.24b}$$

Once more, it is remarkable that the two derivative terms $|\partial\theta_s/\partial s|$ and $|\partial\theta_r/\partial r|$ cancel out in the $\mathscr{F}$ expression. The $\cos\beta_s$ and $\cos\beta_r$ terms are also removed in equation (4.18): this was the reason why vertical derivatives at the source and receiver positions were introduced in equation (4.4).

4.2.11. Gradient derivation for the inversion case (appendix-B)

We explain how to compute the gradient of the objective function J^0_{inv} with respect to the model m_0 (equation 4.11b). The expression for β different from zero is simply given in equation (4.12b). With the Lagrangian formalism, the principle consists of extending the objective function and of introducing the state equations are constrains (Plessix, 2006). The strategy is the same for the classical DSO approach.

$$\begin{aligned}\widetilde{J} = \frac{1}{2}\|a\xi\|^2 &- \big\langle \lambda(\mathbf{x},\mathbf{h}) \,\big|\, \xi(\mathbf{x},\mathbf{h}) - \mathscr{D}_p\xi_0(\mathbf{x},\mathbf{h})\big\rangle \\ &- \left\langle \lambda_0(\mathbf{x},\mathbf{h}) \,\middle|\, \xi_0(\mathbf{x},\mathbf{h}) - \int d\mathbf{s}\, d\omega\, S_0^*(\mathbf{s},\mathbf{x}-\mathbf{h},\omega)R_0(\mathbf{s},\mathbf{x}+\mathbf{h},\omega)\right\rangle \\ &- \big\langle \mu_s(\mathbf{s},\mathbf{x},\omega) \,\big|\, \mathscr{L}_0 S_0 - S_s(\mathbf{s},\mathbf{x},\omega)\big\rangle \\ &- \big\langle \mu_r(\mathbf{s},\mathbf{x},\omega) \,\big|\, \mathscr{L}_0^* R_0 - R_s(\mathbf{s},\mathbf{x},\omega)\big\rangle,\end{aligned} \tag{4.25}$$

where J now depends on $(m,\lambda,\xi,\lambda_0,\xi_0,\mu_s,S_0,\mu_r,R_0)$. The first part in equation (4.25) is the original objective function. The other parts are respectively the definition of the reflectivity variables ξ and ξ_0, and of the forward S_0 and backward R_0 wavefields, solutions of the wave equation with source terms S_s and R_s. The associate adjoint variables are λ, λ_0, μ_s and μ_r. The total gradient reads

$$\begin{aligned}\frac{dJ}{dm_0} = \frac{\partial J}{\partial m_0} &+ \frac{\partial\lambda}{\partial m_0}\frac{\partial J}{\partial\lambda} + \frac{\partial\xi}{\partial m_0}\frac{\partial J}{\partial\xi} + \frac{\partial\lambda_0}{\partial m_0}\frac{\partial J}{\partial\lambda_0} + \frac{\partial\xi_0}{\partial m_0}\frac{\partial J}{\partial\xi_0} \\ &+ \frac{\partial\mu_s}{\partial m_0}\frac{\partial J}{\partial\mu_s} + \frac{\partial S_0}{\partial m_0}\frac{\partial J}{\partial S_0} + \frac{\partial\mu_r}{\partial m_0}\frac{\partial J}{\partial\mu_r} + \frac{\partial R_0}{\partial m_0}\frac{\partial J}{\partial R_0}.\end{aligned} \tag{4.26}$$

In order the avoid to compute the Fréchet derivatives (derivatives of λ, ξ, λ_0, ξ_0, μ_s, S_0, μ_r and R_0 with respect to m_0), we set as a choice and as a definition for the adjoint variables λ, λ_0, μ_s and μ_r, the derivative of J with respect to λ, ξ, λ_0, ξ_0, μ_s, S_0, μ_r and R_0 to zero, leading to

eight equations

$$\lambda = a^2 \xi, \tag{4.27a}$$
$$\xi(\mathbf{x},\mathbf{h}) = \mathscr{D}_p \xi_0(\mathbf{x},\mathbf{h}), \tag{4.27b}$$
$$\xi_0(\mathbf{x},\mathbf{h}) = \int S_0^*(\mathbf{s},\mathbf{x}-\mathbf{h},\omega) R_0(\mathbf{s},\mathbf{x}+\mathbf{h},\omega)\, d\mathbf{s}\, d\omega, \tag{4.27c}$$
$$\mathscr{L}_0^* R_0 = R_s(\mathbf{s},\mathbf{x},\omega), \tag{4.27d}$$
$$\mathscr{L}_0 S_0 = S_s(\mathbf{s},\mathbf{x},\omega). \tag{4.27e}$$
$$\lambda_0(\mathbf{x},\mathbf{h}) = \mathscr{D}_p^* \lambda = \mathscr{D}_p^* a^2 \mathscr{D}_p \xi_0, \tag{4.27f}$$
$$= -32^2 m^2 a^2 \mathscr{D}_z^2 \xi_0 \quad \text{for laterally invariant models,} \tag{4.27g}$$
$$= a^2 \xi_0 \quad \text{for the adjoint version,} \tag{4.27h}$$
$$\mathscr{L}_0^* \mu_s = \int \lambda_0(\mathbf{x}+\mathbf{h},\mathbf{h}) R_0(\mathbf{s},\mathbf{x}+2\mathbf{h},\omega)\, d\mathbf{h}, \tag{4.27i}$$
$$\mathscr{L}_0 \mu_r = \int \lambda_0(\mathbf{x}-\mathbf{h},\mathbf{h}) S_0(\mathbf{s},\mathbf{x}-2\mathbf{h},\omega)\, d\mathbf{h}. \tag{4.27j}$$

equations (4.27a) to (4.27e) are solved in the reverse order. ξ_0 is obtained by crosscorrelating the forward wavefield S_0 with the back-propagated residual wavefield R_0. Then, ξ is obtained by simply applying the differential $\mathscr{D}_p$ operator onto ξ_0. For the λ_0 variable, the adjoint of $\mathscr{D}_p$ should be applied to λ. In a laterally invariant model, $m_{0+} = m_{0-} = m_0$. As $a(h)$ does not depend on $\mathbf{x}$, λ_0 is obtained by applying the second derivative in z (equation 4.27g). As for S_s and R_s, μ_s and μ_r are solution of the same wave equation, but for different source terms for which the modified reflectivity λ_0 is activated by R_0 or S_0 (equations 4.27i and 4.27j). For the final expression of the gradient, both $\mathscr{L}_0$ and $\mathscr{D}_p$ depend on m_0, leading to three terms (equation 4.11b).

4.2.12. Optimal β parameter (appendix-C)

In the 1D case, we demonstrate that oscillatory terms in equation (4.14) cancel out. The strategy consists (1) of splitting the integrals over y from 0 to $z_{\max}$ into two parts, and (2) of integrating by part. According to equation (4.14), G_{inv}^{β} consists of three terms G_1, G_2 and G_3, namely

$$G_1(z) = \frac{\gamma}{v_0} \int_0^{z_{\max}} dy\, (y - z_e)^2 g'_{\text{inv}}(p[y]) g_{\text{inv}}(q[z,y]), \tag{4.28a}$$
$$G_2(z) = \frac{\gamma}{2}(\beta+2)(z-z_e)^2 g^2_{\text{inv}}(p[z]), \tag{4.28b}$$
$$G_3(z) = \gamma \int_0^{z_{\max}} dy\, (y - z_e) g_{\text{inv}}(p[y]) g_{\text{inv}}(q[z,y]), \tag{4.28c}$$

with γ, $p(y)$ and $q(z,y)$ given by

$$\gamma = \frac{4R_e^2}{v_0^{2\beta}}\left(\frac{v_e}{v_0}\right)^2, \tag{4.29a}$$

$$p = 2\left(\frac{z_e}{v_e} - \frac{y}{v_0}\right), \tag{4.29b}$$

$$q = 2\left(\frac{z_e}{v_e} - \frac{z+y+|z-y|}{2v_0}\right). \tag{4.29c}$$

Note that $q(z,y) = p(z)$ if $z \geq y$, and $p(y)$ otherwise. We consider G_1 and first split the integrals between 0 to z and between z to $z_{\max}$

$$\begin{aligned} G_1(z) =& \frac{\gamma}{v_0}\int_0^z (y-z_e)^2 g'_{\text{inv}}(p[y])g_{\text{inv}}(q[z,y])\,dy \\ &+ \frac{\gamma}{v_0}\int_z^{z_{\max}} (y-z_e)^2 g_{\text{inv}}(p[y])g'_{\text{inv}}(q[z,y])\,dy, \end{aligned} \tag{4.30a}$$

$$\begin{aligned} =& \frac{\gamma}{v_0} g_{\text{inv}}(p[z])\int_0^z (y-z_e)^2 g'_{\text{inv}}(p[y])\,dy \\ &+ \frac{\gamma}{v_0}\int_z^{z_{\max}} (y-z_e)^2 g_{\text{inv}}(p[y])g'_{\text{inv}}(p[y])\,dy. \end{aligned} \tag{4.30b}$$

We have simplified $q(z,y)$ depending if $z \geq y$ or not. The first part of equation (4.30b) is integrated by parts, yielding

$$\begin{aligned} &-\frac{\gamma}{2} g_{\text{inv}}(p[z])\Big[(y-z_e)^2 g_{\text{inv}}(p[y])\Big]_0^z + \gamma\, g_{\text{inv}}(p[z])\int_0^z (y-z_e) g_{\text{inv}}(p[y])\,dy \\ =& -\frac{\gamma}{2}(z-z_e)^2 g^2_{\text{inv}}(p[z]) + \gamma\int_0^z dy\,(y-z_e) g_{\text{inv}}(p[y]) g_{\text{inv}}(q[z,y])\,dy, \end{aligned} \tag{4.31}$$

In a similar way, we integrate the second part of equation (4.30b)

$$\begin{aligned} &-\frac{\gamma}{4}\Big[(y-z_e)^2 g^2_{\text{inv}}(p[y]))\Big]_z^{z_{\max}} + \frac{\gamma}{2}\int_z^{z_{\max}} (y-z_e) g^2_{\text{inv}}(p[y])\,dy \\ =& \frac{\gamma}{4}(z-z_e)^2 g^2_{\text{inv}}(p[z]) + \frac{\gamma}{2}\int_z^{z_{\max}} (y-z_e) g_{\text{inv}}(p[y]) g_{\text{inv}}(q[z,y])\,dy. \end{aligned} \tag{4.32}$$

The integration by parts introduced a multiplication by $-v_0/2$. We have supposed that

$$g_{\text{inv}}(p[0]) = g_{\text{inv}}(p[z_{\max}]) = 0.$$

This is reasonable if $z_{\max}$ if large enough. We combined the two parts and get

$$\begin{aligned} \mathrm{G}_1(z) = &-\frac{\gamma}{4}(z-z_e)^2 g_{\mathrm{inv}}^2(p[z]) \\ &+\gamma\int_0^z (y-z_e) g_{\mathrm{inv}}(p[y]) g_{\mathrm{inv}}(q[z,y])\,\mathrm{d}y \\ &+\frac{\gamma}{2}\int_z^{z_{\max}} (y-z_e) g_{\mathrm{inv}}(p[y]) g_{\mathrm{inv}}(q[z,y])\,\mathrm{d}y. \end{aligned} \tag{4.33}$$

The total gradient reads

$$\mathrm{G}_{\mathrm{inv}}^{\beta}(z) = \mathrm{G}_1(z) + \mathrm{G}_2(z) + \mathrm{G}_3(z), \tag{4.34}$$

$$\begin{aligned} &= \frac{\gamma}{2}\left(\beta+\frac{3}{2}\right)\mathrm{G}_2(z) \\ &+ 2\gamma\int_0^z (y-z_e) g_{\mathrm{inv}}(p[y]) g_{\mathrm{inv}}(q[z,y])\,\mathrm{d}y \\ &+ \frac{3\gamma}{2}\int_z^{z_{\max}} (y-z_e) g_{\mathrm{inv}}(p[y]) g_{\mathrm{inv}}(q[z,y])\,\mathrm{d}y. \end{aligned} \tag{4.35}$$

For $\beta = -3/2$, the first oscillatory term is zero. By analogy with the horizontal contraction approach, the remaining terms have a smooth behaviour: for an objective function $\mathrm{J}_1 = 1/2\|h\xi(\mathbf{x},h)\|^2$, the adjoint source wavefield is $\partial \mathrm{J}_1/\xi = h^2\xi$. Modified versions have been proposed such as $\mathrm{sign}(h)h^2\partial\xi/\partial h$ and $h\partial\xi/\partial h$ (Fei and Williamson, 2010; Shen and Symes, 2015). They are not associated any more to a gradient of an objective function, but usually lead to smooth gradients, except for example in the case of a low anomaly velocity (Shen and Symes, 2015). The equivalent in 1D would be $(z-z_e)\partial\xi/\partial z$. This is exactly the term we find in $\mathrm{G}_{\mathrm{inv}}^{\beta}(z)$ when $\beta = -3/2$ (equation 4.35). As a conclusion, we do not prove that $\mathrm{G}_{\mathrm{inv}}^{\beta}(z)$ is smooth, but has a similar behaviour as the terms in the horizontal contraction (Shen and Symes, 2015), while being a gradient of a specific objective function.

This concludes the manuscript submitted to Geophysics.

4.3. Introduction of migration weights in the iterative migration process

In section 4.2, we have introduced an approximate inverse $\mathrm{F}^{\dagger}$ of the extended Born-modelling operator F. Applying this pseudo-inverse to observed data produces CIGs free of migration artefacts, leading to homogeneous gradients of the associated MVA objective function. In this section we explain how $\mathrm{F}^{\dagger}$ can be introduced in the iterative migration process. This modification is expected to accelerate the convergence speed of iterative migration, but not necessarily to stabilise the MVA gradient of the nested optimisation process.

The weighted iterative scheme defined here is different from the one proposed by Hou and Symes (2016a). Their strategy is based on the approximate inverse formula of Hou and Symes

(2015), which can be expressed as a modification of the usual adjoint with two weighting operators:

$$\mathrm{F}^{\dagger} = \mathrm{W}_{\mathrm{mod}} \mathrm{F}^{\mathrm{T}} \mathrm{W}_{\mathrm{data}}. \tag{4.36a}$$

Then Hou and Symes (2016a) propose a modified iterative scheme using the standard linear conjugate gradient algorithm with new definitions for the norms in the extended model and data spaces,

$$\langle \xi \,|\, \chi \rangle_{\mathrm{W}_{\mathrm{mod}}} = \langle \xi \,|\, \mathrm{W}_{\mathrm{mod}} \chi \rangle_{\mathscr{E}}, \tag{4.36b}$$

$$\langle u \,|\, v \rangle_{\mathrm{W}_{\mathrm{data}}} = \langle u \,|\, \mathrm{W}_{\mathrm{data}} v \rangle_{\mathscr{D}_{\mathrm{obs}}}, \tag{4.36c}$$

with the requirement that $\mathrm{W}_{\mathrm{mod}}$ and $\mathrm{W}_{\mathrm{data}}$ are positive definite operators.

The inversion formula derived in section 4.2 cannot be decomposed in this way, so another strategy has to be defined. Here we propose to use the approximate inverse $\mathrm{F}^{\dagger}$ as a right preconditioner. Note that the use of an amplitude-preserving migration operator as a preconditioner has already been studied in other contexts. Sevink and Herman (1996) use it for iterative migration, but do not consider an extended model space. Métivier *et al.* (2015) use the Beylkin migration operator as a left preconditioner in FWI and show that it allows to accelerate the recovery of the short-scale structure of the velocity model.

In the following, we detail a preconditioned iterative scheme based on the inverse defined in section 4.2, first in the general non-linear case, then in the linear-case, corresponding to primaries modelled with the first-order Born approximation.

4.3.1. Preconditioned iterative migration: non-linear case

In the general non-linear case, the migration objective function $\mathrm{J}_0 : \mathscr{E} \mapsto \mathbb{R}$ is defined as (the implicit dependence of J_0 and P on c_0 is omitted in this section and in section 4.3.2 for the sake of clarity),

$$\mathrm{J}_0[\xi] = \frac{1}{2} \left\| \mathrm{MP}[\xi] - \mathrm{P}^{\mathrm{obs}} \right\|^2_{\mathscr{D}_{\mathrm{obs}}} + \varphi[\xi], \tag{4.37}$$

where operator $\mathrm{M} : \mathscr{D} \mapsto \mathscr{D}_{\mathrm{obs}}$ selects the value of calculated data at receiver positions. Preconditioning may be seen as a change of variables. Let us define a new variable $\widehat{\xi} \in \mathscr{D}_{\mathrm{obs}}$ such as

$$\xi = \mathrm{F}^{\dagger} \widehat{\xi}, \tag{4.38}$$

where $\mathrm{F}^{\dagger}$ is the pseudo-inverse of the extended Born modelling operator F, as defined in section 4.2. Replacing ξ by $\mathrm{F}^{\dagger}\widehat{\xi}$ in equation (4.37) leads to a new objective function $\widehat{\mathrm{J}_0} : \mathscr{D}_{\mathrm{obs}} \mapsto \mathbb{R}$,

$$\widehat{\mathrm{J}_0}\left[\widehat{\xi}\right] = \frac{1}{2} \left\| \mathrm{MP}\left[\mathrm{F}^{\dagger}\widehat{\xi}\right] - \mathrm{P}^{\mathrm{obs}} \right\|^2_{\mathscr{D}_{\mathrm{obs}}} + \varphi\left[\mathrm{F}^{\dagger}\widehat{\xi}\right], \tag{4.39}$$

with

$$\begin{cases} \widehat{\mathrm{J}_0}\left[\widehat{\xi}\right] = \mathrm{J}_0[\xi], & (4.40a) \\ \dfrac{\partial \widehat{\mathrm{J}_0}}{\partial \widehat{\xi}}\left[\widehat{\xi}\right] = (\mathrm{F}^{\dagger})^{\mathrm{T}} \dfrac{\partial \mathrm{J}_0}{\partial \xi}[\xi] & (4.40b) \end{cases}$$

The preconditioned iterative migration scheme now consists of finding the value of $\widehat{\xi}$ in the data space minimising $\widehat{J_0}$. The corresponding value in the image space is a minimiser of J_0 and can be computed as $\xi = F^\dagger \widehat{\xi}$. The inclusion of the preconditioner in the minimisation algorithm is easy. The algorithm is initiated with $\xi^{(1)} = 0$ in $\mathscr{E}$, so that the initial value in $\mathscr{D}_{\text{obs}}$ is $\widehat{\xi}^{(1)} = 0$. Then at each iteration, two steps are added to the classic algorithm, so that the value of $\widehat{\xi}^{(n+1)}$ is updated from $\widehat{\xi}^{(n)}$ in four steps:

(i) given $\widehat{\xi}^{(n)} \in \mathscr{D}_{\text{obs}}$, compute the corresponding value $\xi^{(n)} \in \mathscr{E}$ according to $\xi^{(n)} = F^\dagger \widehat{\xi}^{(n)}$;

(ii) compute the value and gradient of the original objective function: $J_0[\xi^{(n)}]$ and $g^{(n)} = \partial J_0 / \partial \xi[\xi^{(n)}] \in \mathscr{E}$;

(iii) using equations (4.40b) and (4.40b), go back to the $\mathscr{D}_{\text{obs}}$-space to determine the value of the new objective function $\widehat{J_0}$, and its gradient $\widehat{g}^{(n)} \in \mathscr{D}_{\text{obs}}$ with respect to $\widehat{\xi}$;

(iv) determine a descent direction $\widehat{d}^{(n)} \in \mathscr{D}_{\text{obs}}$ and a step length $\alpha^{(n)} \in \mathbb{R}$ to update $\widehat{\xi}$,

$$\widehat{\xi}^{(n+1)} = \widehat{\xi}^{(n)} + \alpha^{(n)} \widehat{d}^{(n)}, \tag{4.41}$$

and go back to step (i).

This optimisation algorithm is the same as the one described in chapter 2, except that it is performed in the observed data space instead of the extended model space. Only two additional calculations are needed: one to go from $\mathscr{D}_{\text{obs}}$ to $\mathscr{E}$ (step (i)) where the usual gradient is computed, and one to go back to $\mathscr{D}_{\text{obs}}$ (step (iii)). This operation requires the ability to compute the adjoint $(F^\dagger)^T$ of the approximate inverse $F^\dagger$. The calculation of this operator is detailed in appendix C. It can be interpreted as a weighted modelling favouring small reflection angles. Note that if the algorithm was initiated with a non-zero value $\xi^{(1)}$, we would need to compute the inverse $(F^\dagger)^{-1}$ of $F^\dagger$ to compute the corresponding value in the $\mathscr{D}_{\text{obs}}$-space. Starting with $\xi^{(1)} = 0$, this operator does not need to be computed.

The main computational cost remains the computation of the usual gradient in $\mathscr{E}$ (resolution of seven wave-equations, as described in table 2.2). With preconditioning, four additional wave-equations have to be solved (table 4.2), two for the application of $F^\dagger$ to $\widehat{\xi}^{(n)}$ and two for the application of $(F^\dagger)^T$ to $\partial J_0 / \partial \xi[\xi^{(n)}]$ (appendix C).

$F^\dagger$	$\mathscr{D}_{\text{obs}} \mapsto \mathscr{E}$	2
$(F^\dagger)^T$	$\mathscr{E} \mapsto \mathscr{D}_{\text{obs}}$	2
$(F^\dagger)^T F^\dagger$	$\mathscr{E} \mapsto \mathscr{E}$	3

Table 4.2. – Number of wave-equations to be solved to compute the inverse map $F^\dagger$, its adjoint $(F^\dagger)^T$ and the composition of the two operators. The number indicated here should be multiplied by the number of source positions considered in the acquisition.

In the non-linear case of multiples, the efficiency of the preconditioned strategy will be demonstrated on a numerical example (section 4.4.2). We now consider the special linear case of primaries only, where we explain theoretically why we expect the choice of $F^\dagger$ as preconditioner to accelerate the convergence.

4.3.2. Preconditioned iterative migration: linear case

We now consider the linear case, where the modelling is described by the linear extended Born modelling operator $\mathrm{F} : \mathscr{E} \mapsto \mathscr{D}_{\text{obs}}$ and the regularisation is the ℓ_2-norm. In this specific case, the implementation simplifies and the benefits of the preconditioner are easier to understand.

The classic (equation 4.37) and preconditioned (equation 4.39) versions of the objective function read

$$\begin{cases} \mathrm{J}_0[\xi] = \dfrac{1}{2}\left\|\mathrm{F}\xi - \mathrm{P}^{\text{obs}}\right\|^2_{\mathscr{D}_{\text{obs}}} + a_\varphi \dfrac{1}{2}\left\|\xi\right\|^2_{\mathscr{E}}, & \text{(4.42a)} \\ \widehat{\mathrm{J}_0}\left[\widehat{\xi}\right] = \dfrac{1}{2}\left\|\mathrm{F}\mathrm{F}^\dagger\widehat{\xi} - \mathrm{P}^{\text{obs}}\right\|^2_{\mathscr{D}_{\text{obs}}} + a_\varphi \dfrac{1}{2}\left\|\mathrm{F}^\dagger\widehat{\xi}\right\|^2_{\mathscr{E}}. & \text{(4.42b)} \end{cases}$$

Deriving these expressions with respect to ξ and $\widehat{\xi}$ respectively lead to two linear systems

$$\begin{cases} \left(\mathrm{F}^{\mathrm{T}}\mathrm{F} + a_\varphi \mathrm{I}\right)\xi = \mathrm{F}^{\mathrm{T}}\mathrm{P}^{\text{obs}}, & \text{(4.43a)} \\ (\mathrm{F}^\dagger)^{\mathrm{T}}\left(\mathrm{F}^{\mathrm{T}}\mathrm{F} + a_\varphi \mathrm{I}\right)\mathrm{F}^\dagger\widehat{\xi} = (\mathrm{F}^\dagger)^{\mathrm{T}}\mathrm{F}^{\mathrm{T}}\mathrm{P}^{\text{obs}}. & \text{(4.43b)} \end{cases}$$

The linear system (equation 4.43b) can be solved with a classic conjugate-gradient (CG) algorithm (Nocedal and Wright, 2006, p. 112). Note that this system is defined in the $\mathscr{D}_{\text{obs}}$-space, contrary to the classic linear system (equation 4.43a).

As $\mathrm{F}^\dagger$ is a pseudo inverse of F, the operator $(\mathrm{F}^\dagger)^{\mathrm{T}}\mathrm{F}^{\mathrm{T}}\mathrm{F}\mathrm{F}^\dagger = (\mathrm{F}\mathrm{F}^\dagger)^{\mathrm{T}}(\mathrm{F}\mathrm{F}^\dagger)$ should be close to the identity operator. Then, provided that the regularisation weight a_φ is relatively small, the resolution of this new linear system should be much faster and should require less iterations.

In practice, the resolution of the preconditioned system (4.43b) with the classic CG-algorithm is equivalent to the resolution of the usual system (4.43a) with the preconditioned CG-algorithm (Nocedal and Wright, 2006, p. 119) and the preconditioner $\mathrm{F}^\dagger(\mathrm{F}^\dagger)^{\mathrm{T}} : \mathscr{E} \mapsto \mathscr{E}$. This has the advantage of fewer modifications in the implementation, in particular the resolution is still performed in the $\mathscr{E}$-space. Compared to the usual CG-algorithm, the additional computational cost is the product at each iteration of $\mathrm{F}^\dagger(\mathrm{F}^\dagger)^{\mathrm{T}}$ with a vector of $\mathscr{E}$, requiring three additional wave-equations to be solved (table 4.2).

4.3.3. Preconditioned resolution of the linear adjoint problem

We now consider the computation of the gradient of the MVA objective function studied in chapter 3. Using method A, introducing the preconditioner would change the equations to be solved and would add new contributions to the total gradient, as $\mathrm{F}^\dagger$ and $(\mathrm{F}^\dagger)^{\mathrm{T}}$ depend on c_0. In the following, we consider only method B where the final reflectivity is defined as the exact solution of the linear system $\frac{\partial \mathrm{J}_0}{\partial \xi}(\xi^\infty) = 0$. Then introducing the preconditioner does not change the computation of the gradient, but only allows to find an approximate solution of this linear system in a reduced number of iterations.

Moreover, the preconditioner can also be used to accelerate the resolution of the adjoint system, which is a linear problem

$$\left[\frac{\partial^2 \mathrm{J}_0}{\partial \xi^2}\left[c_0, \xi^{(\mathrm{N}+1)}\right]\right]\eta = \mathrm{A}^{\mathrm{T}}\mathrm{A}\xi^{(\mathrm{N}+1)}, \tag{4.44}$$

where the adjoint variable is noted η from now on. In the linear case, this system is very similar to the direct problem (equation 4.43a),

$$\left(\mathrm{F}^{\mathrm{T}}\mathrm{F} + a_{\varphi}\mathrm{I}\right)\eta = \mathrm{A}^{\mathrm{T}}\mathrm{A}\xi^{(\mathrm{N}+1)}. \tag{4.45}$$

Then the preconditioned CG-algorithm with $\mathrm{F}^{\dagger}(\mathrm{F}^{\dagger})^{\mathrm{T}}$ as preconditioner can be used to solve the adjoint problem, both in the linear and non-linear cases. In the linear case at least, the preconditioner should in theory accelerate the resolution of the adjoint problem. This will be tested in section 4.4.3.

In summary, preconditioning does not change the computation of the gradient of the MVA objective function with method B (equation 3.12). The calculation only uses the last iterates $\xi^{(\mathrm{N}+1)}$ and $\eta^{(\mathrm{N}+1,\mathrm{M}+1)}$ of the direct and adjoint problems. Preconditioning only provides a more efficient way to find a solution to these problems in fewer iterations.

4.4. Synthetic examples

In this section, the behaviour of preconditioned iterative migration is illustrated on simple examples similar to the ones presented in chapters 2 and 3.

We first consider the case of primary reflections only. Note that in this case, iterative migration is not necessarily required as direct inversion (section 4.2) already provides a satisfactory reflectivity image. The primary-only case is studied here to test the efficiency of the preconditioner to accelerate the convergence of iterative migration. Then we consider a case with first-order surface-related multiples to determine if the preconditioner also improves the convergence rate of migration although it has not been designed for multiples. Finally, we investigate the efficiency of the preconditioner for the resolution of the adjoint problem with primaries only. We also compare the background velocity updates obtained with direct and iterative inversions.

4.4.1. Preconditioned iterative migration with primaries only

In this first example, we consider primary reflections only and a simple model with a single reflector located at 300 m depth in a homogeneous velocity model (3000 m/s).

First iterate of the preconditioned iterative scheme

We first compare the results obtained with the inverse formula derived in section 4.2 to the results of a single iteration of preconditioned iterative migration. The corresponding reflectivity images are defined as $\mathrm{F}^{\dagger}\mathrm{P}^{\mathrm{obs}}$ and $\mathrm{F}^{\dagger}(\mathrm{F}^{\dagger})^{\mathrm{T}}\mathrm{F}^{\mathrm{T}}\mathrm{P}^{\mathrm{obs}} = \mathrm{F}^{\dagger}(\mathrm{F}\mathrm{F}^{\dagger})^{\mathrm{T}}\mathrm{P}^{\mathrm{obs}}$, respectively. Note that the approximate inverse has been designed such that $(\mathrm{F}\mathrm{F}^{\dagger})$ is close to the identity operator in the observed data space. Then we expect its transpose $(\mathrm{F}\mathrm{F}^{\dagger})^{\mathrm{T}}$ to be close to identity as well, and thus the first iterate should be close to the result of direct inversion.

Before looking at the reflectivity images, we compare the result of application of the operators $(\mathrm{F}\mathrm{F}^{\dagger})$ and $(\mathrm{F}\mathrm{F}^{\dagger})^{\mathrm{T}}$ to observed data in a too low, correct and too high velocity model (figure 4.18). All results are expected to be very close to observed data. In practice, both operators correctly reproduce input data and residuals are mainly located at large surface offsets. However the transposed version $(\mathrm{F}\mathrm{F}^{\dagger})^{\mathrm{T}}$ yields low-frequency artefacts before the primary event, especially for a too high velocity model. These artefacts might have undesirable effects on CIGs.

We now consider the application of the approximate inverse to three data sets: (1) observed data, (2) result of application of $(FF^{\dagger})$ to observed data and (3) result of application of $(FF^{\dagger})^{T}$ to observed data (figure 4.19). The first case corresponds to direct inversion and the third one to the first iterate of preconditioned iterative migration, while the second case is considered only for comparison. The CIGs obtained in the second case are closer to the result of direct inversion than the first iterate of the preconditioned scheme. In particular the low frequency artefacts observed on reconstructed data result in low frequency energy above the reflector in the case of a too high-velocity. This effect is less visible for a correct and a too low initial velocity. Nonetheless, these CIGs show much improvement compared to the adjoint $F^{T}P^{obs}$, first iterate of the non-preconditioned scheme. In particular, migration artefacts are greatly attenuated.

Eventually we plot the value of the normalised MVA objective function obtained in homogeneous background velocity models after one iteration with and without preconditioning, and the one obtained after direct inversion (figure 4.20). With the inverse, the function is perfectly centred at the correct velocity model. Because of residual energy above the reflector, the minimum after one preconditioned iteration is not at the correct velocity model, but it is still more satisfactory than in the case of a single iteration without preconditioning.

This suggests that a single iteration of preconditioned iterative migration is not as efficient as direct inversion, and that additional iterations are required. This is actually not an issue as the preconditioned iterative scheme is not meant to be used in this primary-only case but in the more complex case of multiples for which iterations are required anyway to attenuate cross-talk artefacts.

Results after a few iterations

We now perform ten preconditioned iterations on the same model (figure 4.21). The main effects of iterations on CIGs are additional deconvolution, strengthening of energy at large values of h (in the case of an incorrect velocity model) and attenuation of residual energy above the reflector. We plot the value of the objective function across iterations obtained with and without preconditioning (figure 4.22). We consider the data misfit, as well as the norm of the residuals associated with the linear system (4.43a): $\left\|\left(F^{T}F + a_{\varphi}I\right)\xi - F^{T}P^{obs}\right\|$, which is actually the value of the gradient of J_0 with respect to ξ. As expected, the preconditioned strategy is faster at the first iteration, but both versions reach the same level of data misfit after a few iterations. As in Hou and Symes (2015), we consider a third possibility: iterative migration is performed without preconditioning but with the approximate inverse as initial guess $F^{\dagger}P^{obs}$. In this case very few progress is made at the first iteration because the approximate inverse is already a very good solution to the minimisation problem. Note that we cannot consider the preconditioned iterative minimisation with the approximate inverse as initial value because this would require the ability to compute $(F^{\dagger})^{-1}$.

Eventually we compute the value of the normalised MVA objective function for several homogeneous background velocity models (figure 4.23) after several iterations of the preconditioned scheme. We saw that the minimum of the function does not correspond to the correct value of c_0 after a single iteration (figure 4.20). This is progressively corrected and after five iterations, the objective function of MVA is stable and minimum for the correct velocity.

In summary, we have shown that a single iteration of preconditioned iterative migration

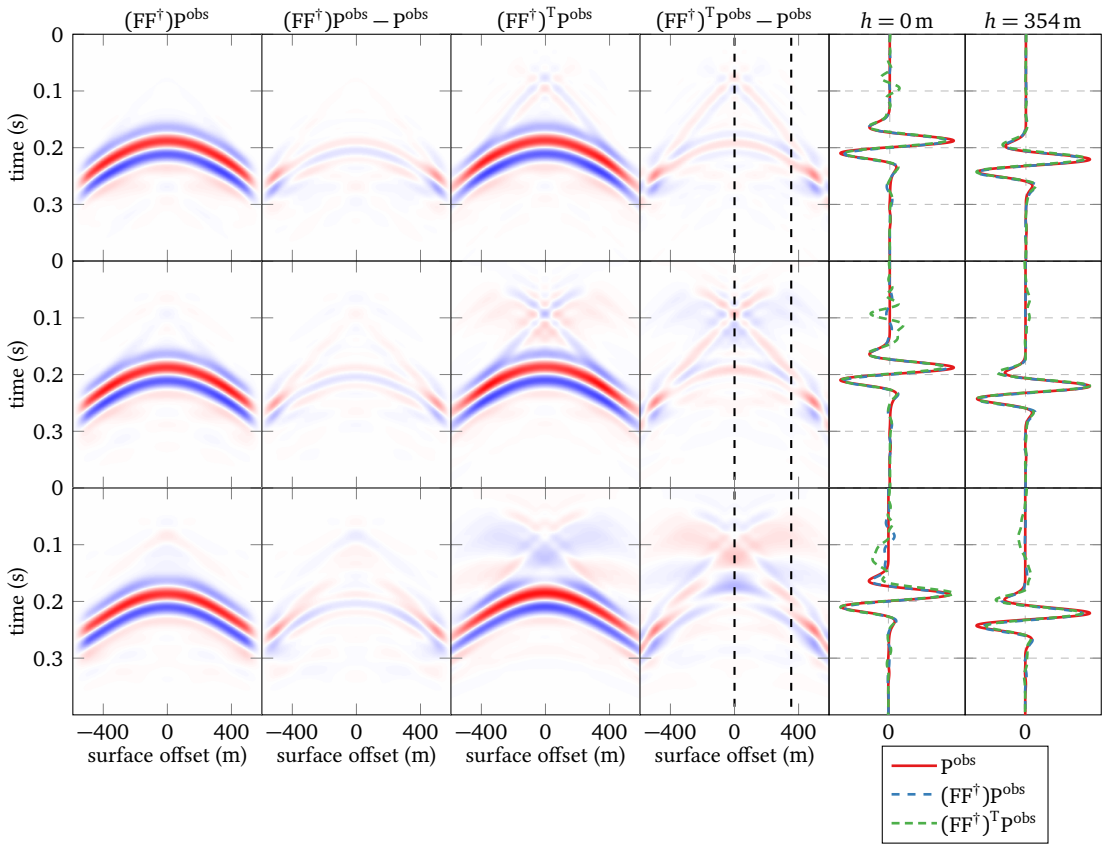


Figure 4.18. – Result of the application of $(FF^\dagger)$ (first column) and $(FF^\dagger)^T$ (third column) to observed data P^{obs}. The corresponding residuals are plotted in the second and fourth column, respectively. All plots share the same colour scale. Observed data are computed in a homogeneous velocity model ($c_0^{ex} = 3000\,m/s$) with a single flat reflector. The velocity model used to compute F and $F^\dagger$ is too low (top, $c_0 = 2500\,m/s$), correct (middle, $c_0 = 3000\,m/s$), and too high (bottom, $c_0 = 3500\,m/s$). The last two columns show traces extracted at positions indicated by dashed lines on the fourth column.

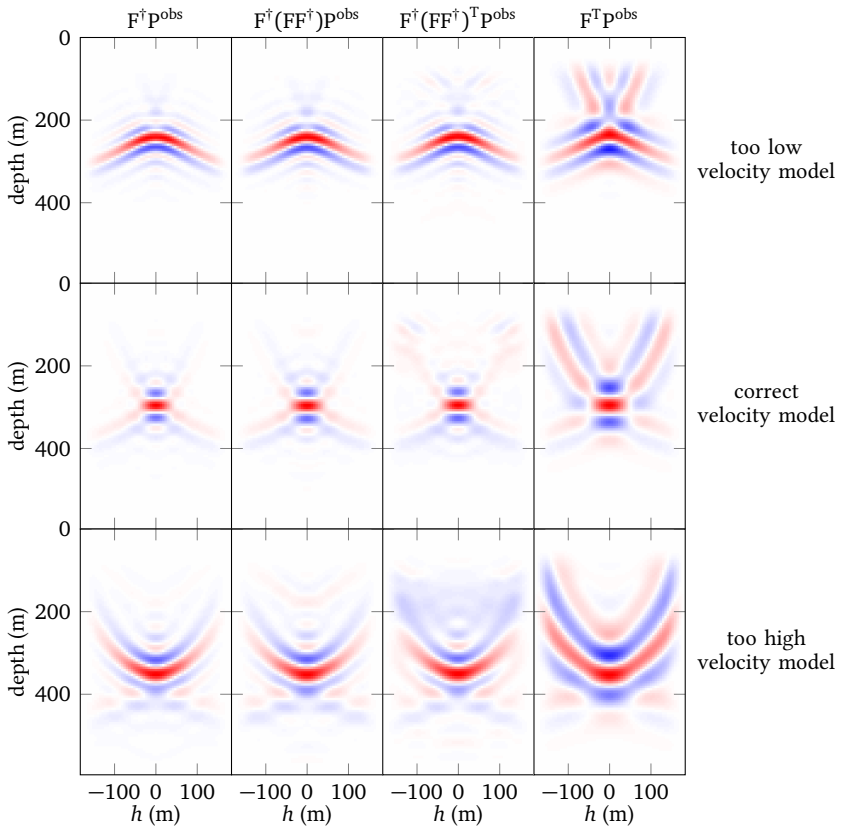


Figure 4.19. – Common Images Gathers obtained in a too low (2500 m/s, top), correct (3000 m/s, middle), and too high (3500 m/s center) background velocity model. CIGs are computed by of application of the pseudo-inverse $F^\dagger$ to observed data (1st column), to observed data premultiplied by $(FF^\dagger)$ (2nd column) and to observed data premultiplied by $(FF^\dagger)^T$ (3rd column). We also consider the application of the adjoint operator F^T to observed data (4th column). The CIGs of the first and third columns correspond to the results obtained by direct inversion as in section 4.2, and to the first iterate of the preconditioned iterative scheme described in section 4.3, respectively. The CIGs of the second and fourth columns are displayed for comparison.

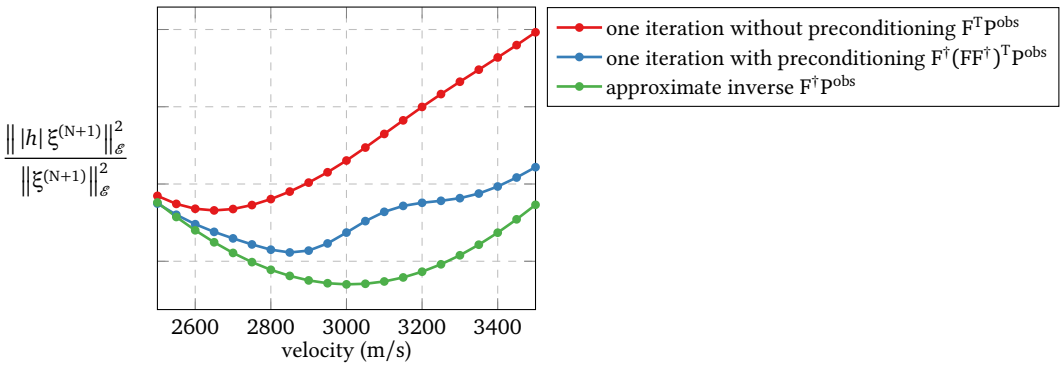


Figure 4.20. – Value of the normalised objective function of MVA obtained in homogeneous velocity models for the approximate inverse and after one iteration with and without preconditioning. The correct velocity is 3000 m/s. The oscillation in the dashed blue curve is due to low frequency energy appearing above the reflector in CIGs especially for too high velocities (third column in figure 4.19).

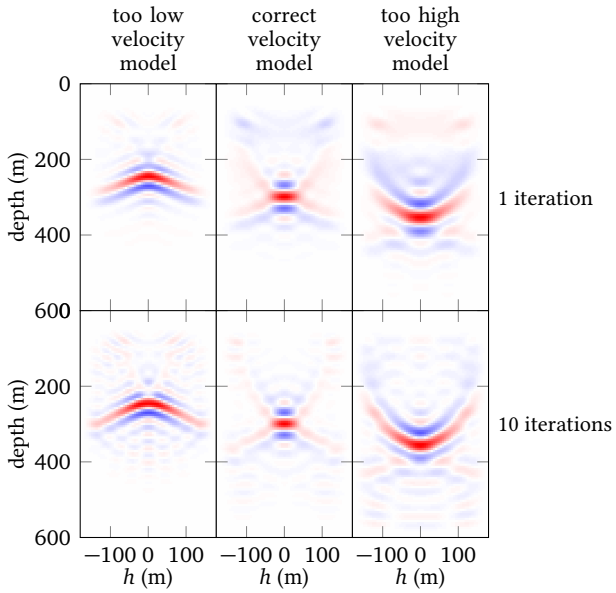


Figure 4.21. – CIGs obtained after one (top) and ten (bottom) iterations of preconditioned iterative migration in a too low (2500 m/s, left), correct (3000 m/s, centre) and too high (3500 m/s, right) velocity model.

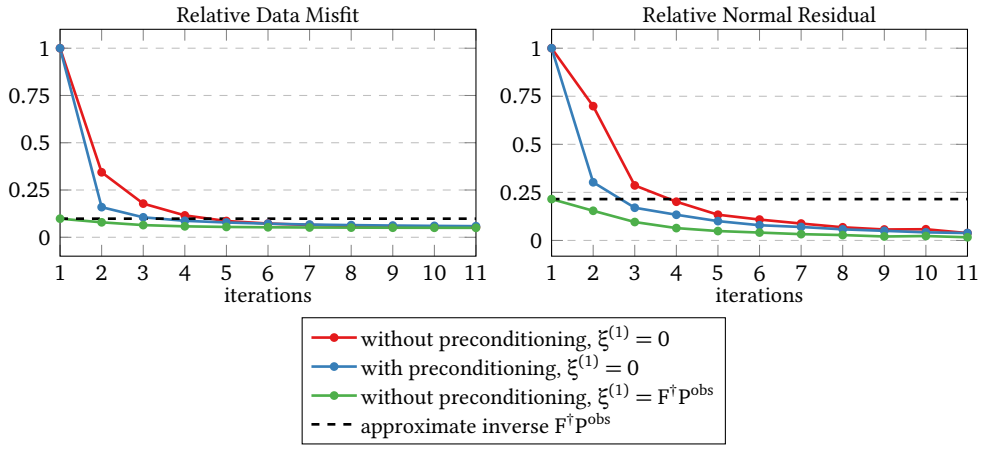


Figure 4.22. – Results of iterative migration in a too slow velocity model (2500 m/s) with primaries only. We plot the relative data misfit (left) and the relative normal residual associated with the linear system (4.43a). Red and blue curves correspond to the case of a zero initial guess ($\xi^{(1)} = 0$) with classical and preconditioned iterative migration, respectively. The green curve corresponds to non-preconditioned migration initiated with the approximate inverse $\xi^{(1)} = F^{\dagger}P^{obs}$. The dashed line corresponds to the value obtained for the approximate inverse $F^{\dagger}P^{obs}$.

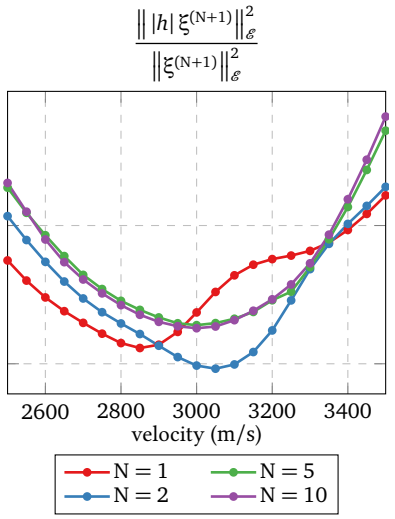


Figure 4.23. – Value of the MVA objective function obtained in homogeneous background velocity models after several iterations with preconditioning. The correct velocity is 3000 m/s.

yields slightly degraded results compared to direct inversion because $(FF^{\dagger})^{T}$ is close but not equal to the identity. Its application to observed data yield small low-frequency artefacts. However the CIGs obtain after a single iteration with the preconditioner is free of migration artefacts, contrary to the non-preconditioned case. Besides the preconditioner also improves the convergence rate of iterative migration.

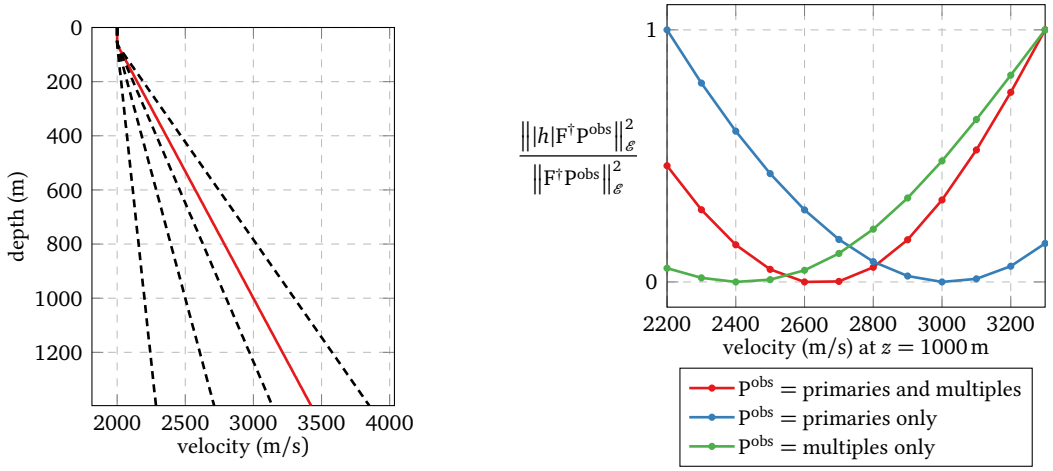
4.4.2. Preconditioned iterative migration with multiples

We now discuss an example with first-order surface-related multiples. We consider an exact model with a background velocity increasing with depth, from 2000 m/s at the surface to 3000 m/s at 1000 m depth (figure 4.24a). The model is 2D but laterally invariant. The reflectivity model consists of a single reflector located at 475 m depth. Contrary to the case of homogeneous models considered in the preceding chapters, we expect events in CIGs corresponding to primaries and multiples (interpreted as primaries) to have different focusing behaviour. In particular, cross-talk artefacts should not be focused for the correct velocity model and should favour too low velocities.

We consider several initial velocity models linearly increasing with depth. All start at the correct velocity at the surface (2000 m/s), but the velocity at 1000 m depth ranges from 2200 m/s to 3300 m/s (figure 4.24a). We first compute CIGs by applying the approximate inverse formula to observed data (figure 4.24c) and plot the corresponding MVA objective function (figure 4.24b). As observed data are computed with the Born approximation, primaries and multiples are available separately and we also plot the value of the objective function obtained for observed data containing primaries only or multiples only with the same range of initial velocity models (figure 4.24a). In the case of primaries only, the objective function is minimum for the correct velocity, similarly to the results of section 4.4.1. Cross-talk artefacts linked to multiples focus for a too low velocity model (figure 4.24c, 3rd column) and curve upward for the correct velocity model (figure 4.24c, 4th column). As a consequence, the minimum of the MVA objective function considering both primaries and multiples is not obtained for the correct velocity model.

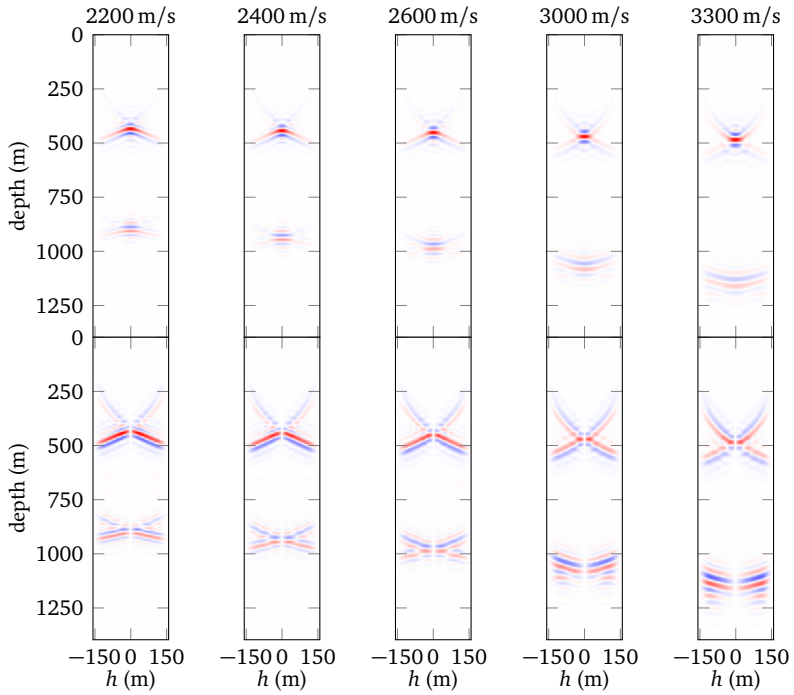
We now perform ten iteration of iterative migration without and with preconditioning using the highest initial velocity model considered in figure 4.24. As in the primaries-only example, we also consider the case of iterative migration without preconditioning starting with the approximate inverse $F^{\dagger}P^{obs}$ as initial guess. To help removing residual energy at large offset, we add regularisation with the Huber norm (Guitton and Symes, 2003). In the first iterations, the migration objective function decreases faster with preconditioning and with the strategy initialised with the approximate inverse (figure 4.26), but the data misfit is similar for the three cases after a few iterations. The final CIGs obtained with the three strategies are very similar (figure 4.25), but the best attenuation of cross-talk artefacts is obtained in the preconditioned case. This is the optimisation strategy used in the following of the study.

Another practical advantage of the preconditioned strategy is related with the determination of an optimal step length. A general issue of linesearch techniques is the choice of an initial guess α_0 initiating the algorithm. Using the preconditioner, the linesearch strategy of Moré and Thuente (1994) returns step lengths in the interval $[0.1, 1]$ at each iteration. Hence $\alpha_0 = 1$ is always a good initial guess. Besides we can even directly set $\alpha = 1$ for the first iteration and



(a) Exact background velocity model (red, solid) and some of the initial models considered (black, dashed). The reflectivity model consists of a single reflector localised at $z = 475\,\mathrm{m}$.

(b) Normalised MVA objective function. Each function has been mapped linearly to the interval $[0, 1]$. The correct model corresponds to a velocity of 3000 m/s at 1000 m depth.



(c) Central CIGs $\xi^{\mathrm{inv}} = \mathrm{F}^\dagger\mathrm{P}^{\mathrm{obs}}$ obtained for different background velocity models (top) and application of the annihilator to these CIGs $\mathrm{A}\xi^{\mathrm{inv}}/\|\xi^{\mathrm{inv}}\|_{\mathscr{E}}$ (bottom). The plots of the bottom row have the same colour scale.

Figure 4.24. – Result of application of the approximate inverse $\mathrm{F}^\dagger$ to observed data containing both primaries and first-order surface-related multiples with several initial background velocity model increasing with depth.

save the computational cost of a linesearch. The first gradient is very close to the approximate inverse, hence setting $\alpha = 1$ yield a reflectivity model $\xi^{(2)}$ very well explaining primaries, so that residuals are mostly due to multiples.

Finally we consider the same range of initial velocity models as in figure 4.24b and we compute the value of the MVA objective function obtained after several iterations of preconditioned iterative migration (figure 4.27). After a single iteration, we obtain a result similar to the direct inversion case with a minimum obtained for a too low velocity model. With iterations cross-talk artefacts are attenuated and the minimum is progressively shifted to the correct velocity model.

4.4.3. Preconditioned resolution of the adjoint problem and associated gradient

Eventually we consider the computation of the gradient of J_1 with method B as described in section 3.2.2. We go back to the primaries-only example of section 4.4.1 and test if the preconditioner accelerates the convergence rate of the adjoint problem. We will show that the results are not fully satisfactory; hence we defer to chapter 5 the numerical applications in the case of multiples (section 5.3.2).

We assume that the direct problem has been solved in seven iterations, meaning that the final reflectivity is $\xi^{(8)}$. The adjoint problem then consists of finding the solution $\eta \in \mathscr{E}$ of

$$\mathrm{H}\eta = b, \tag{4.46}$$

where H is the Hessian matrix of J_0, which is independent of ξ in the linear case, and $b = \mathrm{A}^{\mathrm{T}}\mathrm{A}\xi^{(8)}$ is the image residual. A is the annihilator consisting of a multiplication by the subsurface offset and a power of the background velocity model,

$$[\mathrm{A}\xi](\mathbf{x}, h) = c_0^{\beta}(\mathbf{x})|h|\xi(\mathbf{x}, h). \tag{4.47}$$

We choose $\beta = 3/2$, the value resulting in smooth gradients free of artefacts around the reflector in the case of direct inversion (section 4.2). The adjoint problem is solved with ten iterations of the conjugate gradient algorithm with and without preconditioner. We compare the convergence rate obtained with both strategies. Contrary to the direct problem, there is no norm in the observed data-space $\mathscr{D}_{\text{obs}}$ associated with the resolution of equation (4.46) which can be used as a convergence criteria. Here the conjugate gradient algorithm actually minimises

$$\varphi[\eta] = \frac{1}{2}\langle \eta \,|\, \mathrm{H}\eta \rangle_{\mathscr{E}} - \langle b \,|\, \eta \rangle_{\mathscr{E}}. \tag{4.48}$$

A difficulty with this objective function is that we do not know the value obtained for the ideal solution of the linear problem. Therefore we will also look at the relative normal residuals of the linear system (4.46), defined as

$$\mathrm{J}^{\text{Normal Res.}}[\xi] = \frac{\|\mathrm{H}\xi - b\|_{\mathscr{E}}}{\|b\|_{\mathscr{E}}}, \tag{4.49}$$

which can be interpreted as the normalised norm of the gradient of φ (equation 4.48). Note that contrary to φ, there is no guarantee that the value of the normal residual decreases at each iteration.

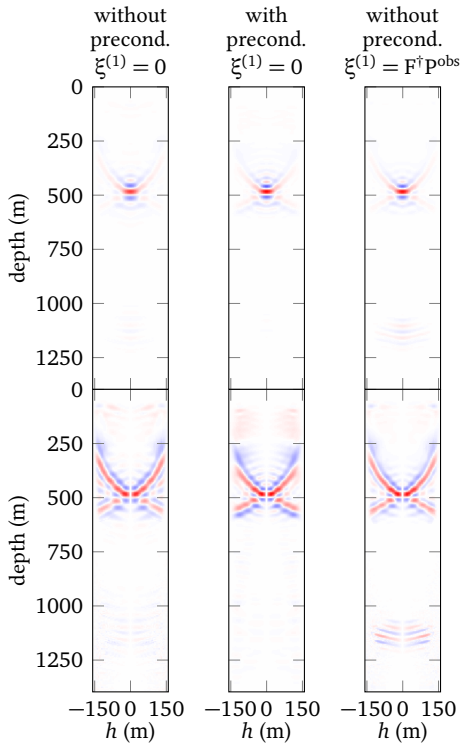


Figure 4.25. – CIGs obtained after ten iterations of migration without (left) and with (centre) preconditioning starting with $\xi^{(1)} = 0$, and ten iterations of migration without preconditioning starting with $\xi^{(1)} = F^\dagger P^{obs}$ (right).

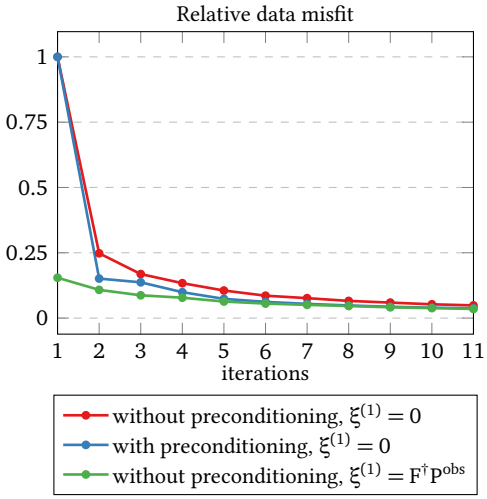


Figure 4.26. – Decrease of the migration objective function using three optimisation strategies, corresponding to the CIGs figure 4.24.

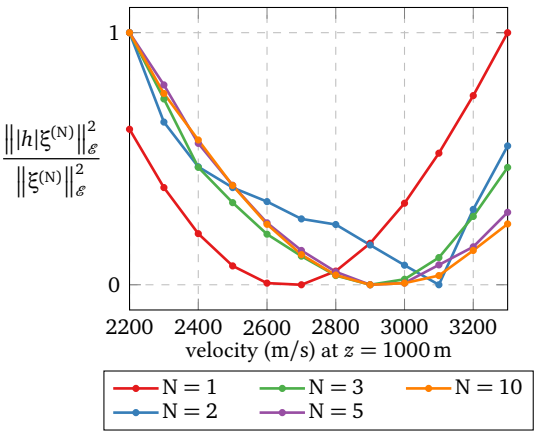


Figure 4.27. – Normalised MVA objective function obtained after several iterations of preconditioned iterative migration. Each function has been mapped linearly to the interval $[0, 1]$. The correct model correspond to a velocity of 3000 m/s at 1000 m depth.

Small regularisation

We first consider a case with small regularisation, as in section 4.4.1. Using the same right-hand side term b, that is the same value for $\xi^{(8)}$, we solve the adjoint problem in ten iterations, without and with preconditioning (figure 4.28). We obtain a much lower convergence rate than for the direct problem, even with the preconditioner. The normal residual associated with the last value $\eta^{(8,11)}$ is just a little lower than for the initial value $\eta^{(8,1)}$. At the last iteration, the adjoint variable $\eta^{(8,11)}$ is very oscillating (figure 4.29), and residuals of the linear system are located mainly near the surface, that is away from the reflector position. This is related to the observations made in chapter 3: weak energy at large values of h, which does not bear relevant kinematic information, has large influence on the resolution of the adjoint problem. Besides, note that the zero value at $h = 0$ in b is not recovered in $H\eta^{(8,11)}$.

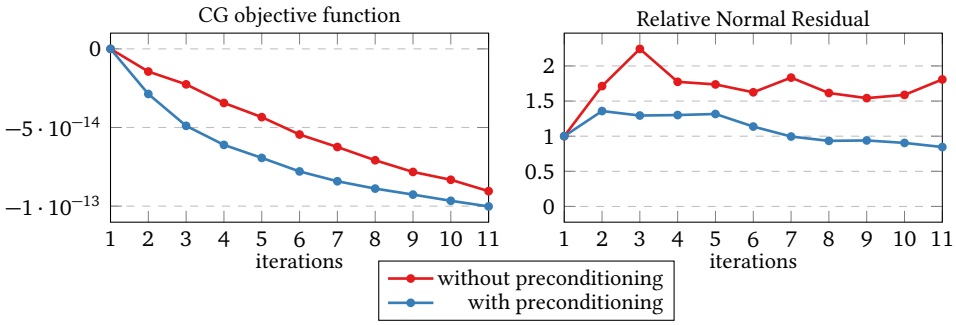


Figure 4.28. – Convergence of the adjoint problem in the case of small regularisation.

An undesirable consequence of the slow resolution of the adjoint problem is that the value of the associated gradient $G^{(7,M)}$ does not converge to a stable final value (figures 4.30 and 4.31). The sign of the gradient may even change across iterations (for example $G^{(7,4)}$ in the preconditioned case). Without preconditioning, the gradients exhibit vertical spurious oscillations very similar to those observed in the preceding chapter. When the adjoint problem is solved with the preconditioner we do not observe these artefacts. However the final value is quite different from the one obtained after direct inversion, which is much smoother and homogeneous (figure 4.32). More investigation is needed to understand these results (chapter 5). In the following we examine if a stronger regularisation helps to overcome these difficulties.

Stronger regularisation

As in chapter 3, we consider increasing the value of the regularisation coefficient a_φ. The convergence of the adjoint problem is improved but still much slower than for the direct problem (figure 4.33). The preconditioner does not clearly improve the convergence rate. This can be explained by the fact that the operator $(F^\dagger)^T(F^TF + a_\varphi I)F^\dagger$ is theoretically close to the identity only if the regularisation weight is not too high. The final value $\eta^{(8,11)}$ is still very oscillating (figure 4.34). Although the residuals of the linear system are weaker in the preconditioned case, the zero value at $h = 0$ in b is not recovered in $H\eta^{(8)}$ and there are still strong residuals away from the reflector.

With stronger regularisation, the spurious oscillations appearing when no preconditioning is applied (figure 4.30) are greatly attenuated (figure 4.35), as already noticed in chapter 3. With

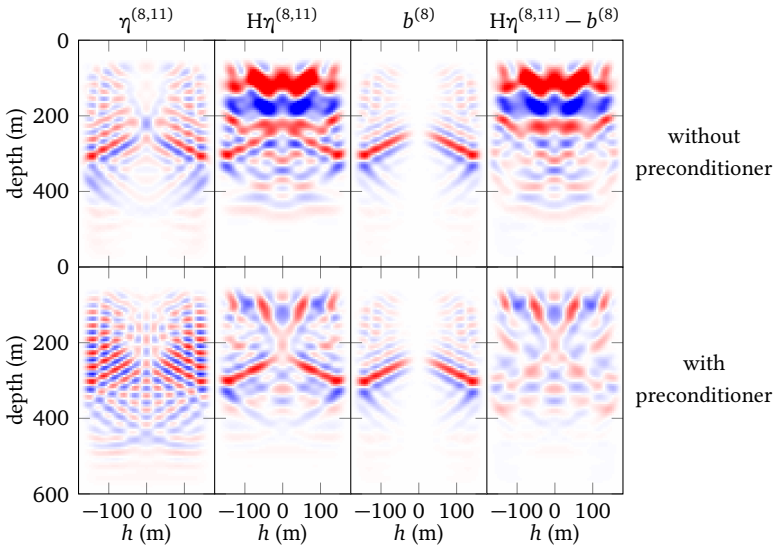


Figure 4.29. – Residuals related to the linear adjoint problem (4.46) after ten iterations. We show the central CIG of the adjoint variable η (1st column) obtained after ten iterations without preconditioning (top) and with preconditioning (bottom). Then we compare $H\eta$ (2nd column) with the right-hand side of the adjoint problem (3rd column) and compare this two vectors. On each line, the three most right plots share the same colour scale.

the preconditioner, the gradients are smoother and more homogeneous (figures 4.35 and 4.36), but do not converge to a stable solution. Besides they are still quite different from the gradient obtained by direct inversion (figure 4.37). In particular there are still residual oscillations around the reflector position.

As a conclusion, we note that although direct inversion and iterative migration yield very similar reflectivity images, the associated gradients are quite different. In particular, the use of the preconditioner does not relax the need for strong regularisation to obtain smooth velocity updates. An other issue associated with the iterative case is the difficult resolution of the adjoint problem, and the instability of the associated gradient. Further investigation is needed to understand this unwanted behaviour and design a more robust strategy. This is the purpose of chapter 5.

4.5. Conclusion

In this chapter, we have introduced an approximate inverse of the extended Born modelling operator. It is similar to the one proposed by Hou and Symes (2015), the main difference being that the derivation presented here uses a linearisation of the phase of $FF^{\dagger}$. The implementation of the approximate inverse resembles the adjoint's one with a similar computational cost. A direct inversion strategy consisting of applying this operator to observed data yields CIGs free of migration artefacts. Introducing a specific power of the background velocity update in the associated MVA objective function results in a smooth and homogeneous gradient free of

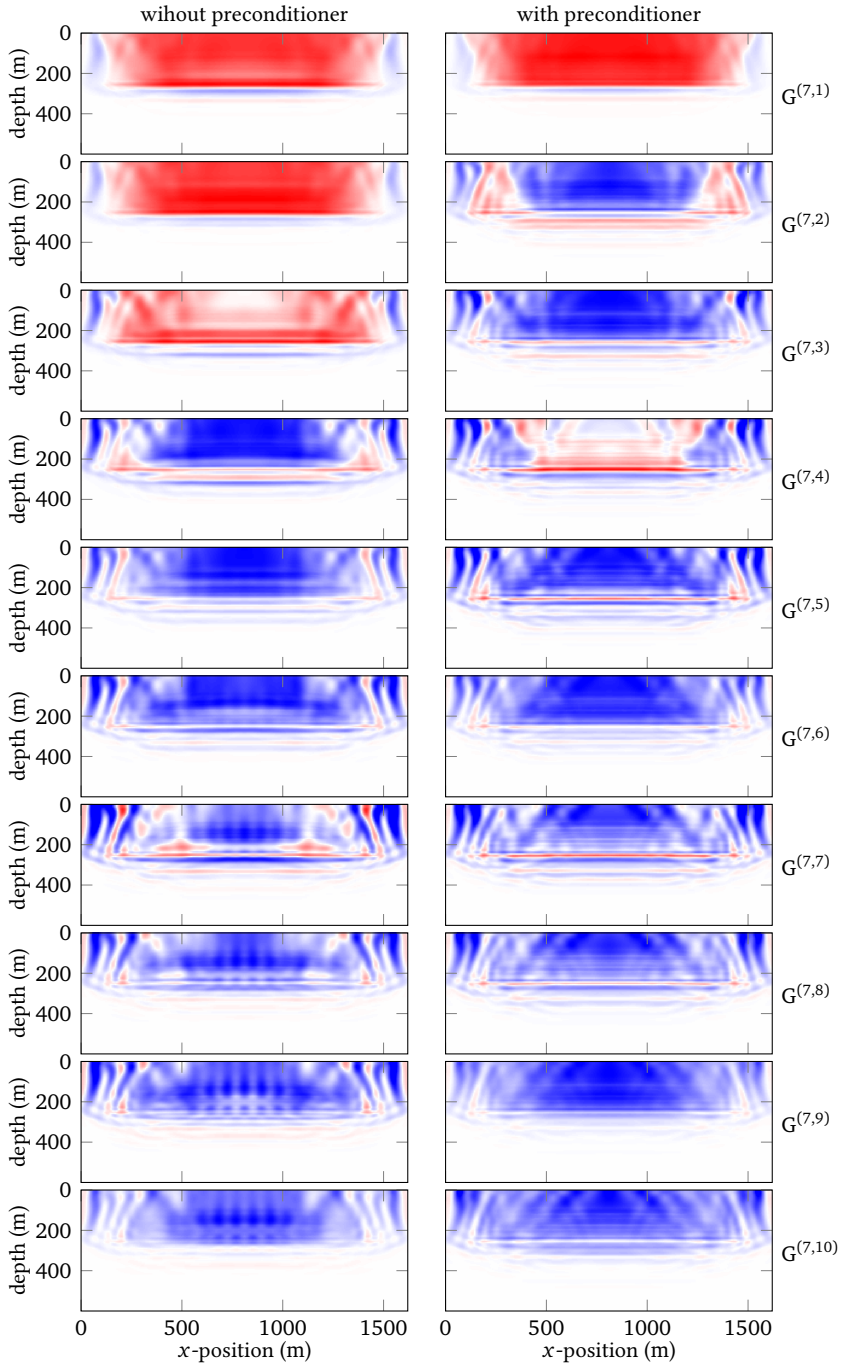


Figure 4.30. – Gradients obtained, in the case of small regularisation, after seven iteration of preconditioned iterative migration for the successive values of the adjoint variable (from top to bottom). The left and right columns correspond to the resolution of the adjoint problem without and with preconditioner, respectively. The gradient is not really stable from one iteration to another.

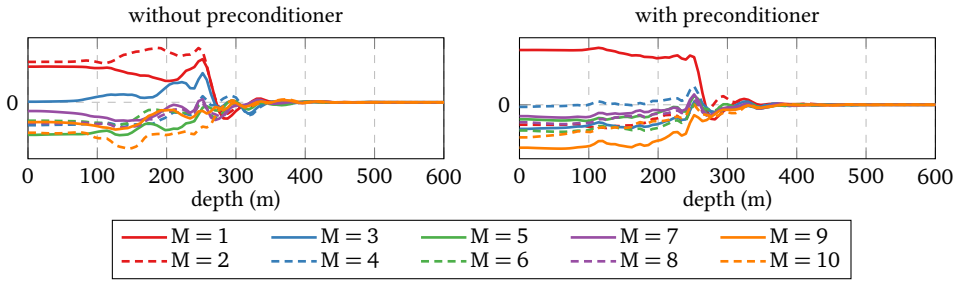


Figure 4.31. – Central trace ($x = 810\,\text{m}$) of the gradients shown in figure 4.30. The value of the gradient above the reflector oscillates from one iteration to another.

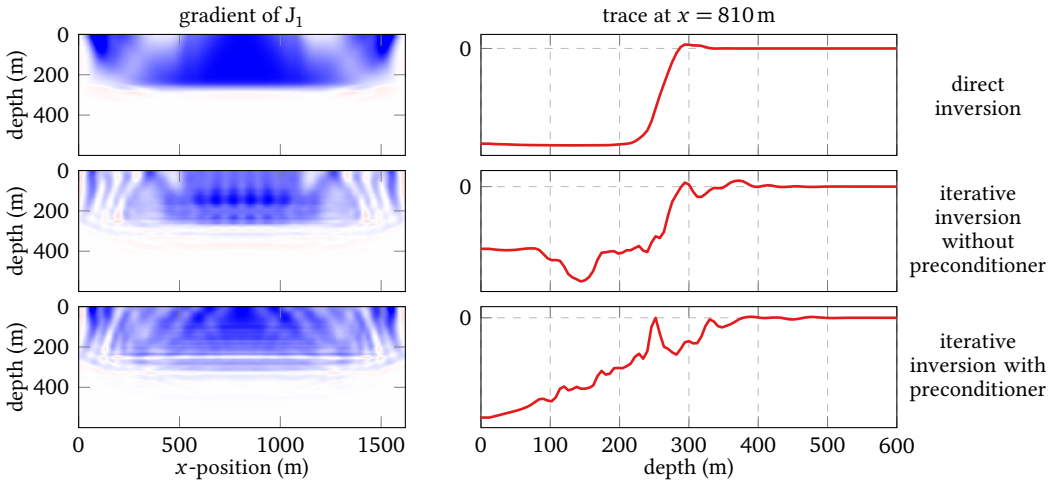


Figure 4.32. – Gradients obtained by direct inversion (top), and after iterative inversion (middle without preconditioning and bottom with preconditioning) with small regularisation (N = 7 and M = 10, corresponding to the bottom row of figure 4.31). The right column shows a trace extracted at the middle position (corresponding to the dashed orange lines in figure 4.31 for the iterative cases).

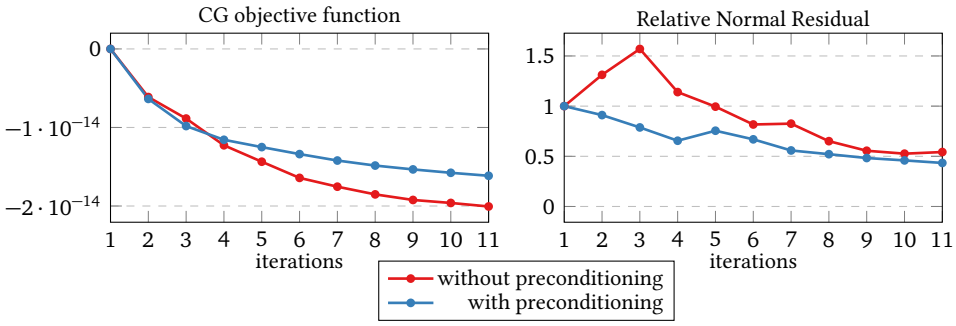


Figure 4.33. – Convergence of the adjoint problem in the case of stronger regularisation.

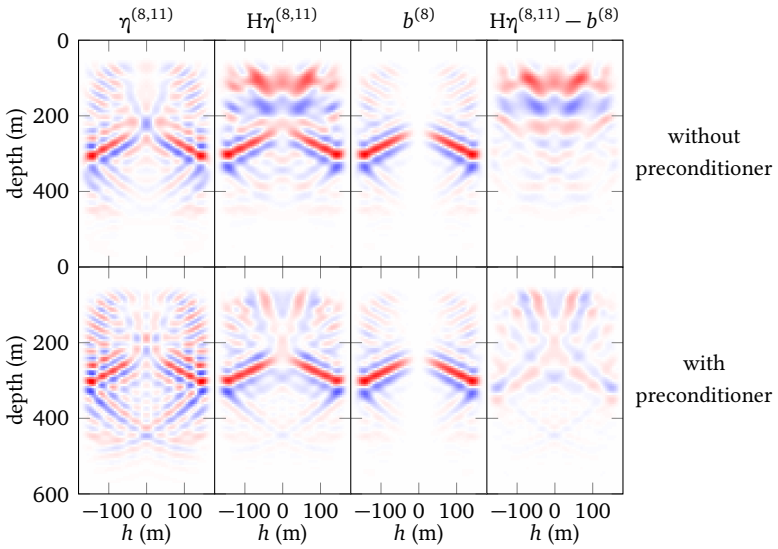


Figure 4.34. – Same as figure 4.29 in the case of stronger regularisation.

oscillations around the reflector.

The approximate inverse operator has been designed for primaries only. Iterative migration is not required any more in this case, at least for regularly sampled input data, but is still needed to handle multiple reflections. The approximate inverse has been introduced as a preconditioner to accelerate the resolution of both the direct and adjoint problems. Numerical examples show that the preconditioner allows a faster convergence of the direct problem. Despite the similarity between the direct and adjoint problems, the convergence rate of the latter is much slower, even with the preconditioner. As a consequence the gradient obtained for successive values of the adjoint problem is not stable and is quite different from the one obtained after direct inversion. Larger regularisation improves the smoothness of the gradient, but this is not a fully satisfactory solution as the adjoint problem still converges much slower than the direct problem. Besides, the preconditioner is not helpful when strong regularisation is applied.

These results may appear somehow surprising as direct inversion and iterative migration provide very similar reflectivity images. If we refer to chapter 3, the reason is that additional deconvolution in the case of iterative migration results in small oscillations in the final reflectivity away from the reflector. These differences allow to further reduce the data misfit but have large influence on the MVA objective function and its gradient, leading to very different background velocity updates.

To better understand these issues and to improve the stability of the gradient computation in the iterative case, we conduct in the next chapter a pure 1D-analysis similar to the one presented in section 4.2.5 and propose a slightly modified MVA objective function leading to a more stable procedure.

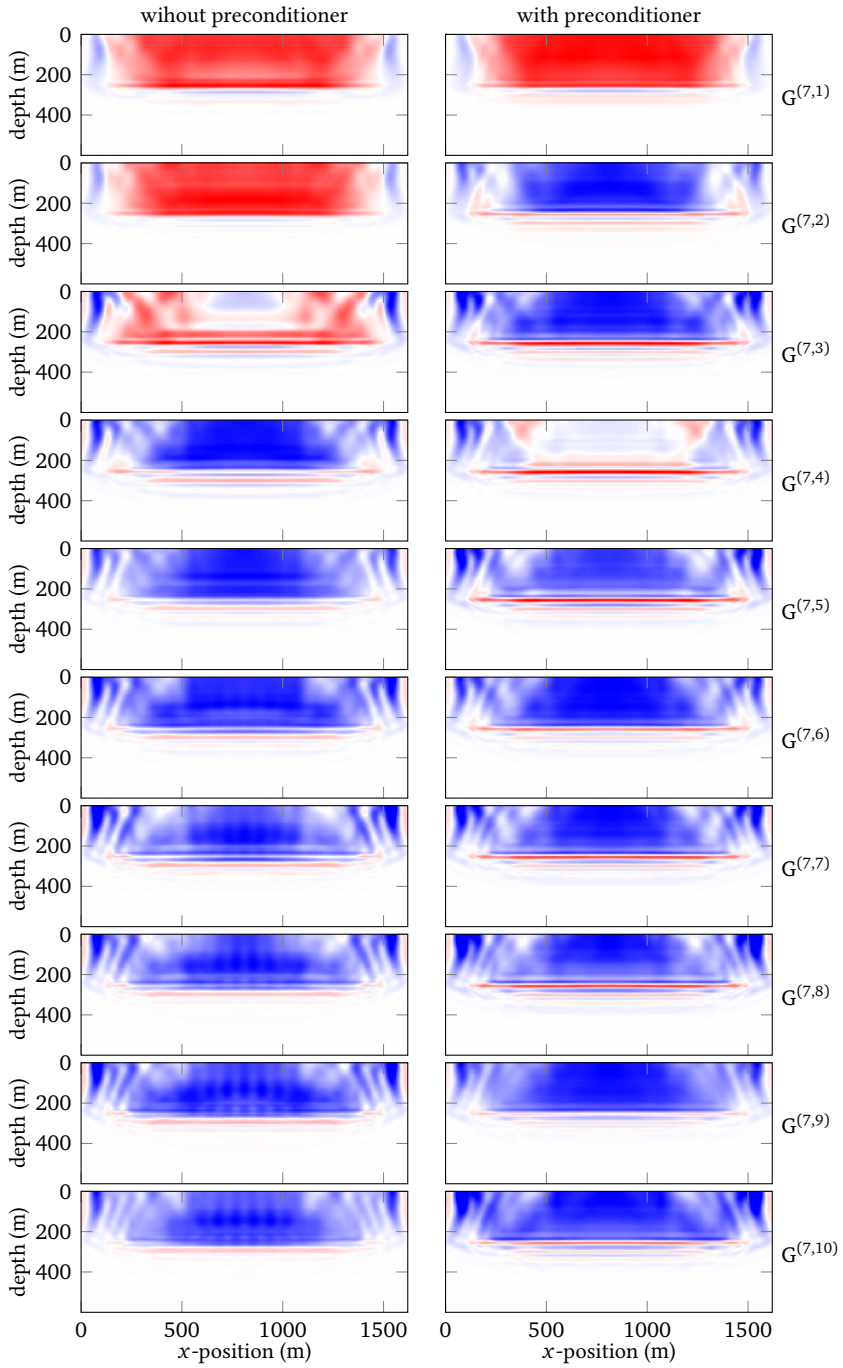


Figure 4.35. – Same as figure 4.28, but with a stronger regularisation.

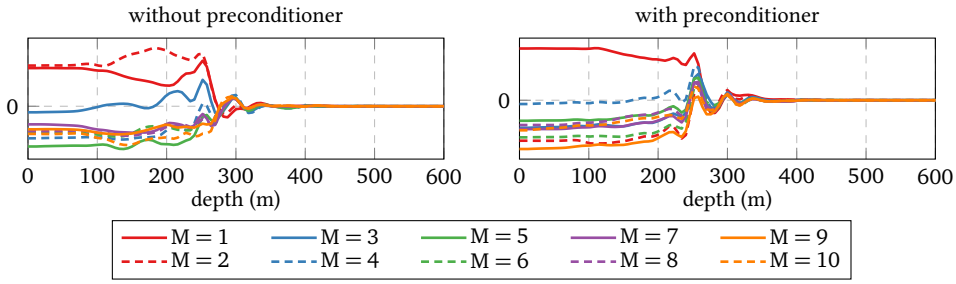


Figure 4.36. – Central trace ($x = 810\,\mathrm{m}$) of the gradients shown in figure 4.35.

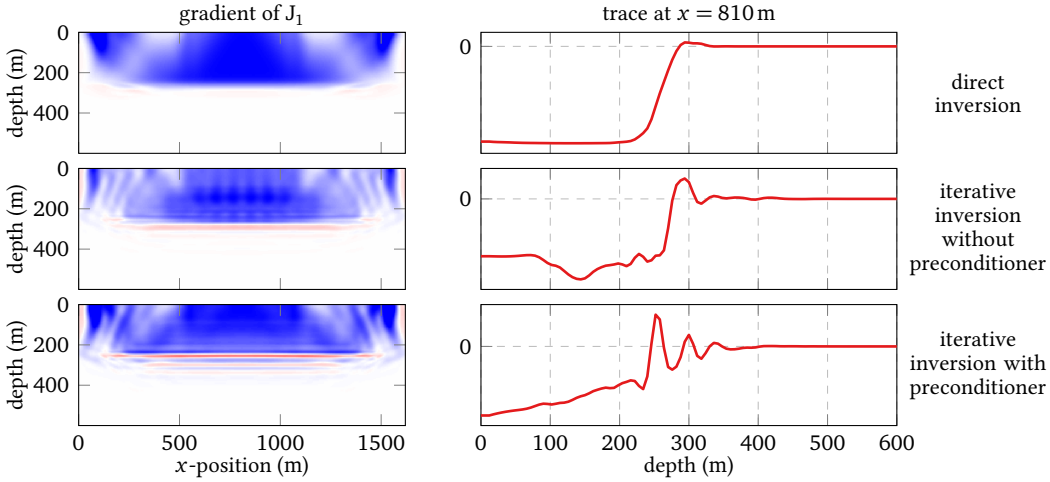


Figure 4.37. – Same as figure 4.37 with stronger regularisation.

Chapter 5.

Regularised Migration Velocity Analysis

Contents

Résumé du chapitre 5

Ce chapitre présente la formulation finale retenue pour l'*analyse de vitesse par migration itérative*. Il débute par une analyse dans un cas purement 1D des gradients de la DSO obtenus par inversion directe (section 4.2) et par la stratégie itérative (suivant la méthode B définie au chapitre 3). Dans le cas purement 1D, l'acquisition ne comprend qu'une seule source et qu'un seul récepteur, et le modèle de réflectivité n'est paramétré que par la profondeur, sans paramètre additionnel équivalent de l'offset en profondeur considéré dans le cas 2D. La formulation de la DSO en 1D n'est valable que dans le cas d'un unique réflecteur : la qualité du modèle de vitesse est évaluée en pénalisant l'énergie éloignée de la position exacte du réflecteur dans le vrai modèle. Ainsi cet analogue 1D est artificiel et ne peut pas reproduire toute la complexité d'une acquisition 2D et du modèle étendu avec l'offset en profondeur. En revanche il présente des caractéristiques similaires au cas 2D concernant l'instabilité du gradient par la méthode itérative pour un coût

de calcul bien moindre. En outre, il permet une représentation des opérateurs de modélisation et de migration sous forme de matrices.

Comme dans le cas 2D, les instabilités dans le calcul du gradient sont liées à l'apparition d'énergie loin du réflecteur et non porteuse d'information cinématique, mais qui influence grandement la fonction coût de la DSO. Pour stabiliser ce calcul, je propose de faire porter la régularisation non pas sur la boucle interne comme étudié au chapitre 3, mais sur la boucle externe. La régularisation prend ici la forme d'un opérateur appliqué au résultat de la migration itérative avant pénalisation par l'annihilateur. Cet opérateur peut être interprété comme un filtre atténuant l'énergie non porteuse d'information cinématique dans le modèle de réflectivité. Il est défini ici par $\mathbf{F}^{\dagger}\mathbf{F}$, composition du pseudo-inverse avec l'opérateur de modélisation. Il est important de noter que $\mathbf{F}^{\dagger}$ est construit de telle sorte que $\mathbf{F}\mathbf{F}^{\dagger}$ soit proche de l'opérateur identité dans le domaine des données. En revanche $\mathbf{F}^{\dagger}\mathbf{F}$ est différent de l'identité dans le domaine image, en particulier il atténue les évènements dans les CIGs qui perturbent l'analyse de vitesse.

L'effet de cette régularisation est illustré dans les cas 1D et 2D. Il est possible de montrer qu'appliquer l'opérateur $\mathbf{F}^{\dagger}\mathbf{F}$ au résultat de la migration itérative produit, dans le cas de primaires seuls, un modèle de réflectivité étendu proche du résultat obtenu par application directe du pseudo-inverse aux données observées. Le calcul du gradient de la fonction coût externe est aussi beaucoup plus stable, sans condition particulière sur le poids donné à la régularisation dans la boucle interne, et le résultat également proche de celui obtenu dans le cas de l'inversion « directe ». L'opérateur $\mathbf{F}^{\dagger}\mathbf{F}$ étant fonction du macro-modèle, des termes additionnels doivent être calculés pour obtenir le gradient. Cependant la résolution du problème adjoint étant accélérée par l'introduction de cet opérateur, le coût numérique total de la méthode est diminué avec l'ajout de la régularisation.

Dans le cas de multiples, la régularisation sur la boucle externe stabilise aussi le calcul du gradient, mais la régularisation sur la boucle interne demeure essentielle. Finalement, une stratégie alternative est étudiée dans le cas des multiples. Elle ne consiste à calculer qu'une seule contribution du gradient total de la boucle externe ; la remise à jour ainsi obtenue est dénommée « gradient tronqué » dans la suite du manuscrit. En particulier la résolution du problème adjoint n'est plus nécessaire, ce qui réduit grandement le coût de calcul de la remise à jour du macro-modèle. Physiquement, cette simplification signifie que la migration itérative est utilisée pour obtenir un modèle de réflectivité dans le domaine étendu sans artefacts de migration et sans artefacts dus aux multiples. Ce modèle de réflectivité est utilisé pour régénérer un nouveau jeu de données constitué de réflexions primaires uniquement, cinématiquement cohérentes avec celles enregistrées dans les données observées. Finalement la stratégie d'inversion directe décrite dans la section 4.2 est appliquée à ce nouveau jeu de données.

5.1. Introduction

In this chapter, we analyse on a simple 1D case the behaviour of iterative migration velocity analysis. The advantage of this simple setting is the possibility to explicitly build modelling and migration operator and to run a large number of iterations. First we summarise the results obtained in 2D in the preceding chapters.

- In chapter 2, iterative migration has been introduced and we have shown that migration artefacts as well as cross-talk artefacts appearing on CIGs at the first iteration are progressively attenuated (figure 2.15);

- In chapter 3, we compared two methods yielding an approximate gradient of the MVA objective function after iterative migration. In the first approach, we derived the exact gradient of an approximation of the ideal objective function. In the second one, we compute an approximate gradient of this ideal objective function. The second approach has a simpler implementation. The gradient is obtained in two steps. First an adjoint variable in the extended model space has to be computed as the solution of a linear problem, even in the presence of multiples. Then the gradient is computed using the last iterate of the direct and adjoint problems. Numerical evidence show that regularisation has to be introduced in iterative migration to ensure the smoothness of the gradient (figures 3.5 and 3.7) and to prevent the apparition of spurious oscillations;

- In chapter 4, we have introduced an approximate inverse of the extended Born modelling operator. It is derived under the high frequency approximation, but involves wave-equations operators only. For data containing primaries only, the application of this pseudo-inverse to observed data yields CIGs free of migration artefacts. The gradient of the associated MVA objective function is smooth and free of oscillations around the reflectors, provided that a specific power of the background velocity is introduced in the annihilator. We showed how this approximate inverse can be used as a preconditioner to speed up the resolution of iterative migration (figure 4.22), needed in the presence of multiples.

However several difficulties remain, investigated in this chapter:

- If the modelling is linear, the direct and adjoint problems are very similar. They consist of the same system with two different source terms,

$$\mathrm{H}\xi = a \quad \text{for the direct problem,} \tag{5.1a}$$
$$\mathrm{H}\eta = b \quad \text{for the adjoint problem,} \tag{5.1b}$$

  where H is the Hessian matrix of the migration objective function. However the resolution of the adjoint problem requires in practice more iterations than the direct problem. Using the approximate inverse operator as a preconditioner, a single iteration already provides a very good solution for the direct problem, but this is not the case for the adjoint problem (figure 4.28);

- Although direct and iterative inversion result in very similar reflectivity images, they lead to quite different background velocity updates (figure 4.32). In the iterative case, the gradient still has oscillations around the reflector and modifying the MVA objective function with a power of the background velocity does not allow to remove these oscillations.

In order to gain a better understanding of these issues and to propose solutions addressing them, we study here a 1D-case defined in the same fashion as in section 4.2. Although the 1D-case cannot reproduce all the features of 2D subsurface-offset extended migration, its reduced dimensionality allows to explicitly build and display modelling and inverse operators as matrices, and to considerably reduce computation time, so that performing tens or hundreds of iterations for solving both the direct and adjoint problems is affordable. For example, it becomes possible to study the distribution of eigenvalues of the normal operator.

In this analysis we consider only the linear case of primaries. We have shown that direct inversion works already well in this case; the objective here is to understand why the iterative case is unstable and to propose a robust solution to this issue. At the end of the chapter, we discuss on a 2D numerical example how the conclusions can be extended to the case of multiples. We begin by presenting the 1D MVA approach. After a brief review of direct inversion with the approximate inverse operator, we detail the computation of the gradient in the case of iterative migration. A simple numerical example illustrates that the 1D and 2D cases have a similar behaviour. Eventually we propose a modification of the iterative MVA procedure to alleviate the difficulties listed above.

In the second part of the chapter, this new strategy is applied on 2D examples, first on primaries only, then on an example with primaries and first-order surface multiples. The objective is to determine if the new approach defined in 1D exhibits the same benefits in the 2D case.

5.2. Analysis on a pure 1D case

5.2.1. Presentation of the 1D case

We consider a pure 1D case, similar to the one studied in section 4.2.5, with a single source and a single receiver located at the same position. In 2D, an extended-model space $\mathscr{E}$ is defined by adding an extension parameter to the model space $\mathscr{M}$. This is not necessary in 1D as the model space $\mathscr{M}$ and the observed data space $\mathscr{D}_{\text{obs}}$ already have the same dimension, meaning that vectors of $\mathscr{E} = \mathscr{M}$ are parametrised only by depth z. However we keep distinct notations $\mathscr{E}$ and $\mathscr{M}$ for consistency with the 2D case. In the following, the reflectivity is defined as a perturbation of the background velocity model[1],

$$\xi(z) = \frac{\delta c(z)}{2c_0(z)}, \tag{5.2}$$

and the reflectivity model consists of a single reflector located at z_{ex} with a reflection coefficient R_{ex},

$$\xi(z) = R_{\text{ex}}\delta(z - z_{\text{ex}}). \tag{5.3}$$

[1] another definition for the reflectivity (e.g. $\xi(z) = 2\delta c(z)/c_0^3(z)$, as in equation 2.8) would modify the gradient expressions, but not the conclusions.

The annihilator $A : \mathscr{E} \mapsto \mathscr{E}$ is diagonal and defined as a multiplication by the distance to the exact reflector position,

$$[A\xi](z) = a(z)\xi(z) = |z - z_{ex}|\xi(z), \tag{5.4}$$

where z_{ex} is the exact depth of the reflector. Note that this 1D-MVA approach is valid for reflectivity models with a single reflector only, but has a similar structure as the classical 2D DSO formulation. Then the MVA objective function is defined in the general form as

$$J_1[c_0] = \frac{1}{2}\left\|AC\xi[c_0]\right\|^2_{\mathscr{E}}, \tag{5.5}$$

where the diagonal operator $C : \mathscr{E} \mapsto \mathscr{E}$ consists of a multiplication by c_0^β. As for the 2D case, using a specific value for β may allow to remove unwanted oscillations in the gradient of J_1 (section 4.2).

Here we consider the case where ξ is defined as the result of iterative minimisation of

$$\begin{aligned} J_0[c_0, \xi] &= \frac{1}{2}\left\|P[c_0, \xi] - P^{obs}\right\|^2_{\mathscr{D}_{obs}} + a_\varphi \varphi[\xi] \\ &= \frac{1}{2}\left\|F[c_0]\xi - P^{obs}\right\|^2_{\mathscr{D}_{obs}} + a_\varphi \varphi[\xi], \end{aligned} \tag{5.6}$$

for a fixed background velocity model c_0, with $F : \mathscr{E} \mapsto \mathscr{D}_{obs}$ the Born modelling operator and $\varphi : \mathscr{E} \mapsto \mathbb{R}$ the regularisation function.

The source wavelet and its approximate inverse are noted Ω and $\widetilde{\Omega}$ respectively. In the numerical example, Ω is a Ricker with maximal frequency 40 Hz (figure 5.1). The exact and the initial velocity model are homogeneous ($c_0^{exact} = 3000\,m/s$ and $c_0 = 2500\,m/s$). The reflectivity consists of a single reflector located at $z_{ex} = 400\,m$ with $R_{ex} = 0.2$. To be consistent with the 2D-implementation, we introduce a taper $T_{\mathscr{E}}(z)$ (figure 5.2) in the $\mathscr{E}$-space by considering that J_0 is function of $(T \cdot \xi)$ instead of ξ. The taper will not be indicated in the following equations for the sake of clarity, but is taken into account in the implementation, including for the gradient expressions.

With the assumption of single scattering, data depend linearly on the reflectivity model and, for a fixed velocity model c_0, we can explicitly construct the (rectangular) matrices (figure 5.3) corresponding to the following operators

- $F : \mathscr{E} \mapsto \mathscr{D}_{obs}$: modelling of primary reflections;
- $F^T : \mathscr{D}_{obs} \mapsto \mathscr{E}$: migration of observed data;
- $F^\dagger : \mathscr{D}_{obs} \mapsto \mathscr{E}$: approximate inverse of F, defined in section 4.2;
- $(F^\dagger)^T : \mathscr{E} \mapsto \mathscr{D}_{obs}$: adjoint of $F^\dagger$, which can be seen as a modified modelling operator.

Given $\xi \in \mathscr{E}$ and $P \in \mathscr{D}_{obs}$, these operators can be expressed with Green's functions, as in

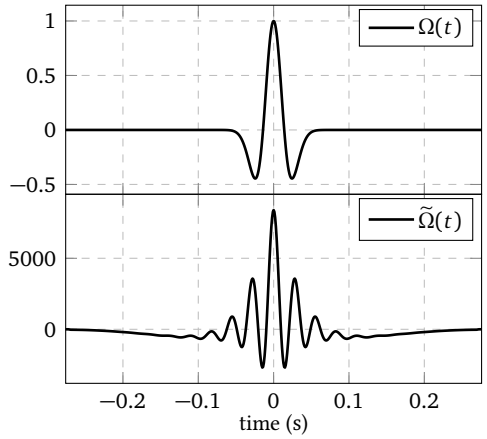


Figure 5.1. – Source wavelet Ω and its approximate inverse $\widetilde{\Omega}$.

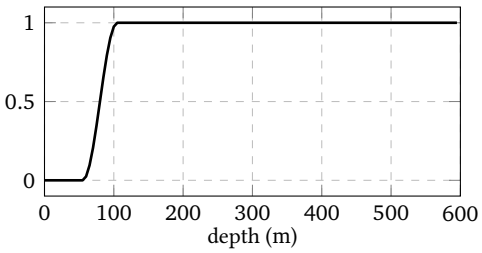


Figure 5.2. – Taper applied in the model space.

equations (4.8) in chapter 4, leading to

$$\begin{cases} \big[\mathrm{F}\xi\big](\omega) = \dfrac{4}{c_0^2(z)} \displaystyle\int_z (\mathrm{i}\omega)^2 \Omega(\omega) \mathrm{G}_0^2(s,z,\omega) \xi(z)\,\mathrm{d}z, & (5.7a) \\ \big[\mathrm{F}^{\mathrm{T}}\mathrm{P}\big](z) = \dfrac{4}{c_0^2(z)} \displaystyle\int_\omega (\mathrm{i}\omega)^2 \Omega^*(\omega) \mathrm{G}_0^{*2}(s,z,\omega) \mathrm{P}(\omega)\,\mathrm{d}\omega, & (5.7b) \\ \big[\mathrm{F}^{\dagger}\mathrm{P}\big](z) = -4\dfrac{\partial}{\partial z} \displaystyle\int_\omega \frac{1}{(\mathrm{i}\omega)} \widetilde{\Omega}^*(\omega) \Big[\frac{\partial}{\partial z_s} \mathrm{G}_0^*(s,z,\omega)\Big]^2 \mathrm{P}(\omega)\,\mathrm{d}\omega, & (5.7c) \\ \big[(\mathrm{F}^{\dagger})^{\mathrm{T}}\xi\big](\omega) = 4 \displaystyle\int_z \frac{1}{(\mathrm{i}\omega)^*} \widetilde{\Omega}(\omega) \Big[\frac{\partial}{\partial z_s} \mathrm{G}_0(s,z,\omega)\Big]^2 \frac{\partial}{\partial z} \xi(z)\,\mathrm{d}z. & (5.7d) \end{cases}$$

In essence, the adjoint and inverse expressions are very similar, except that $\Omega(\omega)$ is replaced by $\widetilde{\Omega}(\omega)$ with a first-order integration in time, and that two vertical derivatives are applied at the image point z and at the surface. The same observation holds for the forward modelling and the adjoint of the inverse operator.

In a homogeneous velocity model c_0, the 1D Green's function reads

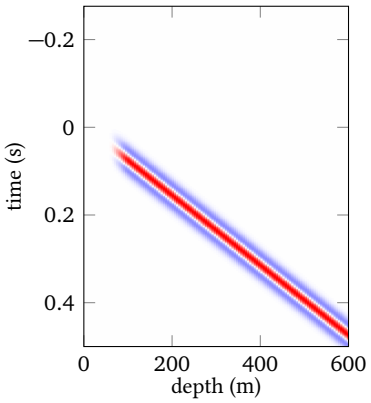
$$\mathrm{G}_0(s,z,\omega) = \frac{c_0}{2\mathrm{i}\omega} \mathrm{e}^{-\mathrm{i}\omega\tau(s,z)}, \tag{5.8}$$

where $\tau(s,z)$ is the traveltime between the source s and the subsurface point z. The geometrical spreading is constant in 1D. Then we can easily derive explicit expressions for the four linear

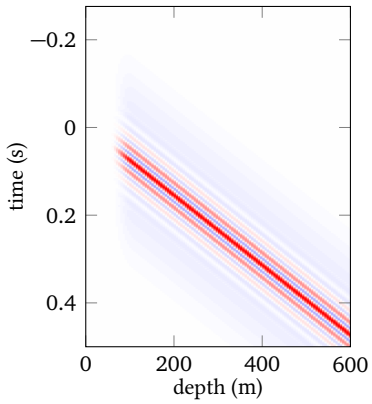
operators

$$\begin{cases} \big[\mathrm{F}\xi\big](\omega) = \displaystyle\int_z \Omega(\omega)\mathrm{e}^{-2\mathrm{i}\omega\tau(s,z)}\xi(z)\,\mathrm{d}z, & (5.9a) \\ \big[\mathrm{F}^{\mathrm{T}}\mathrm{P}\big](z) = \displaystyle\int_\omega \Omega^*(\omega)\mathrm{e}^{2\mathrm{i}\omega\tau(s,z)}\mathrm{P}(\omega)\,\mathrm{d}\omega, & (5.9b) \\ \big[\mathrm{F}^{\dagger}\mathrm{P}\big](z) = \dfrac{2}{c_0}\displaystyle\int_\omega \widetilde{\Omega}^*(\omega)\mathrm{e}^{2\mathrm{i}\omega\tau(s,z)}\mathrm{P}(\omega)\,\mathrm{d}\omega, & (5.9c) \\ \big[(\mathrm{F}^{\dagger})^{\mathrm{T}}\xi\big](\omega) = \dfrac{2}{c_0}\displaystyle\int_z \widetilde{\Omega}(\omega)\mathrm{e}^{-2\mathrm{i}\omega\tau(s,z)}\xi(z)\,\mathrm{d}z. & (5.9d) \end{cases}$$

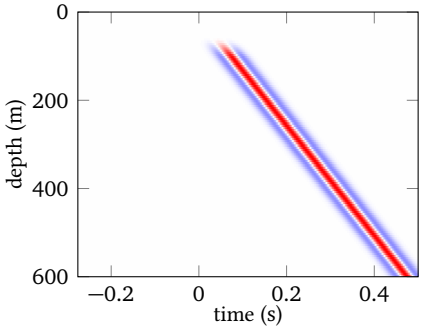
Compared to the adjoint F^{T}, the approximate inverse $\mathrm{F}^{\dagger}$ introduces two modifications: deconvolution of the source wavelet ($\widetilde{\Omega}$ instead of Ω) and correction of the amplitudes by a coefficient $2/c_0$. In 1D there is no need to correct for geometrical spreading nor for uneven illumination.



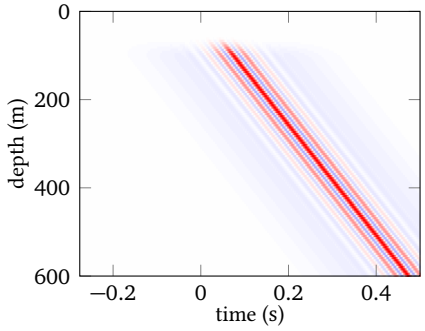
(a) Linear modelling $\mathrm{F} : \mathscr{E} \mapsto \mathscr{D}_{\text{obs}}$.



(b) Adjoint of inverse operator $(\mathrm{F}^{\dagger})^{\mathrm{T}} : \mathscr{E} \mapsto \mathscr{D}_{\text{obs}}$.



(c) Adjoint operator $\mathrm{F}^{\mathrm{T}} : \mathscr{D}_{\text{obs}} \mapsto \mathscr{E}$.



(d) Approximate inverse operator $\mathrm{F}^{\dagger} : \mathscr{D}_{\text{obs}} \mapsto \mathscr{E}$.

Figure 5.3. – Matrix representation of the four linear operators defined in equations (5.9) in the initial velocity model $c_0 = 2500\,\mathrm{m/s}$. The matrices have zero values for $z \leq 100\,\mathrm{m}$ because of the taper introduced in the $\mathscr{E}$-space (figure 5.2).

5.2.2. Direct inversion

Before analysing the iterative case, we briefly review the results obtained with direct inversion. Here the reflectivity is defined as the result of the application of the adjoint F^T or inverse operator $F^\dagger$ to observed data P^{obs}:

$$\begin{cases} \xi_{mig} = F^T P^{obs}, & (5.10a) \\ \xi_{inv} = F^\dagger P^{obs}. & (5.10b) \end{cases}$$

We can control the accuracy of the approximate inverse $F^\dagger$ by applying the modelling operator F to ξ_{inv} (figure 5.4). Recomputed data perfectly match observed data, which is not the case with the adjoint ξ_{mig}. This indicates that $F^\dagger$ is an accurate right inverse of F ($FF^\dagger \simeq I$). As already mentioned in chapter 4, $F^\dagger$ is however not a left inverse of F ($F^\dagger F \neq I$) because the null space of the modelling operator F is not empty. It means that several different reflectivity models can lead to similar data set. For example, applying $F^\dagger F$ to a spike reflectivity (as the red curve in Figure 5.4, left) would yield a smooth reflectivity model. The non triviality of the null space of F is important for the behaviour of the iterative case studied in sections 5.2.3 and 5.2.4.

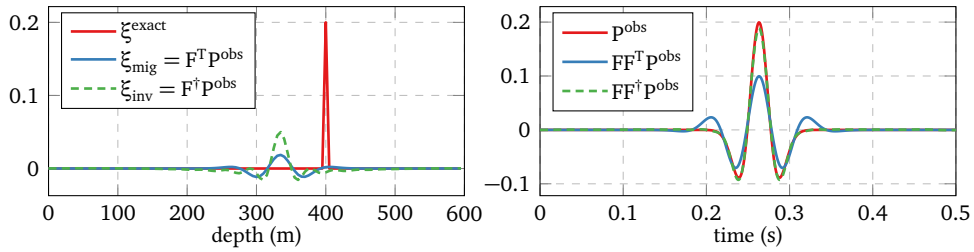


Figure 5.4. – Exact reflectivity model and result of the application of the adjoint F^T and $F^\dagger$ to observed data (left). Observed data and data modelled from the adjoint and inverse reflectivity section (right).

We now derive analytic expressions for the gradient of the MVA objective function obtained when the reflectivity is defined as $F^T P^{obs}$ or $F^\dagger P^{obs}$. The objective is to reproduce the results obtained with the procedure described in section 4.2.11 where wave-equations were solved with a finite difference propagation code. To compute the background velocity update, we construct the following Lagrangians with the adjoint variables η_{mig} and η_{inv} in $\mathscr{E}$.

$$\begin{cases} \overline{J_1}[c_0] = \frac{1}{2}\left\|AC\xi_{mig}\right\|^2_{\mathscr{E}} - \left\langle \eta_{mig} \,\middle|\, \xi_{mig} - F^T P^{obs} \right\rangle_{\mathscr{E}}, & (5.11a) \\ \overline{J_1}[c_0] = \frac{1}{2}\left\|AC\xi_{inv}\right\|^2_{\mathscr{E}} - \left\langle \eta_{inv} \,\middle|\, \xi_{inv} - F^\dagger P^{obs} \right\rangle_{\mathscr{E}}. & (5.11b) \end{cases}$$

The adjoint variables $\eta_{mig/inv}$ are obtained by deriving the Lagrangians with respect to the state variables $\xi_{mig/inv}$,

$$\begin{cases} \eta_{mig} = C^T A^T AC\xi_{mig}, & (5.12a) \\ \eta_{inv} = C^T A^T AC\xi_{inv}, & (5.12b) \end{cases}$$

and finally the gradient of J_1 with respect to c_0 is given by

$$\begin{cases} G_{mig}(y) = (\beta-2)c_0^{2\beta-1}(y)a^2(y)\xi^2_{mig}(y) + \displaystyle\int_z Q_{mig}(y,z)\eta_{mig}(z)\,dz, & (5.13a) \\ G_{inv}(y) = \beta c_0^{2\beta-1}(y)a^2(y)\xi^2_{inv}(y) + \displaystyle\int_z Q_{inv}(y,z)\eta_{inv}(z)\,dz, & (5.13b) \end{cases}$$

where the matrices $Q_{mig} : \mathscr{E} \mapsto \mathscr{M}$ and $Q_{inv} : \mathscr{E} \mapsto \mathscr{M}$ are defined as

$$Q_{mig}(y,z) = \frac{\partial\left[F^T P^{obs}\right](z)}{\partial c_0(y)}, \tag{5.14a}$$

$$Q_{inv}(y,z) = \frac{\partial\left[F^\dagger P^{obs}\right](z)}{\partial c_0(y)}. \tag{5.14b}$$

The term $\beta c_0^{2\beta-1}a^2\xi^2$ in G_{mig} and G_{inv} originates from operator C. Note that there is an additional term $-2c_0^{2\beta-1}a^2\xi^2$ in G_{mig} due to the coefficient $4/c_0^2$ in the definition of the adjoint operator (equation 5.7b).

After derivations reported in appendix D, we obtain the following expressions for Q_{mig} and Q_{inv},

$$Q_{mig}(y,z) = \frac{16}{c_0^3(y)c_0^2(z)}\int_\omega (i\omega)^4\Omega^*(\omega)G_0^*(s,z,\omega)G_0^*(s,y,\omega)G_0^*(y,z,\omega)\,d\omega \tag{5.15a}$$

$$= -\frac{2}{c_0^2}\int_\omega (i\omega)\Omega^*(\omega)e^{(i\omega)\left(\tau(s,z)+\tau(s,y)+\tau(y,z)\right)}P^{obs}(\omega)\,d\omega, \tag{5.15b}$$

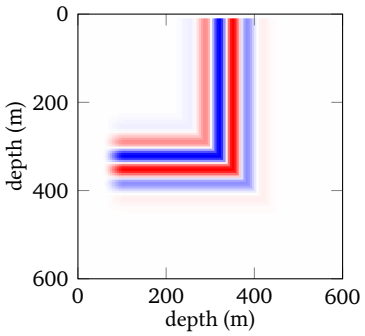
$$Q_{inv}(y,z) = \begin{cases} \dfrac{32}{c_0^2(s)c_0^3(y)c_0(z)}\displaystyle\int_\omega (i\omega)^4\widetilde{\Omega}^*(\omega)G_0^*(s,z)G_0^*(s,y)G_0^*(y,z)\,d\omega & \text{if } y < z, \\ 0 & \text{if } y \geq z, \end{cases} \tag{5.15c}$$

$$= \begin{cases} -\dfrac{4}{c_0^3}\displaystyle\int_\omega (i\omega)\widetilde{\Omega}^*(\omega)e^{(i\omega)\left(\tau(s,z)+\tau(s,y)+\tau(y,z)\right)}P^{obs}(\omega)\,d\omega & \text{if } y < z, \\ 0 & \text{if } y \geq z, \end{cases} \tag{5.15d}$$

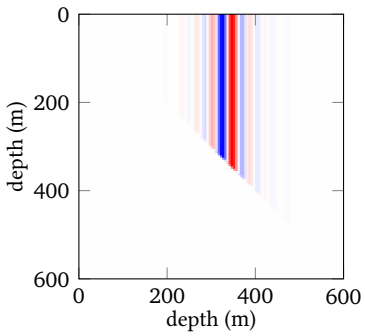
where we have omitted the dependence of the Green's function to ω in equation (5.15c).

In the same way as F^T and $F^\dagger$, Q_{mig} and Q_{inv} differ by a coefficient $2/c_0$ and by the source wavelet (Ω and $\widetilde{\Omega}$, respectively). However an additional difference is that Q_{inv} is upper triangular, contrary to Q_{mig} (figure 5.5). This may be surprising as the high-frequency expressions for the adjoint (5.9b) and pseudo-inverse (5.9c) are very similar. Actually the derivation of Q_{mig} and Q_{inv} starts from equations (5.7b) and (5.7c), which do not use asymptotic approximations. As explained in appendix D, the zero entries below the diagonal of Q_{inv} are due to the derivative with respect to z in equation (5.7c). *This illustrates that the high-frequency approximation of the gradient is not the gradient of the high-frequency approximation.* The high-frequency approximation of the gradients G_{mig} and G_{inv} (equations (5.13)) derived here is different from the gradient

that would be computed from the high-frequency approximation of the reflectivity images (equations 5.9b and 5.9c): in the first case, we replace the Green's function by their expression after having deriving the gradient expression. In the second case, we would use the adjoint-state technique with equations (5.9) as state equations instead of equations (5.7). As equations (5.9b) and (5.9c) are very similar, this approach would lead to similar expressions for operators Q_{mig} and Q_{inv}, without the distinction between the case $y < z$ and $z > y$ (equation 5.15c), and Q_{inv} (figure 5.5b) would not be upper triangular any more.



(a) Operator $Q_{mig} : \mathscr{E} \mapsto \mathscr{M}$.



(b) Operator $Q_{inv} : \mathscr{E} \mapsto \mathscr{M}$.

Figure 5.5. – Matrix representation of the operators Q_{mig} and Q_{inv} defined in equations (5.14a) and (5.14b).

Compared to the adjoint case, the fact that Q_{inv} is zero below the main diagonal changes the shape of the gradient around the reflector depth. In section 4.2, it has been shown that a specific power[2] of β allows to obtain a smooth gradient in the case of the inverse (figure 5.6). Numerical tests (not shown here) indicate that this modification of the MVA objective function does not allow to remove oscillations in the case of the adjoint.

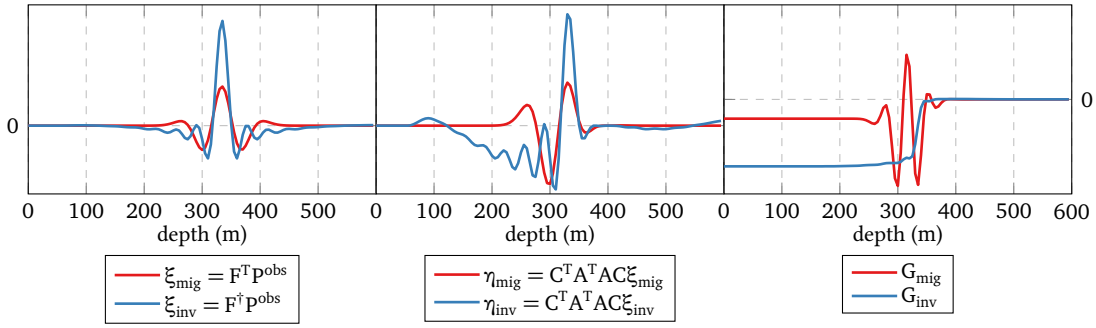


Figure 5.6. – Computation of the gradient of J_1 with respect to c_0 (right panel), comparing the case where the final reflectivity is taken equal to $\xi_{mig} = F^T P^{obs}$ and $\xi_{inv} = F^\dagger P^{obs}$. For this example, we took $\beta = -1/2$.

[2]The optimal value is $\beta = -1/2$ if ξ is defined as a velocity perturbation $\xi = \delta c/(2c_0)$ and $\beta = 3/2$ if ξ is defined as a squared slowness perturbation $\xi = 2\delta c/c_0^3$.

We have derived expression for the MVA gradients for the adjoint and pseudo-inverse operators. In the following section, we analyse the gradient in the iterative case.

5.2.3. Iterative inversion: expression of the gradient

Instead of using the approximate inverse $\mathrm{F}^{\dagger}$, we determine the optimal reflectivity section by minimising iteratively the objective function (5.6). In the case of linear modelling (primaries only) and ℓ_2 regularisation, the minimisation of this objective function is equivalent to the resolution of the linear system

$$\left(\mathrm{F}^{\mathrm{T}}\mathrm{F} + a_{\varphi}\mathrm{I}\right)\xi = \mathrm{F}^{\mathrm{T}}\mathrm{P}^{\mathrm{obs}}, \tag{5.16}$$

which involves the normal operator $\mathrm{F}^{\mathrm{T}}\mathrm{F} : \mathscr{E} \mapsto \mathscr{E}$.

Note that iterative migration is an expensive technique for the case of primaries only as the approximate inverse already provides a good minimiser of J_0. However iterative migration may deal with more complex cases like multiples, for which an inverse formula does not necessarily exist. Here we want to study the shape of the gradient of J_1 after iterative migration and restrict to this linear case for simplicity.

Computation of the gradient of J_1

Following method B defined in chapter 3, we assume that the iterative migration process has reached convergence, meaning that the final reflectivity ξ^{∞} is defined as

$$\left(\mathrm{F}^{\mathrm{T}}\mathrm{F} + a_{\varphi}\mathrm{I}\right)\xi^{\infty} = \mathrm{F}^{\mathrm{T}}\mathrm{P}^{\mathrm{obs}}. \tag{5.17}$$

To compute the gradient of J_1, we define the following Lagrangian with equation (5.17) as constraint and the adjoint state $\eta \in \mathscr{E}$

$$\overline{\mathrm{J}_1}[c_0, \xi^{\infty}, \eta] = \frac{1}{2}\left\|\mathrm{AC}\xi^{\infty}\right\|_{\mathscr{E}}^2 - \left\langle \eta \,\middle|\, \left(\mathrm{F}^{\mathrm{T}}\mathrm{F} + a_{\varphi}\mathrm{I}\right)\xi^{\infty} - \mathrm{F}^{\mathrm{T}}\mathrm{P}^{\mathrm{obs}} \right\rangle_{\mathscr{E}}. \tag{5.18}$$

Deriving this expression with respect to ξ^{∞} leads to the following adjoint equation,

$$\left(\mathrm{F}^{\mathrm{T}}\mathrm{F} + a_{\varphi}\mathrm{I}\right)\eta = \mathrm{C}^{\mathrm{T}}\mathrm{A}^{\mathrm{T}}\mathrm{AC}\xi^{\infty}. \tag{5.19}$$

ξ^{∞} and η are solutions of the same linear problem for two different source terms. However the resolution of the adjoint problem is more difficult and requires more iterations as observed in section 4.4.3. Our interpretation of this different behaviour is that the source term $\mathrm{F}^{\mathrm{T}}\mathrm{P}^{\mathrm{obs}}$ is in the image of the normal operator $\mathrm{F}^{\mathrm{T}}\mathrm{F}$, whereas there is no guarantee that the source term $\mathrm{C}^{\mathrm{T}}\mathrm{A}^{\mathrm{T}}\mathrm{AC}\xi^{\infty}$ is in the image of $\mathrm{F}^{\mathrm{T}}\mathrm{F}$, meaning that a solution to equation (5.19) is less likely to be found, or at least more difficult. This point will be further illustrated in sections 5.2.4 and 5.2.7.

Assuming equation (5.19) is solved perfectly, the ideal gradient is made of three contributions

$$\mathrm{G}^{\infty}(x) = \frac{\beta}{c_0}\left[\mathrm{AC}\xi^{\infty}\right]^2(x) + \frac{\partial}{\partial c_0(x)}\left[\left\langle \eta^{\infty} \,\middle|\, \mathrm{F}^{\mathrm{T}}\mathrm{P}^{\mathrm{obs}} \right\rangle_{\mathscr{E}}\right] + \frac{\partial}{\partial c_0(x)}\left[\left\langle \eta^{\infty} \,\middle|\, \mathrm{F}^{\mathrm{T}}\mathrm{F}\xi^{\infty} \right\rangle_{\mathscr{E}}\right]. \tag{5.20}$$

The first term is due the power of c_0 that has been introduced in the definition of J_1. The second is due to the inner product $\langle \eta \,|\, F^T P^{obs} \rangle$ and has already been analysed in section 5.2.2. Finally we compute the third contribution due to the normal operator,

$$\frac{\partial}{\partial c_0(x)}\Big[\big\langle \eta \,\big|\, F^T F \xi \big\rangle_{\mathscr{E}}\Big] = \frac{\partial}{\partial c_0(x)} \int_z \int_\omega \int_y \eta(z) F^T(z,\omega) F(\omega,y) \xi(y) \,dy \,d\omega \,dz \tag{5.21a}$$

$$= \int_z \int_\omega \int_y \left[\eta(z) \frac{\partial F^T(z,\omega)}{\partial c_0(x)} F(\omega,y)\xi(y) + \eta(z) F^T(z,\omega) \frac{\partial F(\omega,y)}{\partial c_0(x)} \xi(y) \right] dy \,d\omega \,dz \tag{5.21b}$$

$$= \int_z \int_\omega \int_y \left[\eta(z) \frac{\partial F^T(z,\omega)}{\partial c_0(x)} F(\omega,y)\xi(y) + \xi(y) \frac{\partial F^T(\omega,y)}{\partial c_0(x)} F(\omega,z)\eta(z) \right] dy \,d\omega \,dz \tag{5.21c}$$

$$= \int_z Q_1(x,z)\eta(z)\,dz + \int_y Q_2(x,y)\xi(y)\,dy - \frac{2}{c_0}\big[F^T F \xi\big](x)\eta(x) - \frac{2}{c_0}\big[F^T F \eta\big](x)\xi(x), \tag{5.21d}$$

with Q_1 and Q_2 two matrices similar to Q_{mig} (equation 5.14a):

$$Q_1(x,z) = \frac{\partial \big[F^T P_1\big](z)}{\partial c_0(x)} \quad \text{and} \quad Q_2(x,z) = \frac{\partial \big[F^T P_2\big](z)}{\partial c_0(x)}, \tag{5.22}$$

with calculated data $P_1 = F\xi$ and $P_2 = F\eta$.

In practice, the direct and adjoint problems are solved in a finite number of iterations (N and M iterations, respectively), leading to approximate solutions $\xi^{(N+1)}$ and $\eta^{(N+1,M+1)}$ and an approximate gradient $G^{(N+1,M+1)}$ made of seven contributions,

$$\begin{cases}
G_1^{(N,M)}(x) = -\displaystyle\int_y Q_1^{(N+1)}(x,y)\eta^{(N+1,M+1)}(y)\,dy & \text{(5.23a)} \\
G_2^{(N,M)}(x) = -\displaystyle\int_z Q_2^{(N+1,M+1)}(x,z)\xi^{(N+1)}(z)\,dz & \text{(5.23b)} \\
G_3^{(N,M)}(x) = \displaystyle\int_y Q_{mig}(x,y)\eta^{(N+1,M+1)}(y)\,dy & \text{(5.23c)} \\
G_4^{(N,M)}(x) = -\dfrac{2}{c_0}\big[F^T P^{obs}\big](x)\eta^{(N+1,M+1)}(x) & \text{(5.23d)} \\
G_5^{(N,M)}(x) = \dfrac{2}{c_0}\big[F^T F \xi^{(N+1)}\big](x)\eta^{(N+1,M+1)}(x) & \text{(5.23e)} \\
G_6^{(N,M)}(x) = \dfrac{2}{c_0}\xi^{(N+1)}(x)\big[F^T F \eta^{(N+1,M+1)}\big](x) & \text{(5.23f)} \\
G_\beta^{(N)}(x) = \beta c_0^{2\beta-1}(x) a^2(x) (\xi^{(N+1)})^2(x). & \text{(5.23g)}
\end{cases}$$

If the direct (equation 5.17) and adjoint (equation 5.19) system are correctly solved, meaning

that final residuals are small, than for a small regularisation,

$$\begin{cases} P^{(N+1)} \simeq P^{\text{obs}}, & (5.24a) \\ Q_1^{(N+1)} \simeq Q_{\text{mig}}, & (5.24b) \\ F^T F \xi^{(N+1)} \simeq F^T P^{\text{obs}}, & (5.24c) \\ F^T F \eta^{(N+1,M+1)} \simeq C^T A^T A C \xi^{(N+1)}, & (5.24d) \end{cases}$$

and most of the contributions in equation (5.23) cancel out,

$$\begin{cases} G_1 + G_3 \simeq 0, & (5.25a) \\ G_4 + G_5 \simeq 0, & (5.25b) \\ G_6 + G_\beta \simeq 0 \quad \text{if } \beta = -2. & (5.25c) \end{cases}$$

The only remaining contribution is G_2. We expect that it is similar to the gradient obtained with direct inversion. We shall see on a numerical example that this is not the case for reasons explained after.

Preconditioning

To accelerate the resolution of the direct (equation 5.17) and adjoint (equation 5.19) problem, we use a preconditioner based on the approximate inverse $F^\dagger$. We briefly review how the modified preconditioned problem is defined.

The original linear problem can be written in a compact way as

$$H\xi = b, \tag{5.26}$$

with the Hessian $H = F^T F + a_\varphi I$ and the vector $b \in \mathscr{E}$ defined as $b = F^T P^{\text{obs}}$ for the direct problem and $b = C^T A^T A C \xi^\infty$ for the adjoint problem. This problem is solved with the linear conjugate gradient algorithm, which actually minimises the scalar objective function

$$\psi(\xi) = \frac{1}{2} \xi^T H \xi - \xi^T b. \tag{5.27}$$

We introduce the preconditioner $F^\dagger$ in this system with the following change of variables

$$\xi = F^\dagger \widehat{\xi} \tag{5.28}$$

with the new variable $\widehat{\xi} \in \mathscr{D}_{\text{obs}}$. Replacing in equation (5.27) leads to

$$\widehat{\psi}\left(\widehat{\xi}\right) = \frac{1}{2} \widehat{\xi}^T (F^\dagger)^T H F^\dagger \widehat{\xi} - \widehat{\xi}^T (F^\dagger)^T b, \tag{5.29}$$

which corresponds to the linear system

$$(F^\dagger)^T H F^\dagger \widehat{\xi} = (F^\dagger)^T b, \tag{5.30}$$

where $(F^\dagger)^T H F^\dagger \simeq (F^\dagger)^T F^T F F^\dagger = (F F^\dagger)^T (F F^\dagger)$ if the regularisation weight a_φ is not too strong. As $F^\dagger$ has been designed to be an approximate inverse of F ($F F^\dagger d \simeq d$), this operator is expected to be close to the identity (note however that $F^\dagger F \xi$ is not necessary close to ξ).

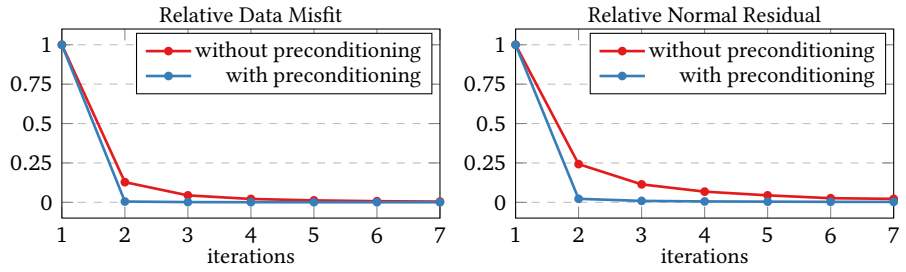
We recall that preconditioning does not change the way the gradient of J_1 is computed. It only provides a more efficient way to compute the solutions $\xi^{(N+1)}$ and $\eta^{(N+1,M+1)}$ of the direct and adjoint problems (equations 5.17 and 5.19).

Before comparing the gradients obtained by the direct (equation 5.23) and the iterative approach (equation 5.13b), we study the convergence rate of the direct and adjoint problems.

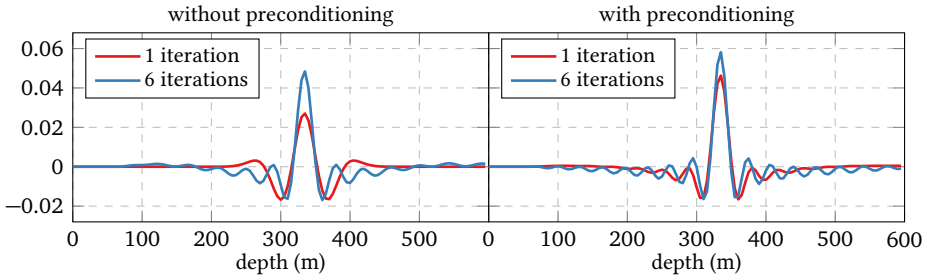
5.2.4. Convergence speed

We now consider the same example as in section 5.2.2 with a single reflector located at 400 m depth and a too low initial velocity. In this first example we solve iteratively both the direct and adjoint problems with a weak regularisation weight a_φ to compare the behaviour of the 1D and 2D cases. In this section, we focus on the convergence speed observed in the resolution of these problems.

We perform six iterations of iterative migration, both without and with preconditioning (figure 5.7). The regularisation function is the ℓ_2-norm and the regularisation weight a_φ is supposed to be small. Similarly to the 2D case, preconditioning significantly accelerates the convergence. After one iteration, the data misfit has been reduced to 0.5 % of its initial value, compared to 13 % for the classical case without preconditioning. Additional iterations only result in further deconvolution of the source wavelet.



(a) Relative data misfit $\|F\xi^{(n)} - P^{obs}\|^2_{\mathscr{D}_{obs}} / \|P^{obs}\|^2_{\mathscr{D}_{obs}}$ (left) and relative normal residual $\|H\xi^{(n)} - b\|_{\mathscr{E}} / \|b\|_{\mathscr{E}}$ (right) across iterations.

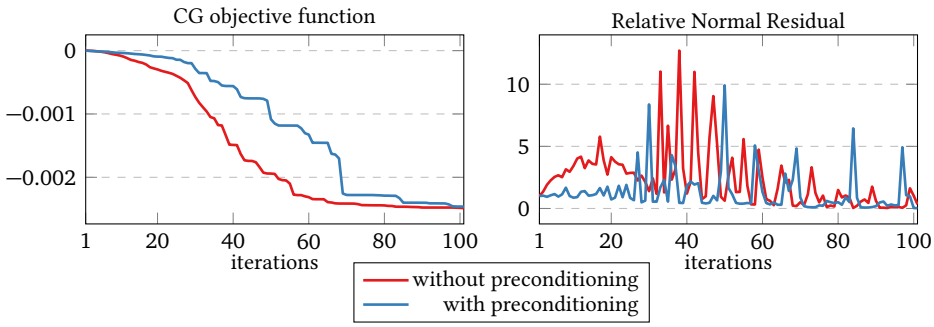


(b) Reflectivity model obtained after one and six iterations, without preconditioning (left), and with preconditioning (right).

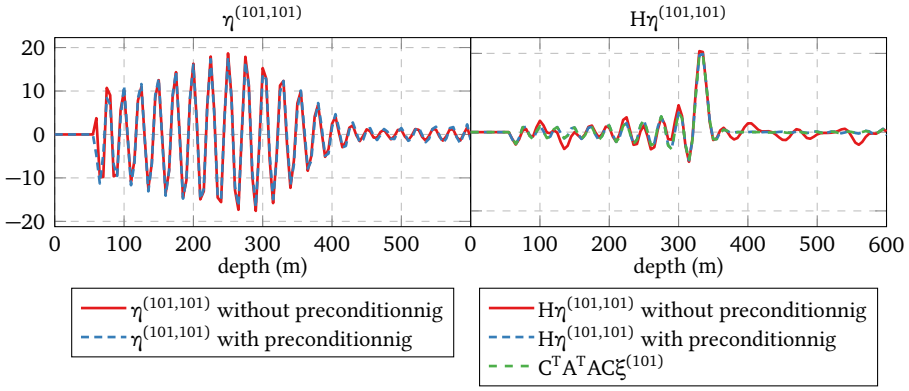
Figure 5.7. – Results of six iterations of iterative migration with and without preconditioning.

The same comparison is performed for the adjoint problem (figure 5.8). We use $N = M = 100$

iterations for both problems to obtain solutions ξ and η as accurate as possible. In this case, preconditioning does not really speed up convergence. The objective function associated with the linear conjugate gradient algorithm (equation 5.27, figure 5.8a, left) decreases slowly in the first iterations, and the normal residual (figure 5.8a, right), which is actually the norm of the gradient of the objective function, is not even monotonous. The final solution $\eta^{(101,101)}$ is very oscillating (figure 5.8b, left), similar to the observations made in the 2D case (figure 4.29). We may have expected to obtain a final value localised around 320 m, corresponding to the main peak in the final reflectivity. However it is a relatively accurate solution to the adjoint problem as $\mathrm{H}\eta^{(101,101)}$ is close to the right-hand side term.



(a) Norm minimised by the conjugate gradient algorithm (left, (equation 5.27) and relative normal residual (right) across iterations. The fact that preconditioning does not accelerate convergence (left) is not satisfactory. The approach proposed in section 5.2.7 will provide a more favourable behaviour.



(b) Adjoint variables obtained after 100 iterations (left). Application of the Hessian matrix H to this result and comparison with the right-hand side term of the adjoint problem (right).

Figure 5.8. – Results of 100 iterations for the adjoint problem. The direct problem has been solved with 100 iterations too. A small ℓ_2-regularisation is applied.

To understand the different behaviour of the direct and adjoint problems, we compute numerically the SVD decomposition of the matrices $\mathrm{F}^{\mathrm{T}}\mathrm{F}$ and $(\mathrm{F}^{\dagger})^{\mathrm{T}}\mathrm{F}^{\mathrm{T}}\mathrm{F}\mathrm{F}^{\dagger}$ (figures 5.9a and 5.9b). To fairly compare the properties of the direct and adjoint problems, we consider normalised versions

of their right-hand side terms (figure 5.9c) and compute their projection on the eigenvectors of the matrices considered in the direct and adjoint problems (figure 5.9d).

The convergence of the conjugate-gradient algorithm accelerates greatly when the eigenvalues are clustered (Nocedal and Wright, 2006, p. 118). In our case, the eigenvalues of the preconditioned system are more concentrated around 1 than the original one. This illustrates that $\mathbf{F}\mathbf{F}^\dagger$ and $(\mathbf{F}^\dagger)^\mathrm{T}\mathbf{F}^\mathrm{T}$ are close to identity operators and is consistent with the observed improved convergence for the direct problem. Note that many eigenvalues are zero due to oversampling of the depth and time axes.

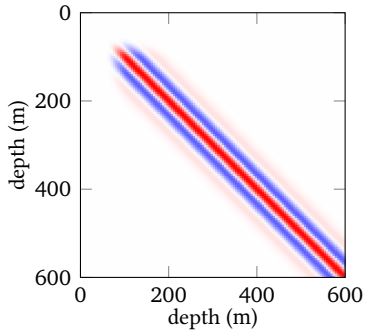
The different behaviour observed for the adjoint problem comes from the right-hand side term. For the direct problem in the preconditioned case, most of the energy of the projection is concentrated on eigenvectors associated with an eigenvalue close to 1, hence a single iteration provides a good solution to the direct problem. In the case of the adjoint problem, the projection of the right-hand side term spreads over a wider range of eigenvectors than for the direct problem, even with preconditioning. More importantly, non-negligible contributions come from eigenvectors associated with small eigenvalues, which explains that more iterations are required to solve the adjoint problem. These small eigenvalues are not all due to the oversampling of the time and depth axis, and using a coarser grid would not solve this issue.

Contrary to the direct problem, the right-hand side term of the adjoint problem cannot be expressed as the application of $\mathbf{F}^\mathrm{T}$ to a vector of $\mathscr{D}_\mathrm{obs}$. In section 5.2.7, we propose a modification of J_1 such that both problem have similar right-hand side terms with a more favourable projection on the eigenvectors of $\mathbf{F}^\mathrm{T}\mathbf{F}$ and $(\mathbf{F}^\dagger)^\mathrm{T}\mathbf{F}^\mathrm{T}\mathbf{F}\mathbf{F}^\dagger$. We first discuss the shape of the MVA gradient and the role of regularisation.

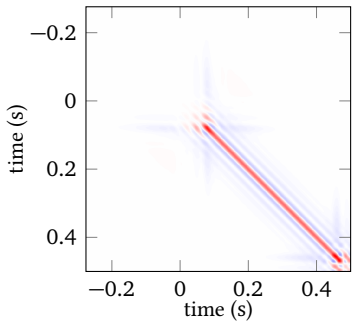
5.2.5. Shape of the gradient

We compute the gradient with the values of ξ and η obtained after resolution of the direct and adjoint problems in 100 iterations each with preconditioning. This ensures that both problems are nearly perfectly solved (figure 5.10). As a consequence, most of the contributions of the gradient cancel with one another as in equations (5.25) (figure 5.11). However the remaining contribution G_2 has a strong negative peak at the reflector position and is strongly oscillating, even below the reflector.

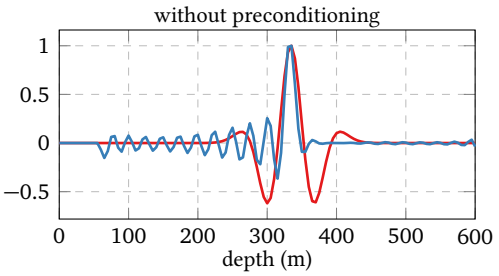
The gradient obtained after iterative inversion is quite different from the one obtained after direct inversion. Contrary to iterative migration, the operator $\mathbf{F}^\dagger$ provides only an asymptotic inverse, as a consequence the final reflectivity $\xi^{(101)}$ and the inverse ξ_inv are not exactly similar, although they both result in very small relative data residuals (0.002 % and 0.25 %, respectively). The iterative result is more oscillating away from the reflector position. These oscillations allow to further reduce the data misfit, but do not carry useful kinematic information, which is contained in the main peak localised at the reflector position. After application of the annihilator, these oscillations are strengthened with respect to the main peak. Small perturbations of the reflectivity model in the null space of F away from the reflector leads to small modifications on the value of J_0, but have a large influence on the gradient of J_1 and may mislead the velocity analysis. This is emphasised in this 1D case with the multiplication by $z - z_\mathrm{ex}$, but the same conclusion holds in 2D. In the following, we investigate two ways of attenuating this undesirable effect, first by regularisation on ξ of the direct problem (section 5.2.6), then by introducing a

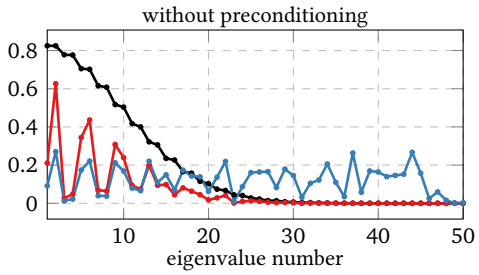


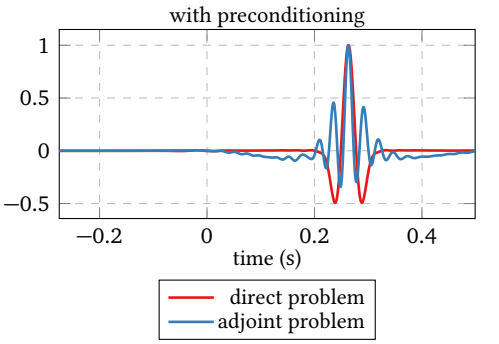
(a) Normal operator $\mathrm{F}^{\mathrm{T}}\mathrm{F} : \mathscr{E} \mapsto \mathscr{E}$ when no preconditioner is applied.



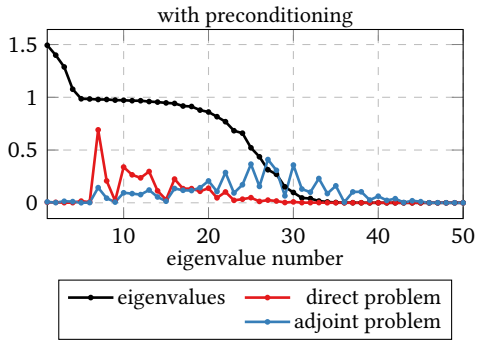
(b) Normal operator $(\mathrm{FF}^{\dagger})^{\mathrm{T}}\mathrm{FF}^{\dagger} : \mathscr{D}_{\mathrm{obs}} \mapsto \mathscr{D}_{\mathrm{obs}}$ when $\mathrm{F}^{\dagger}$ is used as a preconditioner.







(c) Normalised right-hand side term of the direct (red) and adjoint (blue) problems.



(d) Eigenvalue distribution (black) and projection (absolute value) of the right-hand side term of the direct (red) and adjoint (blue) problems.

Figure 5.9. – Analysis of the difference of convergence speed between the direct and adjoint problems.

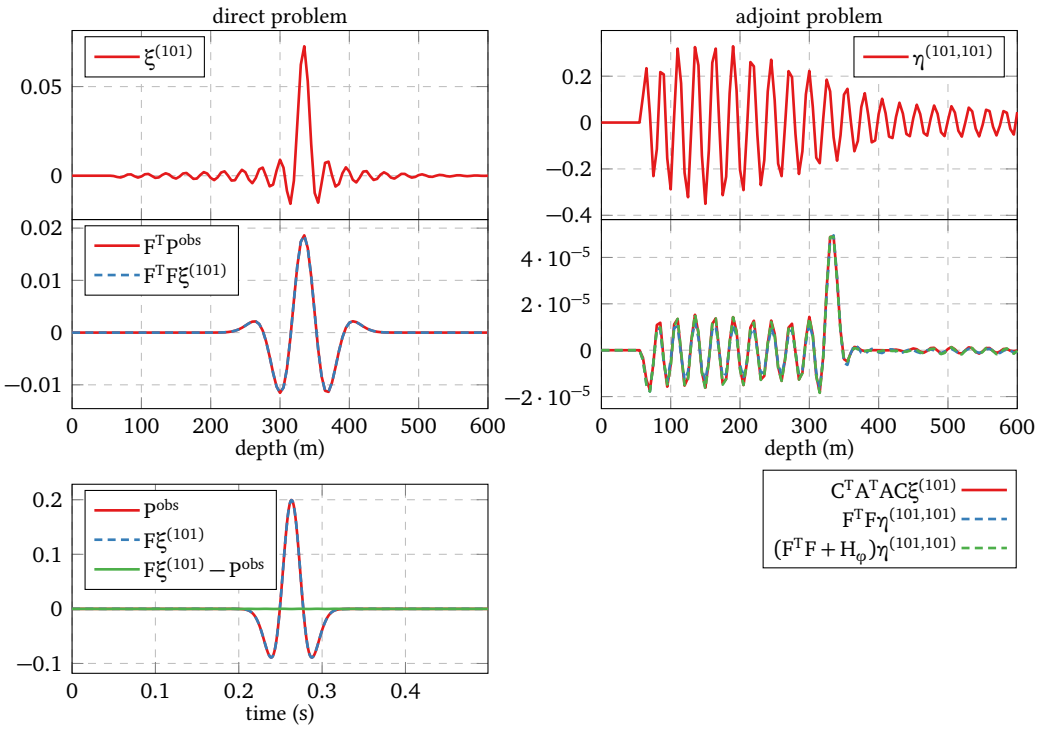


Figure 5.10. – Final solution (top row) of the direct (left) and adjoint (right) problem obtained after 100 iterations. Application of the normal operator to these results and comparison with the right-hand side term (middle row). Comparison of observed data and final calculated data and associated residuals (bottom row).

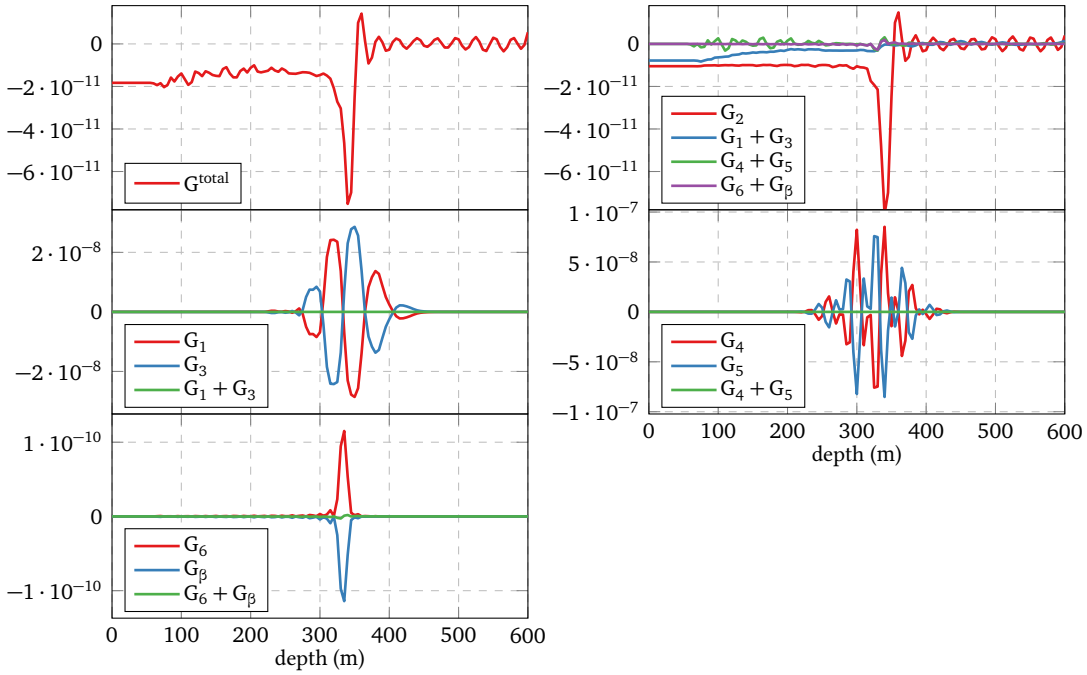


Figure 5.11. – Gradient obtained with $\xi^{(101)}$ and $\eta^{(101,101)}$ (top left) and its decomposition (top right). The three remaining panels illustrate that all contributions but one cancel with one another (equations 5.25).

filter in the definition of the MVA objective function (section 5.2.7).

5.2.6. Influence of regularisation

We consider regularisation of iterative migration as a first remedy to the apparition of oscillations away from the reflector position. In the case of the ℓ_2 norm, we may increase the value of the regularisation weight a_φ. This would yield a reflectivity image with energy more localised, but at the expense of a poorer data misfit. Consequently, the contributions $G_1 + G_3$, $G_4 + G_5$ and $G_6 + G_\beta$ would not be negligible with respect to G_2 any more. Instead we consider using the Huber norm (Guitton and Symes, 2003), an intermediate between ℓ_1 and ℓ_2 norms, for the definition of the regularisation function φ,

$$\varphi[\xi] = \int_z \mathrm{hub}\big(\xi(z)\big)\,\mathrm{d}z, \tag{5.31a}$$

with $\mathrm{hub}(x)$ defined for any real number x as

$$\mathrm{hub}(x) = \begin{cases} \dfrac{x^2}{2s} & \text{if } |x| < s, \\ |x| - \dfrac{s}{2} & \text{if } |x| \geq s. \end{cases} \tag{5.31b}$$

The value for the threshold s needs to be defined. In the following application we choose $s = 0.01$ (dimensionless quantity to be compared to the reflection coefficient defined as $\xi = \delta c / 2c_0$) to distinguish the main peak from the spurious oscillations. Note that the direct problem is not linear any more as the derivative of φ with respect to ξ cannot be written as a matrix-vector product. However the adjoint problem remains linear with a modified regularisation matrix,

$$\big(F^T F + a_\varphi H_\varphi[\xi^\infty]\big)\eta = C^T A^T A C \xi^\infty, \quad \text{with} \quad H_\varphi[\xi] = \frac{\partial^2 \varphi}{\partial \xi^2}[\xi]. \tag{5.32}$$

The matrix H_φ is diagonal, with $1/s$ on the diagonal where ξ^∞ is below the threshold s and zeros elsewhere.

We test the effect of the Huber norm on the same example, with 10 iterations used to solve both the direct and adjoint problems. The final reflectivity image is sparser, but the data misfit remains very low (figure 5.12). The new right-hand side term has a more favourable projection on the eigenvectors of the linear system (figure 5.13), consequently the convergence of the adjoint problem is easier and the preconditioner does bring some acceleration, as it does for the direct problem. The final gradient (figure 5.14) is slightly less oscillating than the one obtained in the previous case but is still affected by a strong artefact at the reflector position. Besides, the determination of a_φ and s is a tedious task. A Pareto analysis (Hansen, 1999) could indicate possible values as a compromise between data misfit and smoothness of the gradient. As this approach is relatively expensive, the value of a_φ and s are in practice determined empirically.

5.2.7. New approach: application of a filter on the final reflectivity

As an alternative to regularisation, we consider applying a filter $K : \mathscr{E} \mapsto \mathscr{E}$ to the result of iterative migration ξ^∞ before measuring defocused energy. The main objective is to remove

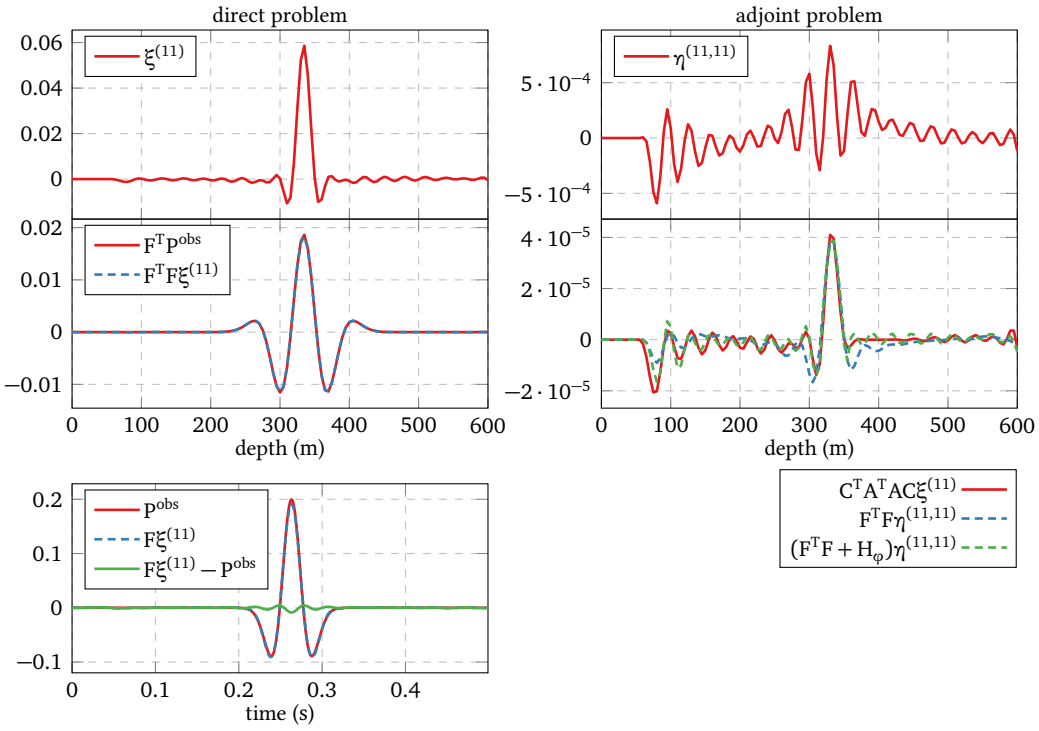


Figure 5.12. – Same as figure 5.10 with the Huber norm as regularisation, and after 10 iterations for both the direct and adjoint problems.

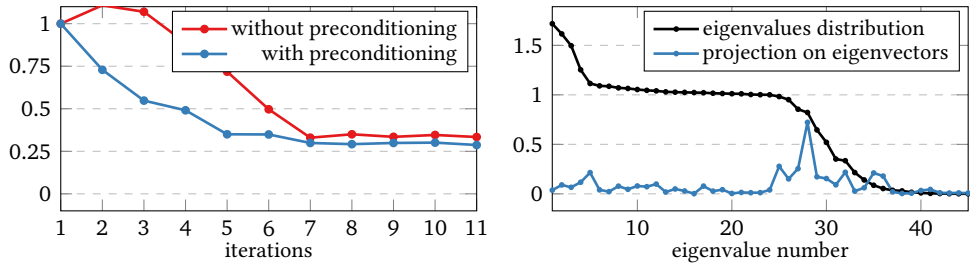


Figure 5.13. – Normal residual across iterations for the adjoint problem (left). Eigenvalue distribution of $(F^\dagger)^T(H + H_\varphi)F^\dagger$ and projection on its eigenvectors of the right-hand side term of the adjoint problem in the preconditioned case.

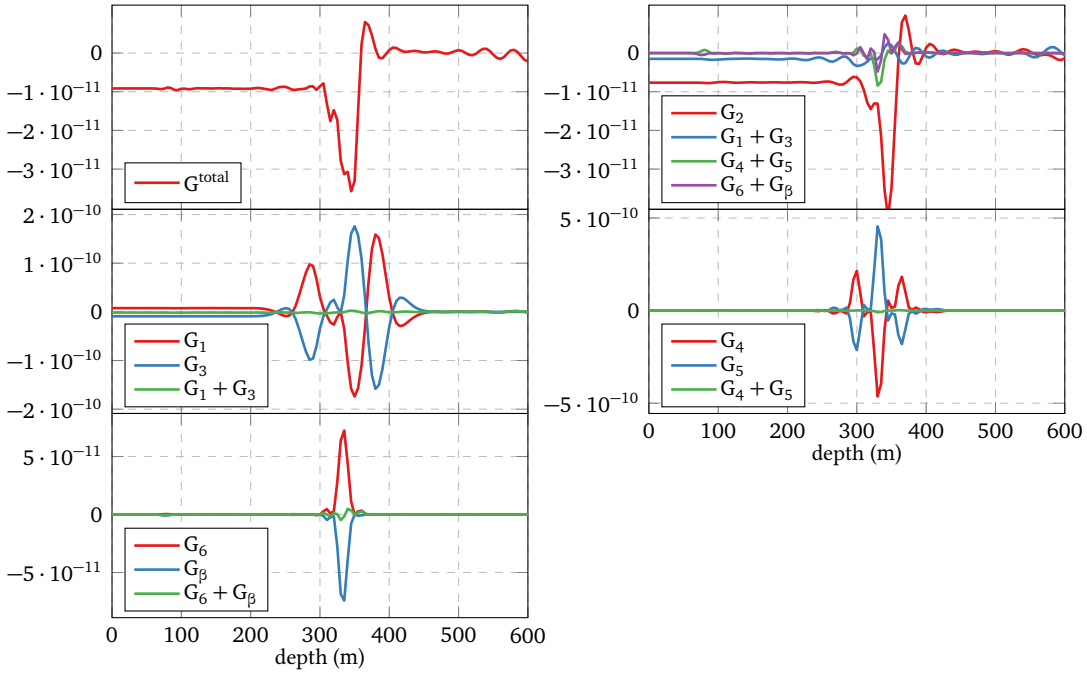


Figure 5.14. – Same as figure 5.11 with the Huber norm as regularisation, and after 10 iterations for both the direct and adjoint problems.

undesired energy in the reflectivity before evaluating the quality of ξ with respect to c_0. This leads to the following objective function

$$J_1[c_0] = \frac{1}{2}\left\|ACK\xi^\infty[c_0]\right\|_{\mathscr{E}}^2. \tag{5.33}$$

In the following we consider two filters K. The second one depend on c_0, leading to additional terms in the velocity update. The gradient of the modified objective function is similar to the regular one. For simplicity, we note $\zeta = K\xi^\infty$ and build the following Lagrangian with the adjoint variables γ and η associated to ζ and ξ^∞, respectively.

$$\begin{aligned}\overline{J_1}[c_0,\xi,\zeta,\gamma,\eta] = \frac{1}{2}\left\|AC\zeta\right\|_{\mathscr{E}}^2 &- \left\langle \gamma \,\middle|\, \zeta - K\xi^\infty \right\rangle_{\mathscr{E}} \\ &- \left\langle \eta \,\middle|\, (F^TF + a_\varphi I)\xi^\infty - F^TP^{obs} \right\rangle_{\mathscr{E}}\end{aligned} \tag{5.34}$$

Deriving this expression with respect to the state variables ζ and ξ^∞ leads to the following adjoint equations,

$$\begin{cases} \gamma = C^TA^TAC\zeta, & \text{(5.35a)} \\ (F^TF + a_\varphi I)\eta = K^T\gamma = K^TC^TA^TACK\xi^\infty. & \text{(5.35b)} \end{cases}$$

The adjoint variable η is solution of a problem similar to equation (5.19), except that K^T is added twice in the right-hand side term. The gradient eventually reads

$$\begin{aligned}\frac{\partial J_1}{\partial c_0(x)} =& \beta c_0^{2\beta-1}\zeta^2(x) - \frac{\partial}{\partial c_0(x)}\left[\left\langle \eta^\infty \,\middle|\, F^T F \xi^\infty \right\rangle_{\mathscr{E}}\right] + \frac{\partial}{\partial c_0(x)}\left[\left\langle \eta^\infty \,\middle|\, F^T P^{obs} \right\rangle_{\mathscr{E}}\right] \\ &+ \frac{\partial}{\partial c_0(x)}\left[\left\langle \gamma \,\middle|\, K \xi^\infty \right\rangle_{\mathscr{E}}\right].\end{aligned} \tag{5.36}$$

The three first terms are similar to the previous case, except that K is now involved in the computation of ζ and η. If K depends on c_0, the fourth contribution is non-zero and should be computed.

Frequency filter independent of c_0

We first consider a filter K which does not depend on c_0, so that the only difference with the usual procedure is the modified right-hand side term of the adjoint problem. The filter is designed to remove the low and high-frequencies components of $\xi^{(N+1)}$ which do not bear relevant kinematic information. In this case, K can be seen as a taper in the wavenumber domain or as a convolution in the depth domain.

We perform 10 iterations without regularisation. The filter is empirically determined such that ζ is close to the result of the adjoint $F^T P^{obs}$ in order to attenuate the spurious oscillations away from the reflector (figure 5.15). The new right-hand side term has a projection on eigenvectors similar to the one obtained for the direct problem, and the convergence of the adjoint problem is almost as fast as for the direct problem (figure 5.16). Ten iterations are sufficient to obtain good solutions to both the direct and adjoint problems (figure 5.17). The final gradient is free of oscillations above and below the reflector, but is still oscillating around the reflector depth (figure 5.18). So this is not yet a fully satisfactory solution. Moreover, in cases more complex than the homogeneous model considered here, the value of K should depend on the spatial positions and on the velocity values. In the next section, we propose a filter that fulfils these requirements.

Using $F^\dagger F$ as a filter on the final reflectivity

We now define a new filter using the approximate inverse $F^\dagger$,

$$K = F^\dagger F. \tag{5.37}$$

It consists first of applying the modelling operator to the result of iterative migration, yielding new data which should be very close to observed data. Then applying the inverse operator should produce a new reflectivity image close to the inverse $F^\dagger P^{obs}$, and the objective function should be close to the one studied in section 5.2.2. Note that $F^\dagger$ has been designed such that $FF^\dagger$ is close to the identity ($F^\dagger F$ is not necessarily identity, though). The filter considered here should be close to the identity as well, but only for a specific range of spatial frequencies. We would like the remaining frequency components to be attenuated. The source term of the adjoint problem now reads $K^T\gamma = F^T(F^\dagger)^T\gamma$ and therefore lives in the image of operator F^T, as does the source term of the direct problem.

With this new definition, K now depends on c_0 and we compute the additional contributions

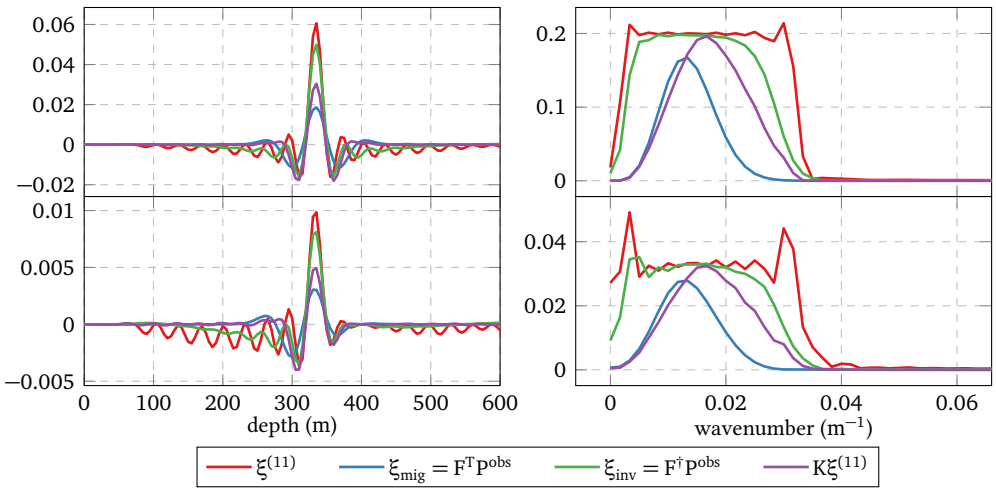


Figure 5.15. – Result (purple) of the application of a filter K to the final reflectivity $\xi^{(11)}$ (red) in the depth (top left) and wavenumber (top right) domain. We display the result of the application of the adjoint (blue) and inverse (green) operator for reference. The bottom row displays the same comparison after application of the annihilator A.

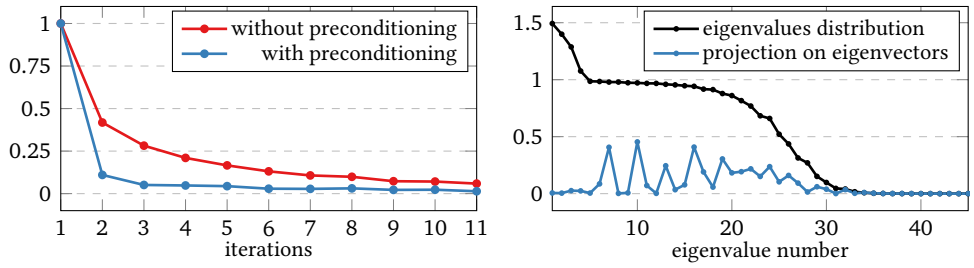


Figure 5.16. – Same as figure 5.13 without regularisation and with the introduction of a filter K in the MVA objective function.

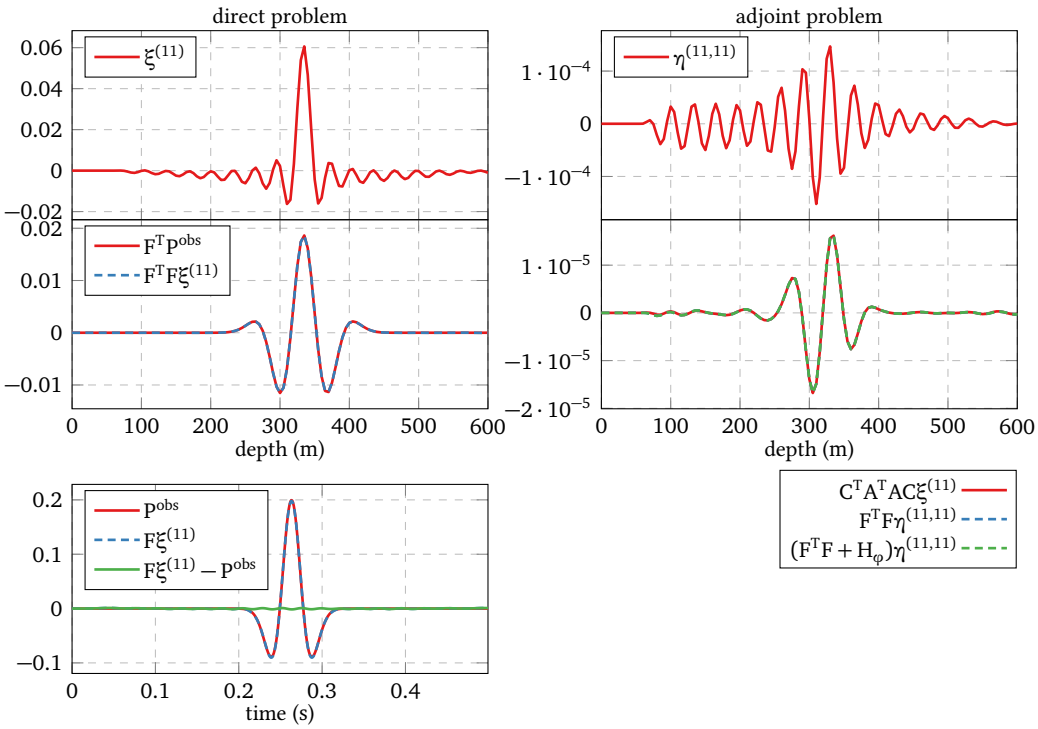


Figure 5.17. – Same as figure 5.12 without regularisation and with the introduction of a filter K in the MVA objective function.

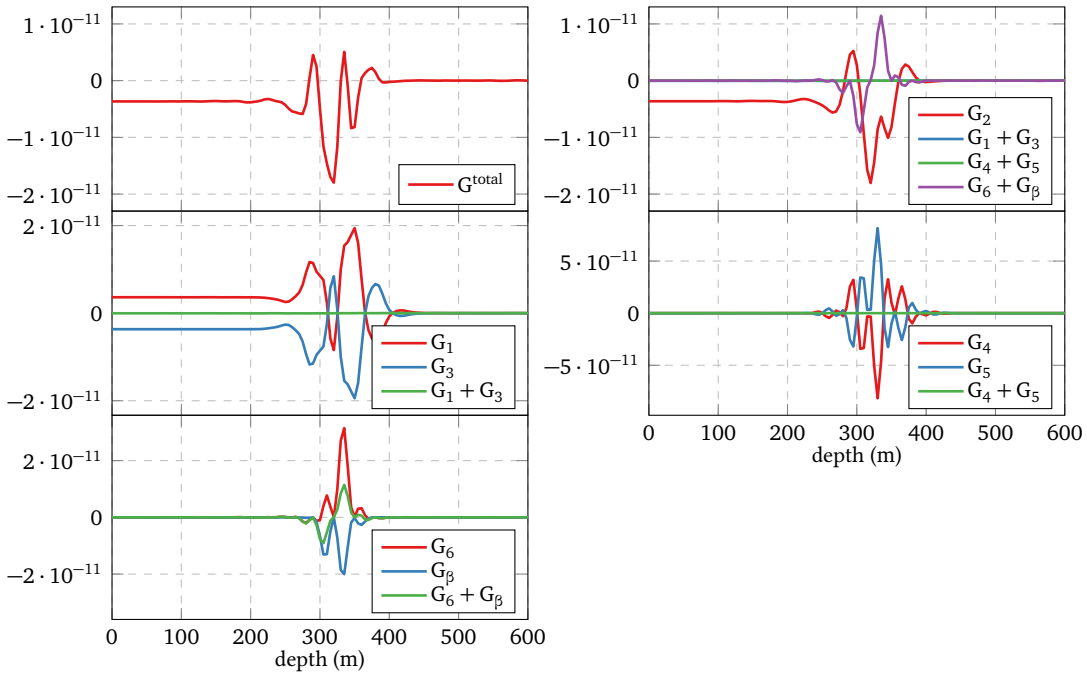


Figure 5.18. – Same as figure 5.14 without regularisation and with the introduction of a filter K in the MVA objective function.

to the gradient (equation 5.36), noted G_7, G_8 and G_9,

$$\frac{\partial}{\partial c_0(x)}\Big[\big\langle \gamma \,\big|\, F^\dagger F\xi \big\rangle_{\mathcal{E}}\Big] \tag{5.38a}$$

$$= \int_z \int_\omega \int_y \left[\gamma(z)\frac{\partial F^\dagger(z,\omega)}{\partial c_0(x)} F(\omega,y)\xi(y) + \gamma(z)F^\dagger(z,\omega)\frac{\partial F(\omega,y)}{\partial c_0(x)}\xi(y)\right] dy\, d\omega\, dz \tag{5.38b}$$

$$= \int_z \int_\omega \int_y \left[\gamma(z)\frac{\partial F^\dagger(z,\omega)}{\partial c_0(x)} F(\omega,y)\xi(y) + \xi(y)\frac{\partial F^{\mathrm{T}}(y,\omega)}{\partial c_0(x)} (F^\dagger)^{\mathrm{T}}(\omega,z)\gamma(z)\right] dy\, d\omega\, dz \tag{5.38c}$$

$$= \underbrace{\int_z Q_7(x,z)\gamma(z)}_{G_7(x)} + \underbrace{\int_y Q_8(x,y)\xi(y) - \frac{2}{c_0}\gamma(x)\big[F^{\mathrm{T}}(F^\dagger)^{\mathrm{T}}\xi\big](x)}_{G_8(x)\qquad\qquad G_9(x)}, \tag{5.38d}$$

where Q_7 and Q_8 are matrices similar to Q_{inv} and Q_{mig}, respectively (equations 5.14a and 5.14b),

$$Q_7(x,z) = \frac{\partial\big[F^\dagger P_7\big](z)}{\partial c_0(x)} \quad \text{and} \quad Q_8(x,z) = \frac{\partial\big[F^{\mathrm{T}} P_8\big](z)}{\partial c_0(x)}, \tag{5.39}$$

with calculated data $P_7 = F\xi$ and $P_8 = (F^\dagger)^{\mathrm{T}}\gamma$.

As in section 5.2.3, we now consider that the direct problem is solved in N iterations and the adjoint problem is solved in M iterations. Then the total gradient can be expressed as the sum

of the following ten contributions,

$$
\begin{cases}
G_1^{(N,M)}(x) = -\displaystyle\int_y Q_1^{(N+1)}(x,y)\eta^{(N+1,M+1)}(y)\,dy & (5.40a)\\
G_2^{(N,M)}(x) = -\displaystyle\int_z Q_2^{(N+1,M+1)}(x,z)\xi^{(N+1)}(z)\,dz & (5.40b)\\
G_3^{(N,M)}(x) = \displaystyle\int_y Q_{mig}(x,y)\eta^{(N+1,M+1)}(y)\,dy & (5.40c)\\
G_4^{(N,M)}(x) = -\dfrac{2}{c_0}\left[F^T P^{obs}\right](x)\eta^{(N+1,M+1)}(x) & (5.40d)\\
G_5^{(N,M)}(x) = \dfrac{2}{c_0}\left[F^T F\xi^{(N+1)}\right](x)\eta^{(N+1,M+1)}(x) & (5.40e)\\
G_6^{(N,M)}(x) = \dfrac{2}{c_0}\xi^{(N+1)}(x)\left[F^T F\eta^{(N+1,M+1)}\right](x) & (5.40f)\\
G_7^{(N)}(x) = \displaystyle\int_z Q_7^{(N+1)}(x,z)\left[C^T A^T ACK\xi^{(N+1)}\right](z)\,dz & (5.40g)\\
G_8^{(N)}(x) = \displaystyle\int_y Q_8^{(N+1)}(x,y)\xi^{(N+1)}(y)\,dz & (5.40h)\\
G_9^{(N)}(x) = -\dfrac{2}{c_0}\left[C^T A^T ACK\xi^{(N+1)}\right](x)\left[K^T\xi^{(N+1)}\right](x) & (5.40i)\\
G_\beta^{(N)}(x) = \beta c_0^{2\beta-1}\left[ACK\xi^{(N+1)}\right]^2(x), & (5.40j)
\end{cases}
$$

with $P_7^{(N+1)} = F\xi^{(N+1)}$ and $P_8^{(N+1)} = (F^\dagger)^T C^T ACK\xi^{(N+1)}$. Assuming that both problems are perfectly solved, that the regularisation weight is small and that K is close to the identity, eight contributions cancel out,

$$
\begin{cases}
G_1 + G_3 \simeq 0, & (5.41a)\\
G_4 + G_5 \simeq 0, & (5.41b)\\
G_2 + G_8 \simeq 0, & (5.41c)\\
G_6 + G_9 \simeq 0, & (5.41d)
\end{cases}
$$

and the two remaining contributions G_7 and G_β are close to the two contributions of the direct inversion gradient (equation 5.13b).

Still using the same numerical example, we solve the direct and adjoint problems in 10 iterations with very small regularisation. Applying the filter yields a reflectivity section close to the direct inverse ξ_{inv} (figure 5.19). The convergence of the adjoint problem is not as good as with the previous "hand-made" filter, but is still better than in the reference case (figure 5.20), and 10 iterations already provide a good solution (figure 5.21). Eventually we obtain a gradient which is very close to the one obtained after direct inversion (figure 5.22). Some unwanted oscillations are still visible. They are progressively attenuated if more iterations are performed both on the direct and adjoint problems (figure 5.23). Contrary to the preceding examples, applying a gaussian blur to the gradient obtained after 10 iterations would yield a velocity

update free of anomaly at the reflector position, which is preferable to additional iterations in the perspective of more expensive 2D applications.

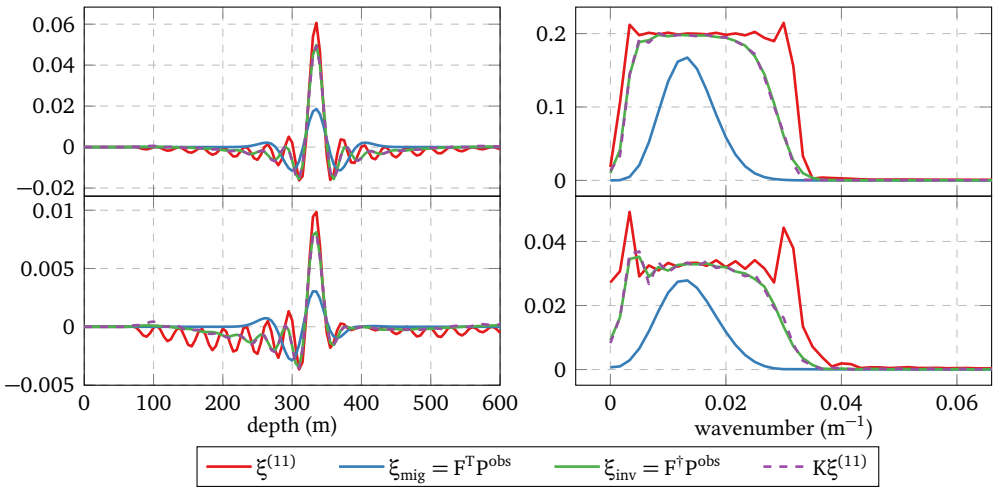


Figure 5.19. – Same as figure 5.15 for the filter $K = F^\dagger F$. The curves corresponding to ξ_{inv} and $K\xi^{(11)}$ are almost superimposed.

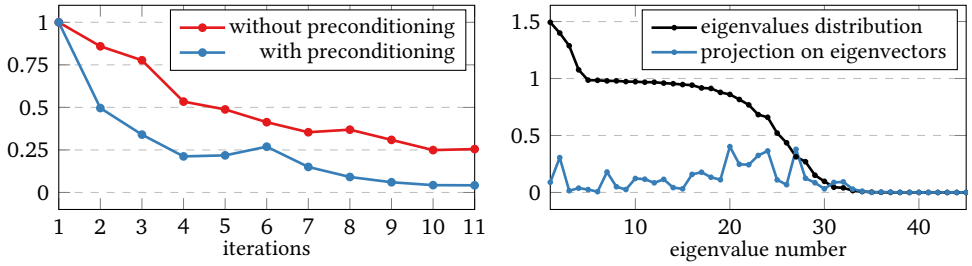


Figure 5.20. – Same as figure 5.16 for the filter $K = F^\dagger F$.

Finally, we investigate the stability of the gradient with respect to the number of iterations used to solve both the direct and adjoint problems. We compute the gradient obtained after each iteration on $\xi^{(n)}$ with and without the filter $K = F^\dagger F$. We also study the influence of the preconditioner on the stability of the gradient computation. For the sake of simplicity, we consider that the same number of iterations is performed to solve the adjoint problem ($M = N$). We plot the mean value of the gradient obtained between $z = 0$ m to 250 m, that is above the reflector (figure 5.24). Without filter and without preconditioner, the value of the gradient is quite chaotic and its sign may change from an iteration to the next one. This is consistent with observations made by Huang (2016). Here we provide a numerical example where the sensitivity is so large that it depends on the precise number of iterations. The results are slightly better in the preconditioned case but still very oscillating. We obtain a much more stable behaviour when the filter is introduced in the objective function, the mean value being constant after 30 iterations.

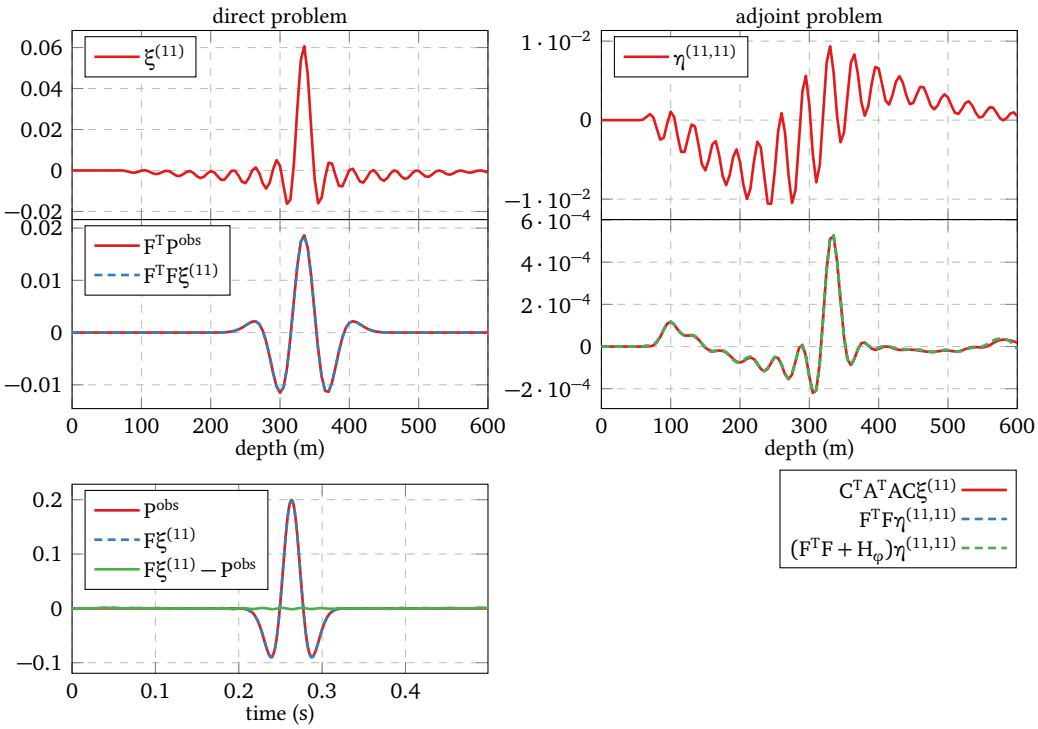


Figure 5.21. – Same as figure 5.17 for the filter $K = F^\dagger F$.

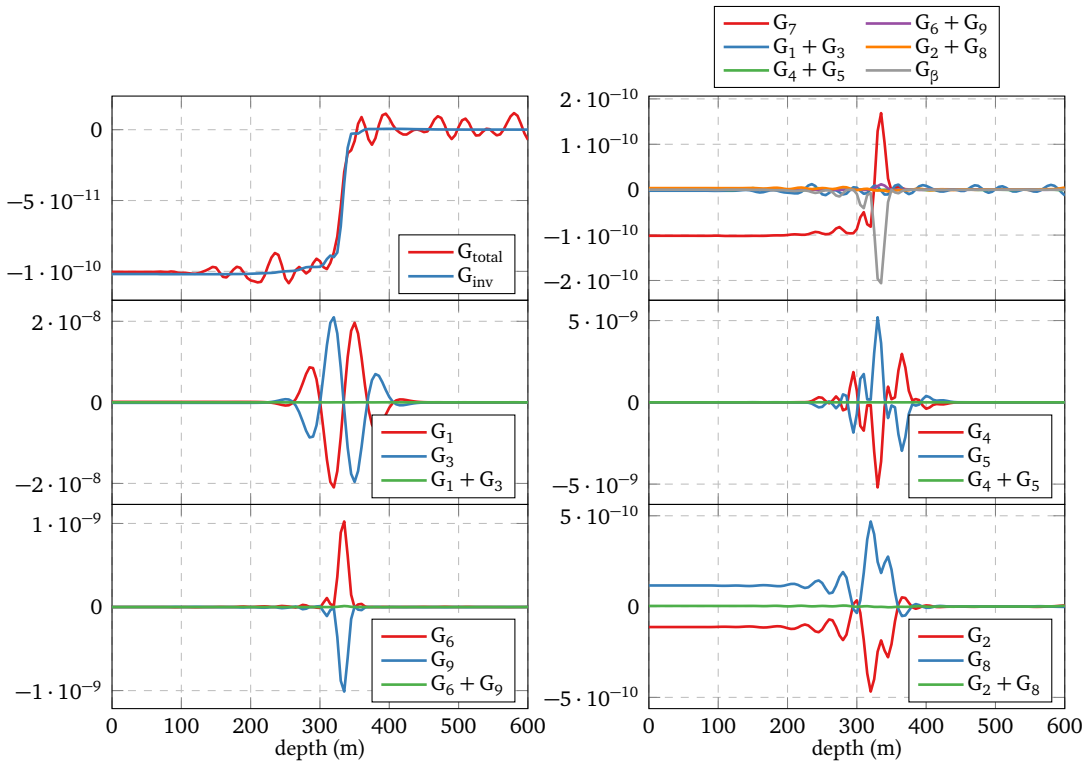


Figure 5.22. – Same as figure 5.18 for the filter $K = F^{\dagger}F$.

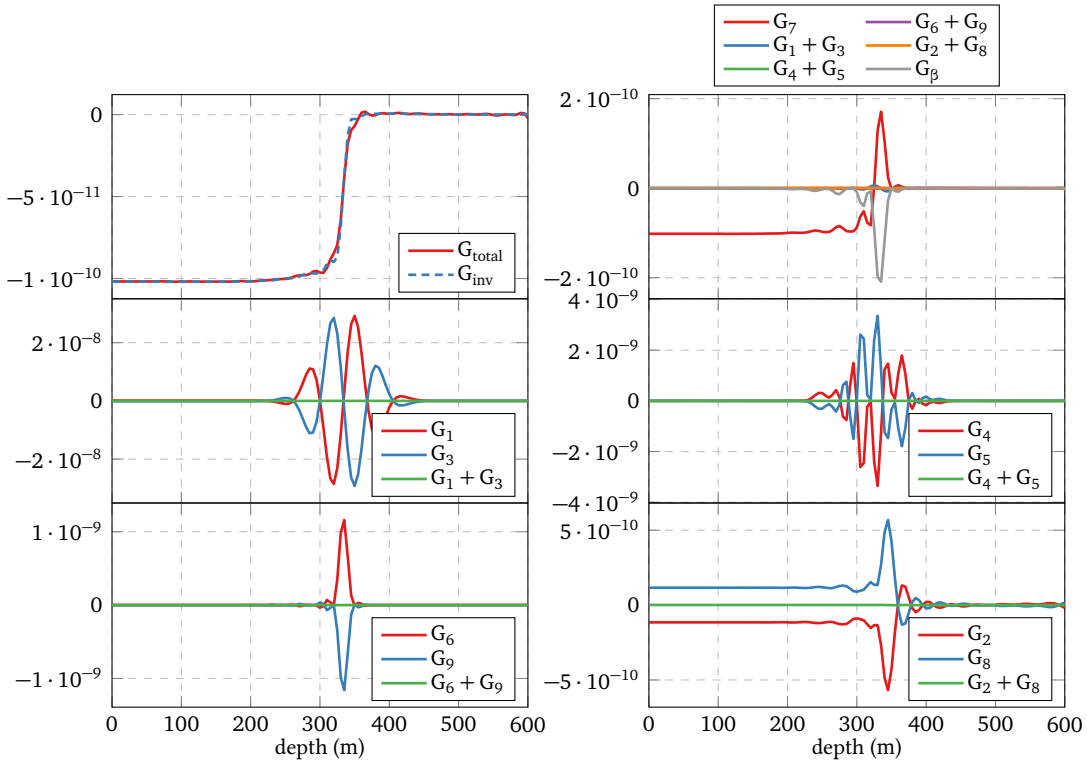


Figure 5.23. – Same as figure 5.22 except that 100 iterations (instead of 10) are performed to solve both the direct and adjoint problems.

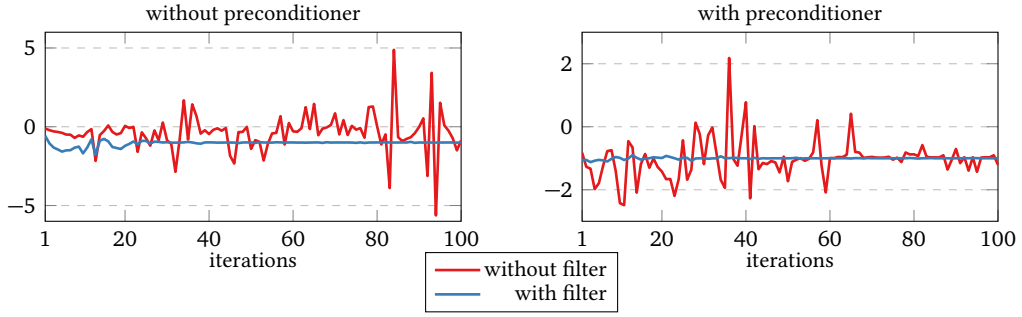


Figure 5.24. – Mean value of the gradient between z =0 m to 250 m (above the reflector where it is supposed to be constant with depth) obtained without and with the filter K = $F^{\dagger}F$ after performing N =1 to 100 iterations of iterative migration. The adjoint problem is solved with the same number of iterations as the direct problem (M = N). Small ℓ_2 regularisation is applied. For each case we plot the result obtained without (left) and with (right) preconditioner introduced in the resolution of both direct and adjoint problems. Each plot is normalised individually by the mean value obtained between iterations 50 to 100. The expected mean value is negative as this is the case of the blue curves. Locally the sign of the gradient is incorrect (positive values, red curves).

In summary, the introduction of the filter $K = F^\dagger F$ in the definition of the MVA objective function is essential to improve the stability of the gradient obtained after iterative migration and provides background velocity updates similar to those obtained after direct inversion, while keeping the possibility to perform iterations. We now want to test the applicability and efficiency of this new approach in the 2D case.

5.3. 2D-Examples

In this section, we apply the new approach consisting of introducing a filter in the objective function of MVA:

$$J_1(c_0) = \frac{1}{2}\left\|A[c_0]K[c_0]\xi^{(N+1)}[c_0]\right\|^2_{\mathscr{E}}. \tag{5.42}$$

We will consider only the filter $K = F^\dagger F$ based on the approximate inverse $F^\dagger$. From now on, the reflectivity is defined as a squared slowness perturbation ($\xi = 2\delta c/c_0^3$). In equation (5.42), the annihilator consists of a multiplication by $|h|$ and also includes a multiplication by a power of the background velocity model c_0^β, with $\beta = 3/2$.

In section 5.3.1, we consider primaries only and an example similar to the one of chapter 4. The objective is to determine if the introduction of the filter K has the same benefits as in the pure 1D case. We compare the results obtained with and without filtering final CIGs and expect the filter to improve the convergence rate of the adjoint problem and to yield gradients similar to those obtained after direct inversion.

Then we move on to an example with primaries and first-order surface-related multiples (section 5.3.2). This case was not considered in the 1D-analysis of the preceding section. We want to determine if the new approach extend to the case of multiples and yield coherent velocity updates.

5.3.1. Primaries only

We first consider an example with primaries only similar to the one presented in section 4.4.3, to compare two approaches, without and with the filter $K = F^\dagger F$. A single flat reflector is located at 300 m depth in a too low homogeneous velocity model (2500 m/s). The exact background velocity model is homogeneous too (3000 m/s). The model is 1620 m large and 450 m deep and is discretised on a 6 m × 6 m grid. Sources are located at each point of the surface with receiver also at every grid point within a maximum surface offset of ±540 m.

We perform seven iterations to solve the direct problem with preconditioning and small ℓ_2 regularisation, leading to $\xi^{(8)}$. After application of $K = F^\dagger F$ to this final reflectivity, we obtain a CIG very close to the one obtained by direct inversion (figure 5.25). In particular, spurious oscillations located above and below the reflector have been greatly attenuated. In $\xi^{(8)}$, we observe around $z = 350$ m spurious events with an opposite curvature from the main event. We do not have a formal explanation for these events which may not be due to edge effects as PML and tapers are implemented on each edge of the model. Note that they are greatly attenuated by the filter.

We now consider the adjoint problem. With the filter, the right-hand side term reads $K^T A^T A K \xi^{(8)}$ instead of $A^T A \xi^{(8)}$. As in 1D, this new vector can be expressed as $F^T Q$, result

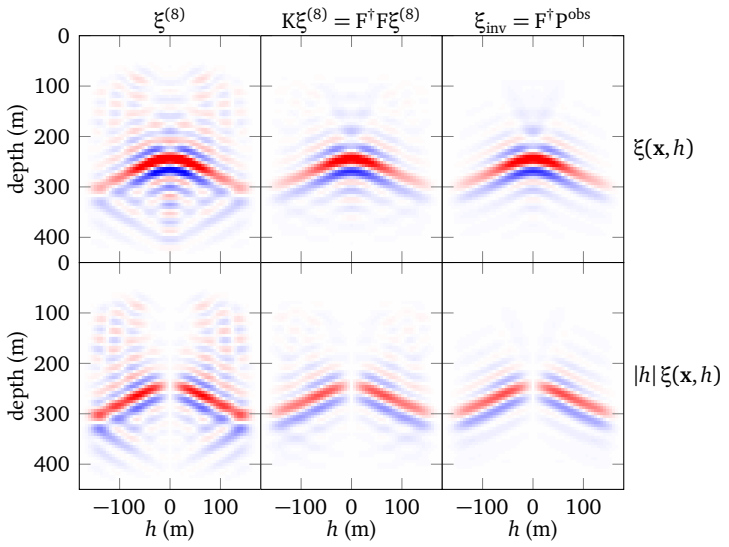


Figure 5.25. – Central CIG (top) obtained after sever iteration of preconditioned iterative migration (left). After application of $F^\dagger F$, we obtain a CIG (middle) very close to the result of direct inversion (right). The bottom row display the same CIGs multiplied by the absolute value of the subsurface offset. The same colour scale is used in each row.

of application of the adjoint to a data set $Q = (F^\dagger)^T A^T A K \xi^{(8)} \in \mathscr{D}_{obs}$. As a consequence it is similar to the right-hand side term of the direct problem $F^\dagger P^{obs}$ with an event curved upward looking like usual migration artefacts in addition to the downward event corresponding to the true reflector (figure 5.26, third column). The adjoint problem is solved in 10 iterations, with and without preconditioning for comparison. With the application of the filter, the adjoint problem is easier to solve and converge much faster, contrary to the original case (figure 5.27). Moreover, the preconditioner accelerates the convergence rate, which is also satisfactory. Note however that the convergence rate is still slower than for the direct problem. The last iterates $\eta^{(8,11)}$ in the two cases are both very oscillating (figure 5.26, first column), but in the new approach, the residuals of the linear system are much smaller (figure 5.26, fourth column).

We now consider the sequence of gradients of J_1 with respect to c_0 associated to the successive values of adjoint variables obtained without and with the filter when the adjoint problem is solved with preconditioning. With the filter, this sequence converges to a stable gradient and the final value is reached after five iterations (figures 5.28 and 5.29). Moreover, the final gradient exhibit small oscillations, but is very close to the one obtained by direct inversion (figure 5.30). This result is very similar to the one obtained in 1D (figure 5.22).

As in the 1D case, the gradient can be decomposed into several parts. But contrary to the 1D case where we identified ten contributions (equation 5.40), the reflectivity has been defined here as a square slowness perturbation $\xi = 2\delta c/c_0^3$. With this parametrisation, the contributions G_4, G_5, G_6 and G_9 in equation (5.40) are zero. The remaining contributions are $G_7 + G_\beta$, $G_2 + G_8$ and $G_1 + G_3$ (figure 5.31, right column, top, middle and bottom plots respectively). As $\xi^{(8)}$ and $\eta^{(8,11)}$ are good solutions of the direct and adjoint problems respectively, the second and third

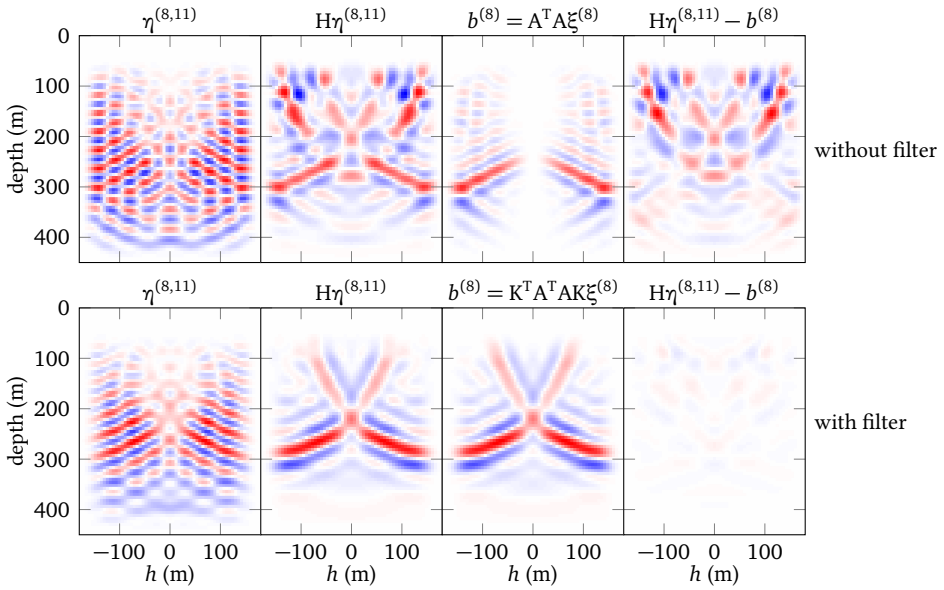


Figure 5.26. – Central CIG of the last iterate $\eta^{(8,11)}$ obtained in the resolution of the adjoint problem (1st column) and application of the Hessian H to $\eta^{(8,11)}$ (2nd column). We also display the right-hand side term $b^{(8)}$ of the linear system (3rd column) and the corresponding residual (4th column). We consider the case where no filter is introduced in the MVA objective function (top, similar to figure 4.29), and the case with a filter $K = F^\dagger F$ based on the approximate inverse (bottom). The three most right plots of each line share the same colour scale.

contributions are small with respect to the first one. The latter originates from the operator $F^\dagger$ in the filter and has a physical meaning. It is the gradient that would be obtained by applying the direct inversion strategy to primary data re-computed with first-order Born modelling from the final reflectivity $F\xi^{(8)}$. As this data set is very close to observed data, this contribution is very close to the gradient obtained with the direct inversion approach (figure 5.30, top). Although this contribution is not the gradient of an objective function, we discuss in section 5.3.3 an alternative strategy using it as background velocity update.

Finally we comment on the edge effects visible on the left and right side of the gradient (figure 5.32, bottom left). Their extension (x =0 m to 550 m and x =1000 m to 1620 m) is approximately equal to the value of the maximum surface offset (540 m, here). Note that in the reflectivity section, the inhomogeneous part only is half of the maximum surface offset (figure 5.32, top left). If we keep the same value for the maximum surface offset and extend the model laterally, the homogeneous central part is extended, but not the part altered by edge effects (figure 5.32, right). Therefore, in this example as well as in the following, we focus on the smoothness and coherency of the gradient in the central part. The attenuation of these artefacts is not specifically addressed in this thesis. We observe that tapers on source and on receivers positions, properly included in the gradient computation, help mitigating this undesirable effect. This issue deserves nonetheless further investigation.

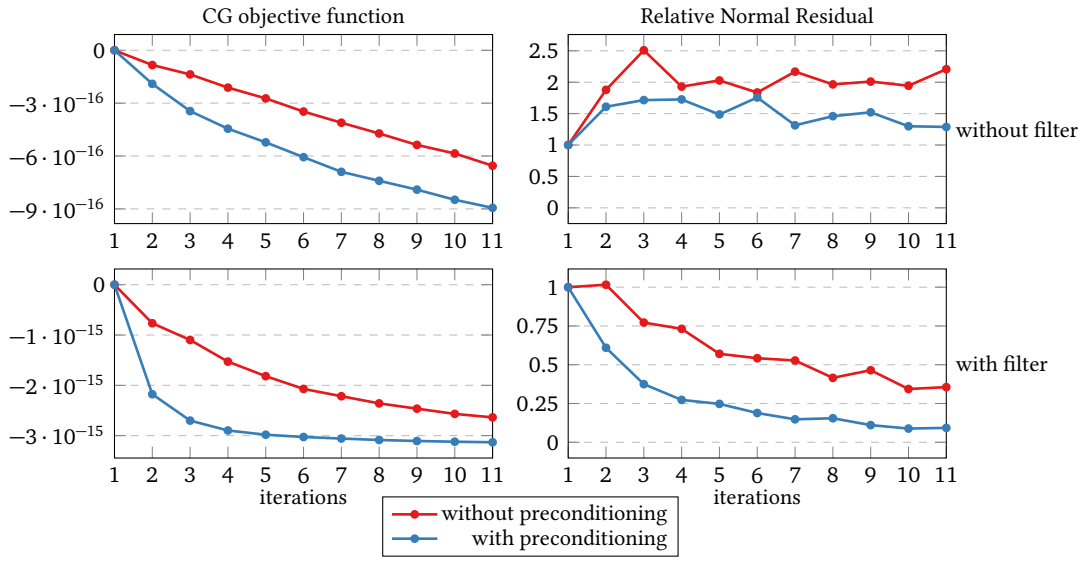


Figure 5.27. – Convergence rate of the adjoint problem without (top) and with (bottom) application of the filter $F^{\dagger}F$. The left column shows the decrease of the norm associated with the linear conjugate gradient algorithm, the right column shows the norm of residuals of the linear system.

In summary, the introduction of the filter has a similar effect in the pure 1D and the 2D cases. It improves the convergence rate of the adjoint problem and the associated gradient is close to the result obtained after direct inversion. Also, the introduction of the filter relax the requirement of sufficiently strong regularisation illustrated in the examples of chapters 3 and 4 where no filter was considered.

5.3.2. Primaries and Multiples

We want to test this new approach on a case with first-order surface-related multiples. We consider an example similar to the one studied in section 4.4.2. The exact model is laterally invariant with a single reflector located at 300 m depth and a background velocity model increasing with depth (figure 5.33), leading to two events in observed data (figure 5.34). The initial background velocity model is also increasing with depth, with the correct velocity at the surface but with a too low gradient. In all the following examples, we add a multiplication by z in the definition of the annihilator to emphasise the impact of multiples. This weight is taken into account in the derivation of the gradient.

For reference, we first apply the direct inversion strategy. If we remove multiples from observed data, we obtain a single event in CIGs, with a downward curve and the gradient is negative and homogeneous above the reflector, which is consistent with the too low initial velocity model (figure 5.35a). With both primaries and multiples in observed data, a new event corresponding to the multiple interpreted as a primary appears around 600 m depth (figure 5.35b). The actual multiple travelled twice in the exact model between the surface and

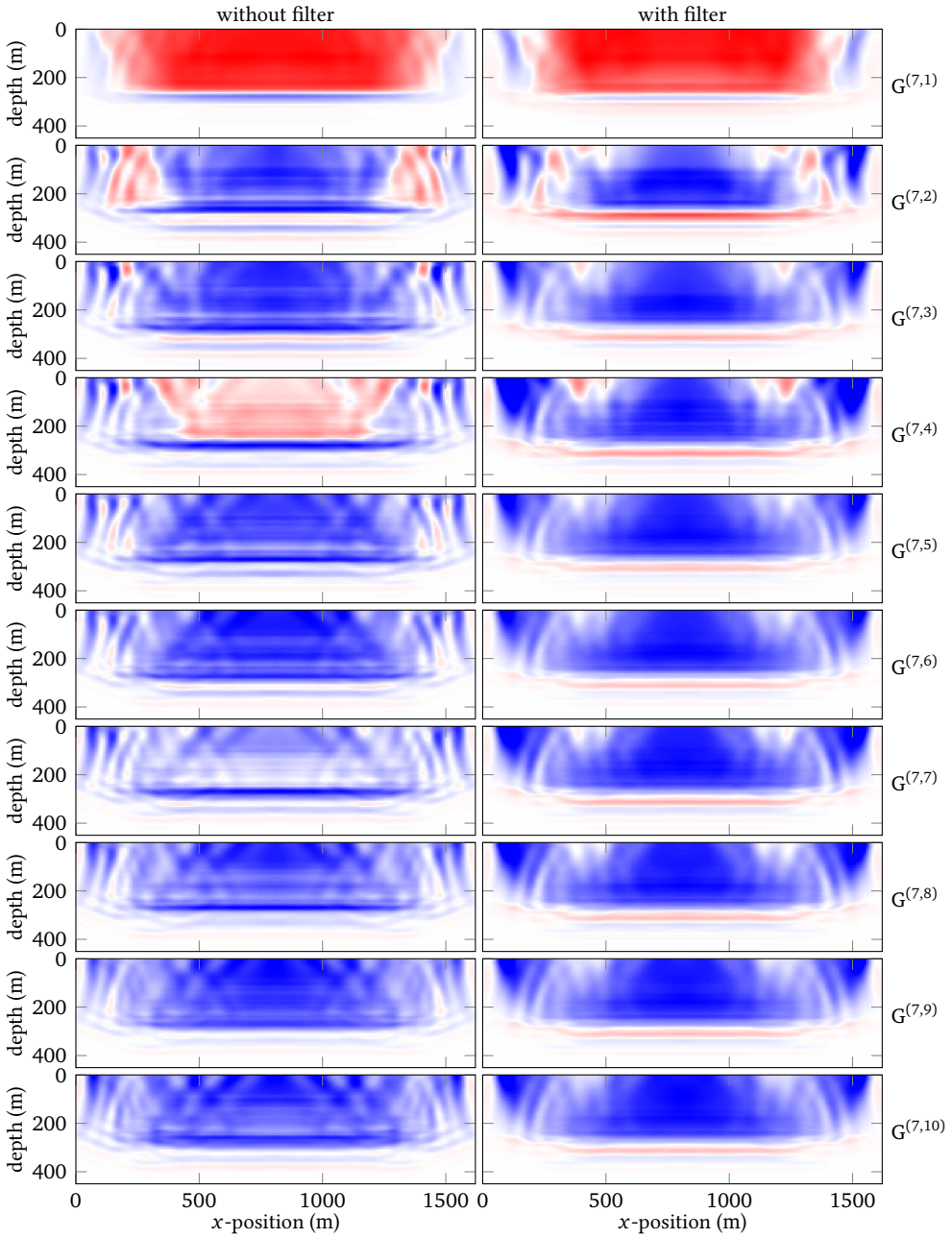


Figure 5.28. – Gradient $G^{(7,M)}$ obtained with $\xi^{(8)}$ and with the successive adjoint variables $\eta^{(8,M+1)}$ without (left) and with (right) application of the filter $K = F^{\dagger}F$. The left column is close to the right column of figure 4.30 where a very similar model was considered.

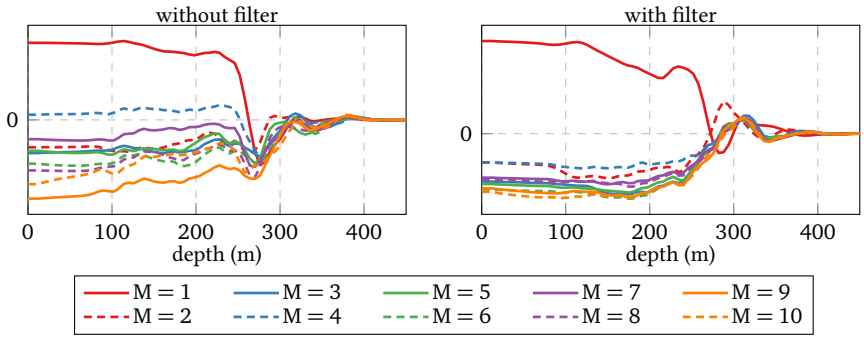


Figure 5.29. – Section of the gradients at the middle x-position obtained with the final reflectivity $\xi^{(8)}$ and the successive values of the adjoint variables $\eta^{(8,M+1)}$, without (left) and with (right) the filter $K = F^{\dagger}F$.

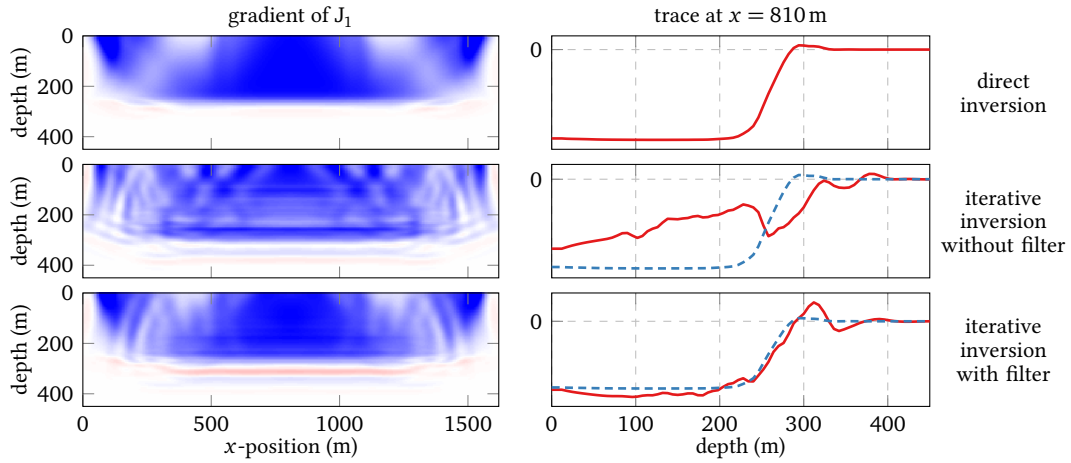


Figure 5.30. – Gradient obtained after direct inversion (top), and after iterative inversion without (middle) and with (bottom) filter. For the iterative case, seven and ten iterations are performed to solve the direct and adjoint problems respectively. The right column displays a section of the gradient at the middle position $x = 810\,\mathrm{m}$. The blue dashed line is the gradient obtained after direct inversion (top) and is displayed for comparison. The middle row is close to the bottom row of figure 4.32 where a very similar model was considered.

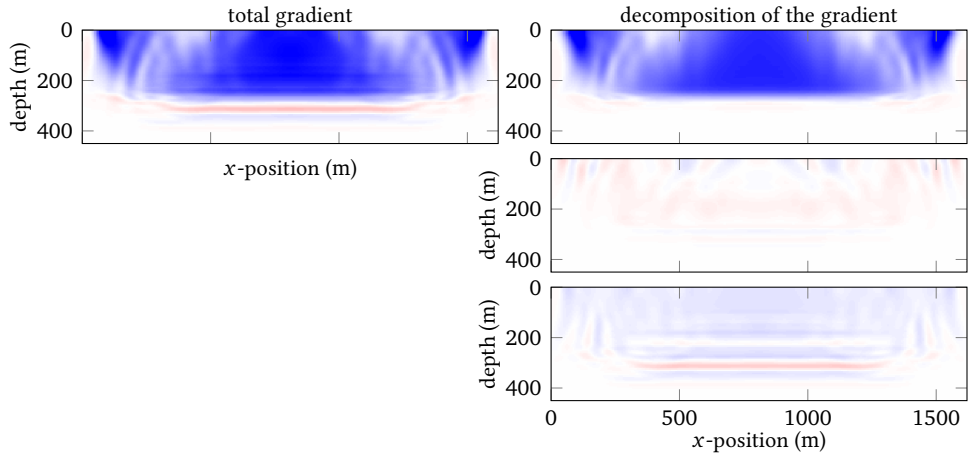


Figure 5.31. – Total gradient obtained with the filter (top left, similar to the bottom left plot in figure 5.30. The gradient is decomposed into three part (right column). Using the notations of equations (5.40), the first is due to the approximate inverse ($G_7 + G_\beta$, top), the second should be zero if both direct and adjoint problems are perfectly solved ($G_2 + G_7$, middle) and the third should be zero if the direct problem is perfectly solved and the data residuals zero ($G_1 + G_3$, bottom).

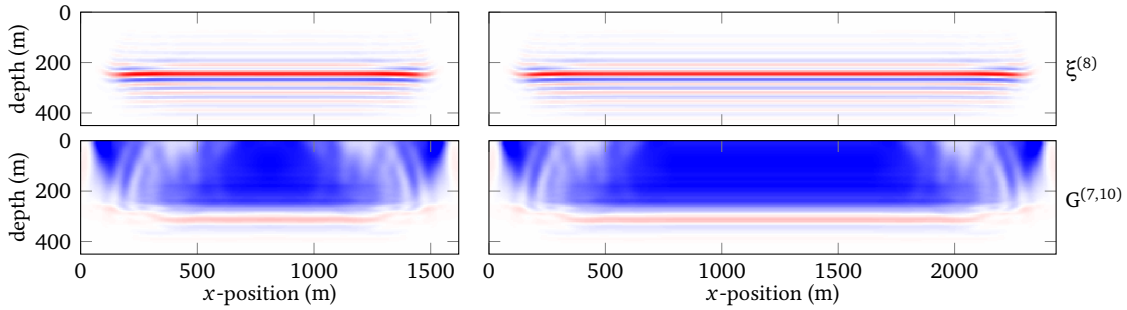


Figure 5.32. – Section of the reflectivity $\xi^{(8)}$ at $h = 0\,\text{m}$ (top) and gradient $G^{(7,10)}$ (bottom) obtained with $x_{\max} = 1620\,\text{m}$ (left, same gradient as figure 5.30 bottom left), and with $x_{\max} = 2430\,\text{m}$. Extending the model laterally for a similar value of the maximum surface offset results in a larger homogeneous central part, the edge effects remaining similar.

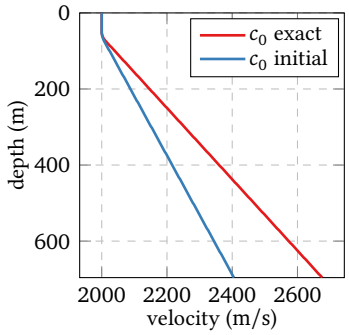


Figure 5.33. – Exact (red) and initial (blue) background velocity model. The reflectivity model consists of a single reflector located at 300 m depth.

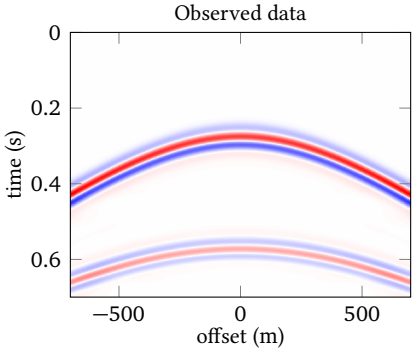


Figure 5.34. – Observed data obtained for a source located at the middle position on the surface in the exact model of figure 5.33.

the exact reflector position. It is interpreted as a primary travelling in the initial velocity model between the surface and a depth roughly twice the one of the exact reflector. Hence, this "imaginary" primary has "seen" a higher velocity than the true multiple. As a consequence, part of the energy of the cross-talk artefact is defocused with an upward curvature and the gradient has positive contribution above the reflector

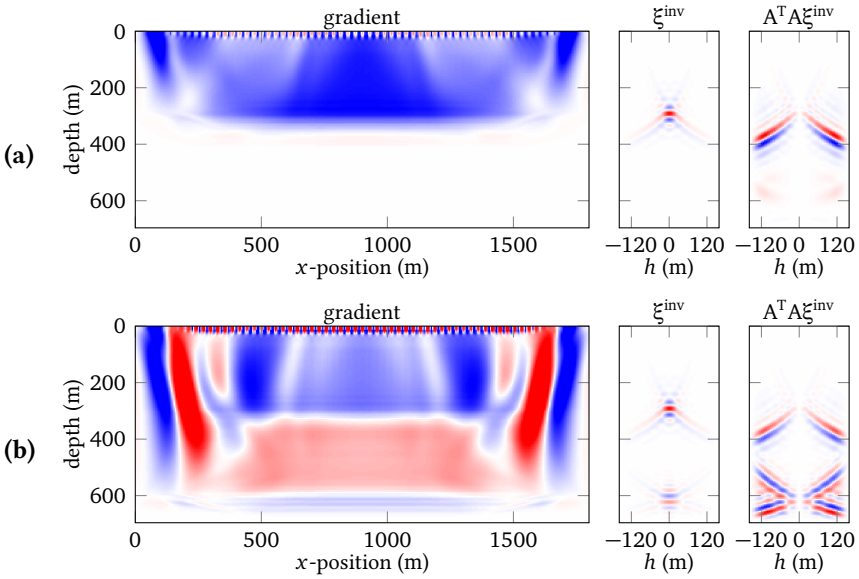


Figure 5.35. – Gradient and CIGs obtained in the initial model shown in figure 5.33, after direct inversion with observed data containing (a) primaries only or (b) both primaries and multiples.

We now consider iterative inversion with the filter $F^\dagger F$ introduced in the definition of the

MVA objective function (equation 5.42). First, we study the case of observed data containing only the primary reflection (figure 5.36). We obtain a gradient very similar to the case of direct inversion (figure 5.35a), which indicates that the new approach works well in non-homogeneous models as well.

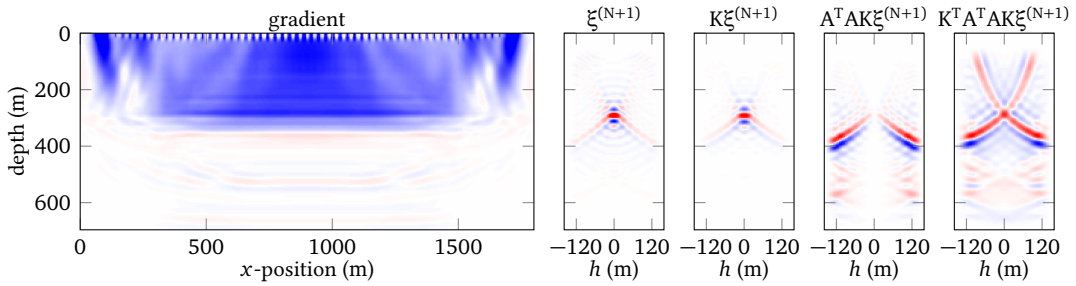


Figure 5.36. – Gradient and CIGs obtained in the initial model shown in figure 5.33, after iterative inversion with observed data containing primaries only. The direct and adjoint problems are solved in $N = 7$ and $M = 10$ iterations respectively, and with small ℓ_2 regularisation.

We now consider the full data set with the primary reflection and the first-order surface multiple and use preconditioned non-linear optimisation as in section 4.4.2. The linear adjoint problem reads in the general case

$$\left[\frac{\partial^2 J_0}{\partial \xi^2}\left[c_0, \xi^{(N+1)}\right]\right]\eta = A^T K^T K A^T \xi^{(N+1)}, \tag{5.43}$$

the difference with the linear case being that the Hessian matrix depends on the final value of ξ. This operator reads

$$\begin{aligned}\frac{\partial^2 J_0}{\partial \xi^2} =& \frac{\partial^2 P_3}{\partial \xi^2}(P_1 + P_3 - P^{obs}) + \left(\frac{\partial P_1}{\partial \xi}\right)^T\left(\frac{\partial P_1}{\partial \xi}\right) + 2\left(\frac{\partial P_1}{\partial \xi}\right)^T\left(\frac{\partial P_3}{\partial \xi}\right) + \left(\frac{\partial P_3}{\partial \xi}\right)^T\left(\frac{\partial P_3}{\partial \xi}\right) \\ &+ a_\varphi \frac{\partial^2 \varphi}{\partial \xi^2},\end{aligned} \tag{5.44}$$

where we have omitted the dependency to $\xi^{(N+1)}$ and operator M for readability. Because of the first term in this expression, the Hessian may not be positive definite in case of too strong residuals. This is an issue as the conjugate gradient algorithm is designed for positive definite matrix only. This problem should not appear if iterative migration has converged, meaning that residuals are small, or if enough regularisation is added. An additional safeguard consists of exiting the algorithm as soon as a negative curvature is encountered (Métivier *et al.*, 2013).

Using small ℓ_2 regularisation and ten iterations to solve the adjoint problem, we first compare the gradient obtained after 1 and 10 iterations of migration (figures 5.37a and 5.37b). In the first case, residuals are still strong and a negative curvature has been encountered at the 10th iteration. Therefore the final gradient is actually $G^{(10,9)}$. Because of the cross-talk artefact in the CIG, the gradient is non-zero below the reflector. After ten iterations, the artefact is weaker and is further attenuated by the application of the filter. However, the gradient is not homogeneous

and exhibits strong oscillations around the reflector. To obtain a more satisfactory result, we first increase the number of iterations for the resolution of the direct and adjoint problems (30 and 20, respectively), but the gradient is still far from being homogeneous (figure 5.37c). We notice that even after several iterations and application of the filter, the annihilator still strengthen residual energy at large values of h below the reflector (figure 5.37c, 4th column). This suggests that these areas may not be well constrained. As a remedy, we add a Huber norm to the migration objective function J_0 with a bigger weight on these parts of the CIGs:

$$\varphi[\xi] = \int_{\mathbf{x}} \int_{h} \mathrm{hub}\big(h \cdot z \cdot \xi(\mathbf{x}, h)\big)\, \mathrm{d}h\, \mathrm{d}\mathbf{x}, \tag{5.45}$$

and we use again 10 iterations to solve the direct and adjoint problems (figure 5.37d). Although the final CIGs look very similar to the one obtained in the previous case, this modification greatly improves the gradient, which is homogeneous with the correct sign above the reflector and very weak below.

As in the 1D case, we decompose the gradient into three parts (figure 5.38). The first one (figure 5.38, 1st column) is due to the operator $F^\dagger$ in the definition of J_1 (equation 5.42). As already mentioned in the linear case, this is the gradient that would be obtained by applying the direct inversion approach to observed primaries re-computed from the final reflectivity free of cross-talk artefacts. As a consequence this contribution is very similar to the one obtained by applying direct inversion to primary only (figure 5.35a), provided that the final reflectivity correctly explains observed data. The second contribution (figure 5.38, 2nd column) is directly related to data residuals, and its share in the final gradient decreases as far as the number of iterations increases (figures 5.38a to 5.38c). With regularisation, the final data residuals may be degraded, and this contribution may be not negligible any more (although it is still quite small in the example of figure 5.38d). Eventually we notice that the third part (figure 5.38, 3rd column) is relatively strong in the four cases and does not necessarily add a coherent contribution to the gradient. Contrary to the linear case, it does not vanish when the direct and adjoint problems are perfectly solved.

We conclude from this example that the new strategy consisting of adding a filter in the definition of the MVA objective function is also efficient in the case of data containing first-order surface multiples. The filter attenuates unwanted energy at large values of h as in the primaries only case and helps attenuating residual energy of cross-talk artefacts. Note however that contrary to the primaries only case, regularisation remains essential to iterative migration for a proper attenuation of cross-talk artefacts and consistent background velocity updates.

5.3.3. Alternative strategy

Both in the primaries only case and in the presence of multiples, we have isolated a contribution to the total gradient due to the operator $F^\dagger$ in the filter K and showed that it is very close to the result of direct inversion applied on primaries only (figure 5.31, top right and figure 5.38, 1st column). We propose as an alternative strategy to use this contribution only to update the background model. In this new approach, the MVA objective function to be minimised still reads

$$J_1[c_0] = \frac{1}{2}\left\| A[c_0]F^\dagger[c_0]F[c_0]\xi^{(N+1)}[c_0] \right\|_{\mathscr{E}}^2. \tag{5.46}$$

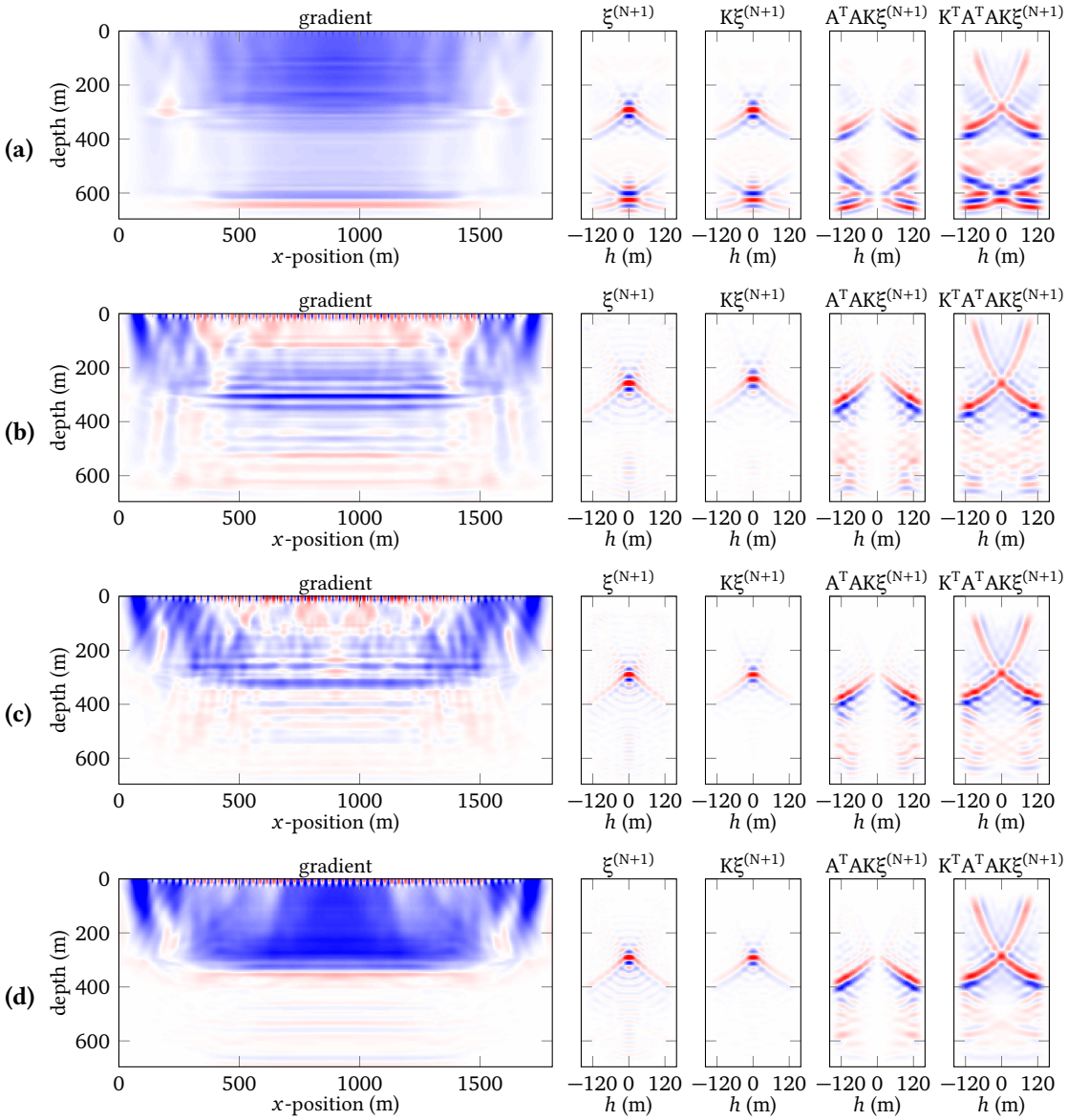


Figure 5.37. – Gradient and CIGs obtained after iterative inversion with observed data containing both primaries and multiples. Different optimisation strategies are considered, regarding regularisation and the number N and M of iterations used to solve the direct and adjoint problems:

(a) small ℓ_2 regularisation, $N = 1$, $M = 9$;
(b) small ℓ_2 regularisation, $N = 10$, $M = 10$;
(c) small ℓ_2 regularisation, $N = 30$, $M = 20$;
(d) stronger Huber regularisation, $N = 10$, $M = 10$.

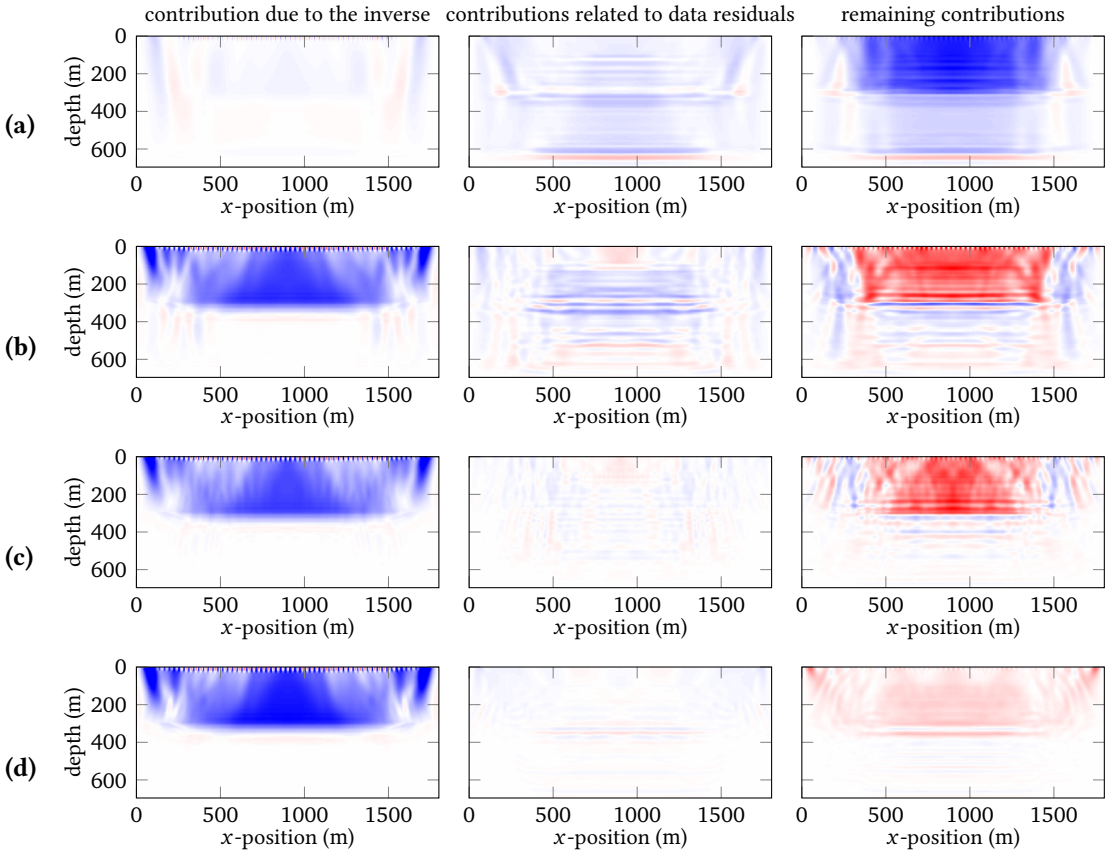


Figure 5.38. – Decomposition of the gradients of figure 5.37 into three parts, similar to the one presented in figure 5.31 in the case of primaries only. The first contribution (left) can be interpreted as the gradient that would be obtained by applying the direct inversion strategy to recalculated primaries remodelled with the last reflectivity image of the iterative migration processing. The second contribution (middle) is directly related to data residuals and should be zero if iterative migration has converged. The remaining contributions (right) are difficult to interpret and are not necessarily zero if convergence is reached for both the direct and adjoint problems.

We briefly rederive an expression for the total gradient of the objective function (5.46) with respect to c_0 and identify the contribution which is considered in the new strategy. We define the following Lagrangian,

$$\overline{J_1} = \frac{1}{2}\left\|A\zeta\right\|_{\mathscr{E}}^2 - \left\langle \gamma \,\middle|\, \zeta - F^\dagger P^{\text{calc}} \right\rangle_{\mathscr{E}} - \left\langle Q \,\middle|\, P^{\text{calc}} - F\xi^{(N+1)} \right\rangle_{\mathscr{D}} - \left\langle \eta \,\middle|\, \frac{\partial J_0}{\partial \xi}\left[c_0, \xi^{(N+1)}\right] \right\rangle_{\mathscr{E}} \tag{5.47}$$

The value of the adjoint variables γ, Q and η are obtained by zeroing the partial derivatives of the Lagrangian (5.47) with respect to the state variables ζ, P^{calc} and $\xi^{(N+1)}$,

$$\begin{cases} \gamma = A^T A \zeta^{(N+1)}, & \text{(5.48a)} \\ Q = (F^\dagger)^T \gamma, & \text{(5.48b)} \\ \left[\frac{\partial^2 J_0}{\partial \xi^2}\left[c_0, \xi^{(N+1)}\right]\right]\eta = F^T Q, & \text{(5.48c)} \end{cases}$$

and the total gradient of the objective function (5.46) with respect to c_0 is obtained as the partial derivative of the Lagrangian (5.47) with respect to c_0,

$$\frac{\partial J_1}{\partial c_0} = \overbrace{\frac{\partial}{\partial c_0}\left[\frac{\left\|A\zeta\right\|_{\mathscr{E}}^2}{2}\right] + \frac{\partial}{\partial c_0}\left[\left\langle \gamma \,\middle|\, F^\dagger P^{\text{calc}} \right\rangle_{\mathscr{E}}\right]}^{\text{“truncated gradient”}} + \frac{\partial}{\partial c_0}\left[\left\langle Q \,\middle|\, F\xi^{(N+1)} \right\rangle_{\mathscr{D}}\right] + \left[\frac{\partial^2 J_0}{\partial \xi \partial c_0}\left[c_0, \xi^{(N+1)}\right]\right]\eta. \tag{5.49}$$

In the new approach, the background velocity update is defined as the sum of the first two contributions in (5.49), the two remaining contributions being dropped. Therefore this strategy will be referred to as “*truncated gradient*” in chapter 6.

We now explain the meaning of this new approach. We apply iterative migration to a data set containing both primaries and first-order surface multiples to derive an extended reflectivity free of cross-talk artefacts. This extended image is used to re-compute primary reflections under the first-order Born approximation. The value of this wavefield at receiver positions is kinematically consistent with the primaries contained in the original observed data set and defines a new data set contains primaries only. The velocity analysis is performed on this data set as if it were observed data, using the direct inversion strategy presented in section 4.2.

The background velocity update used in the “truncated gradient” strategy is not the gradient of an objective function, but this definition has several advantages. First, it requires less computational effort as there is no need to solve the adjoint problem (5.48c) any more. Moreover the need to add proper regularisation and the difficult choice of parameters of the regularisation function is relaxed. Finally note that re-modelled primaries can be computed for source positions not present in the original acquisition, for example in the case of an irregular or incomplete acquisition. In the latter case, multiples may provide extra illumination to reconstruct a more detailed reflectivity model and compute primary reflections not recorded in the original acquisition. Also, the new strategy allows to provide the direct inversion strategy with data acquired with dense source and receiver coverage, which is one of the hypotheses made in the derivation of the approximate inverse $F^\dagger$.

5.4. Conclusion

We have studied the behaviour of iterative migration velocity analysis on a simple 1D case and observed the same properties as in the 2D case. Oscillations coming from the source deconvolution away from the reflector are amplified by the annihilator, leading to a very oscillating image residual. As a consequence, the resolution of the adjoint problem is slow and the background velocity update oscillating.

We have proposed two modifications of the original procedure to alleviate these issues. First we have shown that adding a Huber norm to the migration objective function improves the properties of the adjoint problem and the associated gradient. A disadvantage of this solution is that two parameters have to be chosen empirically. Then, we have proposed to introduce a filter in the MVA objective function, so that spurious oscillations away from the reflector are attenuated. We showed that a filter based on the approximate inverse $\mathbf{F}^{\dagger}$ yields a gradient very similar to the one obtained after direct inversion. With this new objective function, the right-hand side terms of both the direct and adjoint problems can be expressed $\mathbf{F}^{\mathrm{T}}\mathbf{Q}$, and the adjoint problem converges much faster. As the filter depends on c_0, additional terms have to be computed, as costly as two iterations of migration. *However, the filter allows to solve the adjoint problem in a reduced number of iterations, so the filter actually reduces the computational expense of the method.*

An application to a simple example suggests that this new approach works well in 2D as well. Compared to the examples of chapter 4, the adjoint problem converge faster (figure 5.27), as well as the associated sequence of gradients (figure 5.27). In this primary-only case, the need for sufficiently strong regularisation is also relaxed. However, regularisation remains essential in the case of multiples to obtain coherent velocity updates.

In the case of multiples, we have proposed an alternative strategy consisting of using the final result of iterative migration to generate a new data set made of primaries only. The direct inversion approach (section 4.2) is then applied on this new data set. Although the background velocity update defined with this strategy is not the gradient of an objective function any more, this technique considerably reduces the computational expense of the method, and relaxes the difficult requirement to find suitable regularisation parameters. Furthermore it may allow to regenerate reconstructed data with an acquisition different from observed data. For example, source positions missing in initial observed data can be added in the new data set and provide the dense source and receiver coverage assumed in the derivation of the approximate inverse.

Chapter 6.

Application to synthetic data sets

Contents

Résumé du chapitre 6

L'objectif de ce chapitre est de tester la robustesse de l'approche définie au chapitre 5 lorsque le code de modélisation utilisé pour générer les données observées est différent de celui utilisé lors de l'inversion. Dans l'optique d'applications sur données réelles (absentes de cette étude), il s'agit d'aller au-delà de la situation de « crime inverse » des tests numériques réalisés dans les chapitres précédents.

Dans la première partie de ce chapitre, je considère quatre modifications du code de modélisation des données observées, tout en gardant le même code de modélisation pour réaliser l'inversion, et j'étudie l'influence de ces modifications sur les gradients calculés avec l'inversion directe et itérative, avec et sans multiples :

A – absence de basses fréquences dans l'ondelette de source ;

B – modification de la forme de l'ondelette de source ;

C – modélisation dans un modèle de vitesse lisse mais avec un modèle de densité variable dont les discontinuités sont responsables de réflexions primaires et multiples, modélisées avec une approximation de Born du second ordre ;

D – modélisation des données avec densité et vitesse variables par différences finies sans approximation de Born et avec une condition de surface libre pour la modélisation de multiples de surface.

L'absence de basses fréquences dans l'ondelette de source a une influence négligeable sur le calcul du gradient (test A). En revanche, l'estimation de la forme de l'ondelette de source est essentielle pour la migration itérative dans le cas des multiples (en revanche, cela ne pose pas de difficultés dans le cas de primaires seuls) : si elle est trop éloignée de celle utilisée pour générer les données observées, la migration itérative ne parvient pas à construire un modèle de réflectivité expliquant à la fois les ondes réfléchies primaires et multiples (test B). L'utilisation de perturbations du modèle de densité au lieu de perturbations du modèle de vitesse pour générer les données observées (test C) n'affecte pas la cohérence des gradients. La différence principale est l'estimation de l'amplitude d'évènements correspondant à des grands angles de réflexions : elle se traduit dans le gradient par un poids différent donné aux contributions des réflecteurs peu profonds. Enfin l'utilisation d'une surface libre pour la modélisation des multiples de surface (test D) reste une difficulté, en particulier la présence de « ghosts » aux positions des sources et des récepteurs. Ils sont ici pris en compte dans la modélisation des données calculées en utilisant une dérivée seconde de l'ondelette de source, mais cette approximation n'est valide que pour des angles d'incidence nuls à la surface et la migration itérative peine à trouver un modèle de réflectivité expliquant correctement à la fois les primaires et les multiples.

Dans la deuxième partie du chapitre, des données synthétiques sont modélisées à partir d'un modèle par blocs (vitesse et densité variables) comprenant une anomalie lente de vitesse pour réaliser une inversion du modèle de vitesse. Les données observées sont modélisées avec des versions lissées de ces modèles et l'approximation de Born du second ordre (perturbations de vitesse et de densité). Lors de l'inversion, le modèle de densité est supposé constant. En pratique un lissage du gradient par un noyau gaussien est appliqué avant remise à jour du macro-modèle

de vitesse. Dans un premier temps, seules les réflexions primaires sont considérées et la stratégie d'inversion directe est utilisée. Après quinze itérations, l'anomalie de vitesse est retrouvée, les CIGs sont aussi bien focalisés que dans le modèle exact, et la fonction coût atteint un niveau proche, légèrement inférieur, à celui obtenu dans le modèle exact. Dans un deuxième temps, les réflexions multiples sont ajoutées aux données observées et l'inversion est relancée à partir du même macro-modèle initial, les CIGs étant toujours calculés avec la stratégie d'inversion directe. Comparé au premier cas, de nouveaux évènements, non focalisés pour le bon modèle de vitesse, apparaissent dans les CIGs. Ainsi, au cours de l'inversion, la valeur de la fonction coût associée à l'analyse de vitesse atteint des valeurs bien plus faibles que celle obtenue dans le modèle exact. Le modèle de vitesse final est différent de celui obtenu dans le cas de primaires seuls et ne fait pas apparaître clairement l'anomalie de vitesse. Le modèle retrouvé réalise un compromis entre la focalisation dans les CIGs des évènements correspondant aux vrais réflecteurs et la focalisation des artefacts dus aux multiples. Finalement un troisième cas est traité : la remise à jour du modèle de vitesse est calculée par la méthode du « gradient tronqué » après cinq itérations de migration itérative. La migration itérative permet d'atténuer les artefacts dus aux multiples et le modèle final retrouvé est proche de celui obtenu dans le cas de primaires seuls avec l'inversion directe. Cela indique que les multiples ont été correctement interprétés au cours de l'inversion.

La dernière partie du chapitre est consacré au cas d'un trou d'acquisition, avec des sources manquantes au milieu du modèle. Les réflexions multiples peuvent potentiellement apporter une information complémentaire dans la partie centrale du modèle, non illuminée par les réflexions primaires. Le modèle de vitesse exact est latéralement invariant et un modèle homogène est utilisé comme modèle initial pour calculer le premier gradient (le macro-modèle n'est pas remis à jour dans cet exemple). On constate une absence d'énergie dans la partie centrale du modèle de réflectivité obtenue par inversion directe dans le cas où les données observées ne comprennent que des réflexions primaires. Il en va de même pour le gradient associé. Il est intéressant de remarquer que, même en ne considérant que les réflexions primaires, la migration itérative permet de combler une partie du trou d'acquisition dans l'image de réflectivité et dans la remise à jour du macro-modèle. En revanche l'introduction des réflexions multiples (à la fois dans les données observées et dans les données calculées) ne permet pas de combler d'avantage le trou d'acquisition, même après vingt itérations sur le modèle de réflectivité.

6.1. Introduction

The aim of the thesis was to provide a more robust MVA procedure based on iterative migration, able to deal with multiple reflections. In chapters 2 to 5, we have progressively built a strategy, whose final formulation can be summarised in four steps:

- preconditioned iterative migration to determine the reflectivity section $\xi^{(N+1)}$ best explaining observed data. The preconditioner is a pseudo-inverse $F^{\dagger}$ of the extended modelling operator F. It has been derived for the case of primaries only (chapter 4) and is similar to the one proposed by Hou and Symes (2015);
- application of the filter $K = F^{\dagger}F$ to $\xi^{(N+1)}$ to remove unwanted energy at large values of h that does not have any meaning in terms of wave propagation kinematics. The filtered reflectivity $\zeta = K\xi^{(N+1)}$ is used as input of the usual MVA objective function

$$J_1[c_0] = \frac{1}{2}\left\|A\zeta[c_0]\right\|^2_{\mathscr{E}}; \tag{6.1}$$

- computation of the image residual $K^T A^T A\zeta$ and resolution of a linear adjoint problem (even in the case of multiples) for the adjoint variable η

$$\left[\frac{\partial^2 J_0}{\partial \xi^2}\left[c_0, \xi^{(N+1)}\right]\right]\eta = K^T A^T AK\xi^{(N+1)}; \tag{6.2}$$

- computation of the approximate gradient $G^{(N,M)}$, using the last iterate $\xi^{(N+1)}$ and the last iterate $\eta^{(N+1,M+1)}$, as

$$G^{(N,M)} = \left[\frac{\partial^2 J_0}{\partial \xi \partial c_0}\left[c_0, \xi^{(N+1)}\right]\right]\eta^{(N+1,M+1)}. \tag{6.3}$$

We have also introduced in section 5.3.3 an alternative strategy for the derivation of a background velocity update after iterative migration. It consists of using the extended reflectivity image resulting from iterative migration to recompute primary reflection data. Then the macro-model update is obtained with the direct inversion strategy applied to this new data set free of multiples, as if it was observed data. The background velocity update obtained with this method is only one of the contribution to the gradient of the objective function (6.1) with respect to c_0, and is therefore referred to as "*truncated gradient*" in this chapter. The advantage of this strategy is that there is no need to solve the adjoint problem any more, so that the computational expense of the method is considerably reduced. Furthermore the direct inversion strategy has been shown to provide smooth consistent gradients in a wide variety of cases (section 4.2.6). Therefore we expect the *truncated gradient* to benefit of its favourable behaviour also in the case of multiples.

Currently the iterative migration strategy has been applied only on observed data modelled with a second-order Born approximation, that is with the same modelling tool as the one used in inversion. This is referred to as the "inverse crime" (Wirgin, 2004) in inverse problems literature.

The purpose of this chapter is first to evaluate the robustness of the approach when synthetic observed data are obtained with a different modelling code. This may allow us to determine possible limitations of our approach in the perspective of real data applications (not considered in this thesis). The second objective is to go beyond the computation of the first gradient and run several non-linear iterations to update the background model. Finally we examine the ability of our MVA strategy to benefit from the additional information contained in first-order surface-related multiples.

In the first part of this chapter (section 6.2), we run four tests, named A, B, C and D, summarised in table 6.1. In tests A and B, we model observed data with the constant density acoustic wave-equation and a *modified source wavelet*, the Ricker wavelet still being used for inversion. In tests C and D, observed data are modelled with a *variable density* acoustic propagation code, first under the Born approximation, then with a full finite-difference modelling and a *free-surface condition*. For each test, we compute six background velocity updates using different strategies (the letters in the list below correspond to the labels of subfigures in section 6.2):

(a) direct inversion on primary reflection data only;
(b) direct inversion on data containing both primaries and first-order surface-related multiples;
(c) iterative inversion on primary reflection data only;
(d) iterative inversion on data containing both primaries and first-order surface-related multiples;
(e) smoothed version of the gradient obtained in (c) after application of a gaussian blur;
(f) "truncated gradient" obtained in the case of primaries, as described above;
(g) same as (e) in the case of multiples (i.e. smoothed version of (d));
(h) same as (f) in the case of multiples.

The result of direct inversion (a) on primaries only is used as reference. The case (b) is considered to assess how multiples misinterpreted as primaries alter the result of (a). Then we compare the result of iterative inversion on multiples (d) with (a) to assess the robustness of our approach. To determine if potential failures of iterative inversion in the presence of multiples are due to multiples or to the iterative strategy, we also compute the gradient obtained by iterative inversion in the case of primaries only (c). Smoothing is commonly applied before updating the velocity model, therefore we show a smoothed version ((e) and (g)) of the gradient obtained after iterative inversion (d).

The same exact (in principle unknown) velocity model is used for tests A, B and C. Those tests are compared to a reference case where the same code is used for the modelling of observations and for inversion. The model is 1800 m large and 700 m deep, laterally invariant and discretised on a 4.8 m × 4.8 m grid. The background velocity model is increasing with depth from 2000 m/s at the surface to approximately 2700 m/s at 700 m depth (figure 6.1, left). The initial velocity model is similar to the exact model except that velocities are underestimated (from 2000 m/s at the surface to 2400 m/s at the maximum depth). Sources are located every 6 grid points at the surface with receivers at each grid point on the surface within ±700 m around the source. The model used in test D is different and will be presented in section 6.2.5.

For all tests A, B, C and D, we define an annihilator similar to the one used at the end of

test	section	figures	source wavelet	modelling tool for observed data
ref.	6.2.1	6.3 and 6.4	Ricker wavelet	2nd-order Born approximation
A	6.2.2	6.6 and 6.7	Ricker wavelet without low frequencies	2nd-order Born approximation
B	6.2.3	6.9 and 6.11	time derivative of Ricker wavelet	2nd-order Born approximation
C	6.2.4	6.16 and 6.19	Ricker wavelet	2nd-order Born approximation density perturbation instead of velocity perturbation
D	6.2.5	6.23 and 6.24	Ricker wavelet	finite difference and free-surface condition with variable density

Table 6.1. – Series of tests performed in chapter 6. The coloured cells correspond to the modification introduced in the modelling code for observed data compared to the modelling code used during inversion.

chapter 5,

$$A\xi(z,x,h) = |h| \cdot z \cdot c_0^{\beta}(z,x)\xi(z,x,h). \tag{6.4}$$

It consists of a multiplication by the absolute value of the subsurface offset to penalise defocused energy as well as a power $\beta = -3/2$ of the background velocity model which has been shown in chapter 4 to attenuate spurious oscillations appearing in the MVA gradient around reflectors' positions. To emphasise the influence of deeper reflectors which may be hidden by cross-talk artefacts, we add a multiplication by the depth z in the definition of the annihilator.

In the second part of this chapter (section 6.3), a blocky velocity model and a variable density model both with lateral variations are defined to generate observed data under a second-order Born approximation. We run several outer iterations on the background velocity model and compare the results obtained by direct inversion on primary reflection data only and the "truncated gradient" strategy applied on data containing both primaries and first-order surface multiples.

Finally we study in section section 6.4 an example with an incomplete acquisition where multiples may bring additional information compared to primaries. We want to test the ability of iterative MVA to use this information.

6.2. Robustness of Iterative Migration Velocity Analysis

This section presents and discusses results obtained when a different modelling tool is used for the modelling of synthetic observed data and for inversion. In tests A, B, and C (sections 6.2.2 to 6.2.4), reflections are due to a dense model perturbation (velocity model perturbation or density model perturbation). Before presenting the results of these tests, we show the velocity

updates obtained in the reference case where the same modelling tool is used to compute observed data and calculated data.

6.2.1. Reference case

In the reference case, the reflectivity model is laterally invariant with rapidly varying values between 100 m and 500 m depth, and zero values above and below (figure 6.1, right). Using a Ricker wavelet with maximum frequency 40 Hz, we obtain observed data with several events overlapping with one another (figure 6.2). The first gradient of the MVA objective function is expected to be negative above $z = 500\,\text{m}$ and zero below.

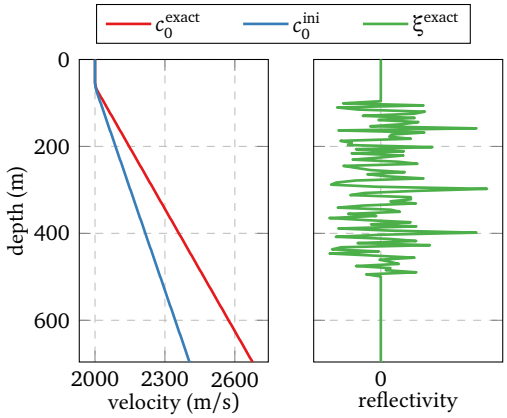


Figure 6.1. – Exact and initial background velocity model (left) and dense exact reflectivity model (right).

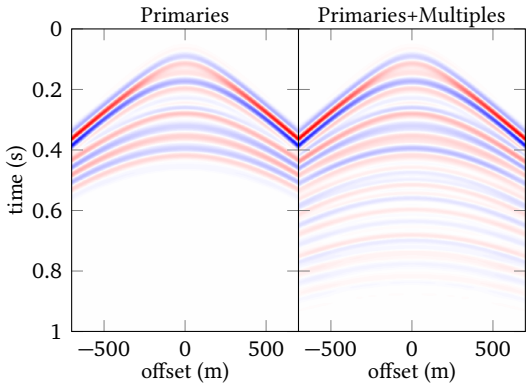


Figure 6.2. – Observed data computed in the model of figure 6.1 under the first-order Born approximation (left) and a second-order Born approximation (right).

We first compute the gradients obtained after direct inversion. In the primaries only case, all events in CIGs have a downward curvature (figure 6.3a). As expected, the gradient is negative above the deeper reflector and zero below. When direct inversion is applied to both primaries and first-order surface multiples, additional cross-talk artefacts appear above and below the deeper reflector with an upward curvature (figure 6.3b). As a consequence positive values appear on the gradient below the deeper reflector as well as strong edge effects.

In the case of primaries only, the combination of iterative migration and filtering with $\mathrm{F}^\dagger\mathrm{F}$ leads to CIGs very similar to those obtained by direct inversion (figure 6.3c). The associated gradient has small non-zero values below the reflector but is quite close to the result shown in figure 6.3a. Finally in the case of multiples, iterative migration succeeds in removing cross-talks artefacts resulting in filtered CIGs very similar to the previous case (figure 6.3d). The associated gradient has a consistent negative value above the deeper reflector and spurious positive values below are attenuated compared to figure 6.3b. These results are satisfactory, and smoothing the gradients yields consistent background velocity updates for subsequent inversion (figures 6.4e and 6.4g). Note also that the "truncated gradients" (figures 6.4e and 6.4g) are very similar to the gradient of figure 6.3a.

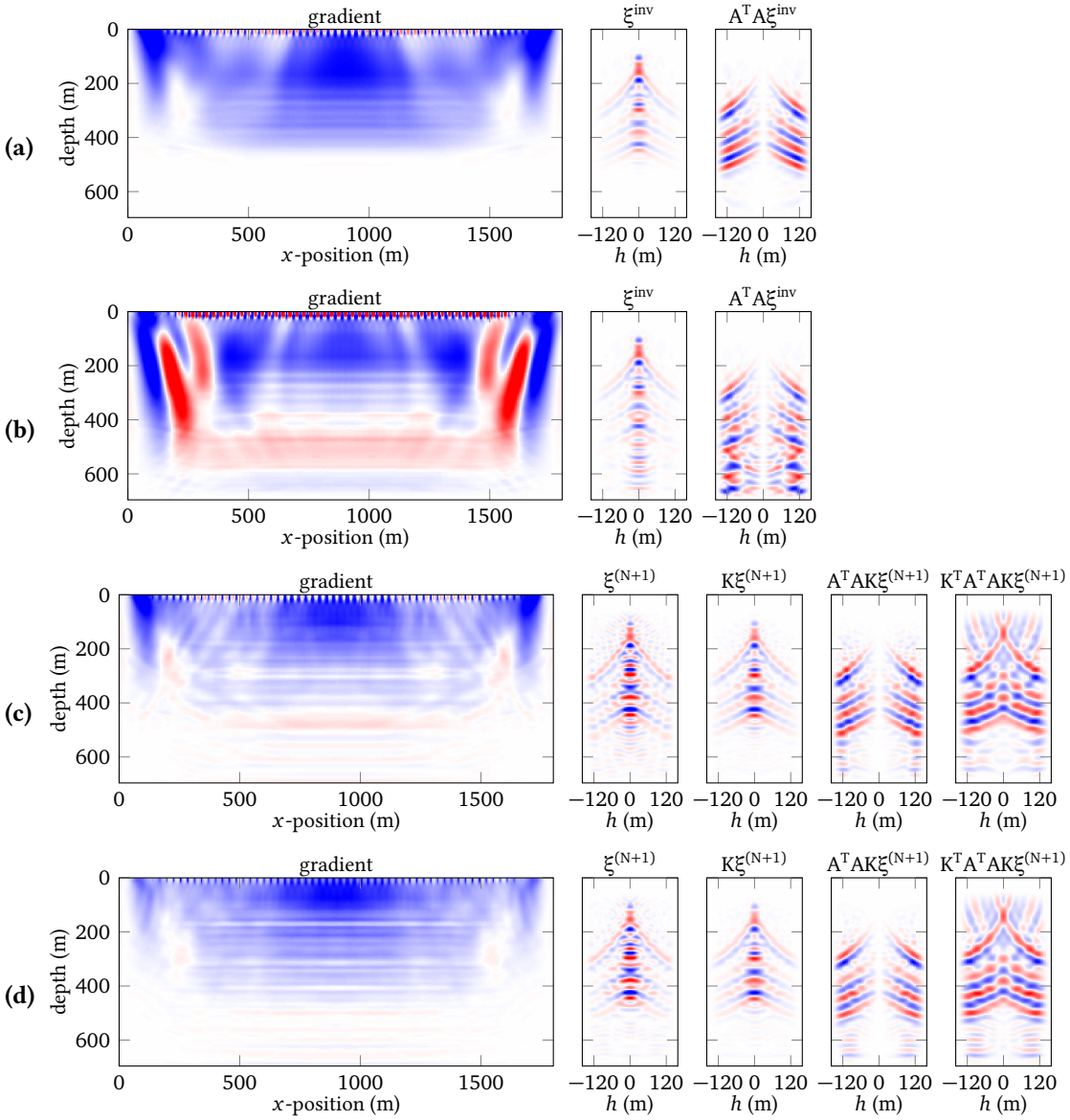


Figure 6.3. – Gradient and CIGs obtained with the model of figure 6.1 with different inversion strategies with and without multiples:

(a) direct inversion on primaries only;
(b) direct inversion on primaries and multiples;
(c) iterative migration on primaries only (N = 7 and M = 10);
(d) iterative migration on primaries and multiples (N = 10 and M = 8).

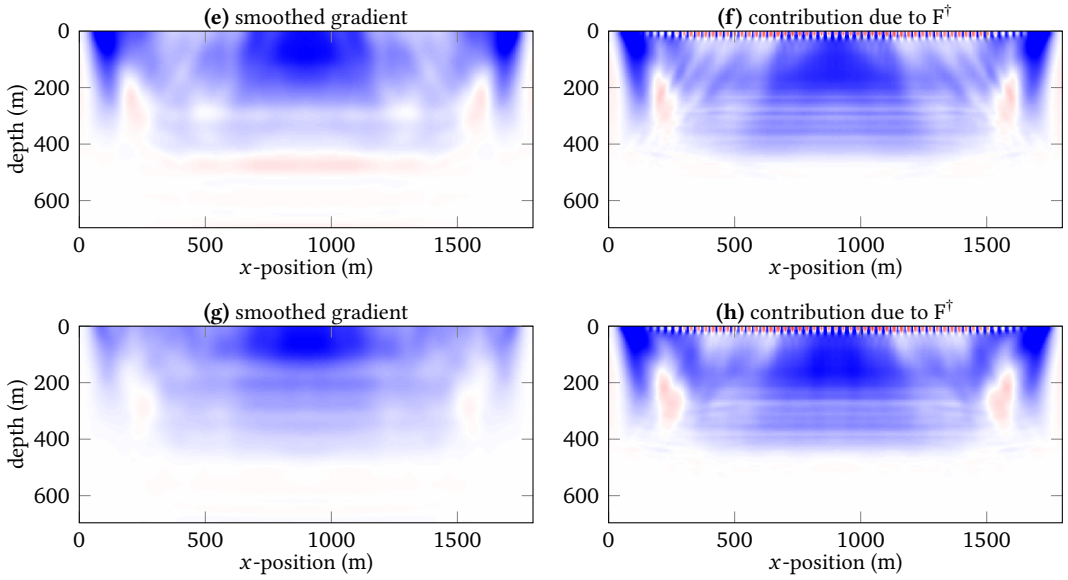


Figure 6.4. – Velocity updates obtained by smoothing the total gradient obtained after iterative inversion (left) and by keeping only the contribution due to the pseudo-inverse (right), in the case of primaries (top), and with both primaries and multiples (bottom).

In summary we obtain results very similar to the case of a single reflector studied in section 5.3.2. However, the gradients presented here are less homogeneous. This can be interpreted as a superposition of the contributions due to the different events in the CIGs, although the gradient obtained with all the events in the CIGs is not formally a linear combination of the gradient obtained from CIGs containing single events. We also observe strong edge effects for all gradients. As already mentioned in chapter 5 (see also figure 5.32), we are mainly interested in the central part of the gradient. With a larger model and the same acquisition parameters, the edge effects would remain similar while the central part would be extended. In these examples we have considered relatively small models to limit the computational cost. We now begin the series of tests summarised in table 6.1 and focus on the central part of the gradients.

6.2.2. Test A – Sensibility to the lack of low frequencies in observed data

In this first test, we use two different source wavelets for observations and inversion. Low frequencies up to 5 Hz are removed from the Ricker wavelet for observations (figure 6.5). The original Ricker wavelet is used during inversion.

In the eight cases (figures 6.6 and 6.7), we obtain results very similar to the reference case (figures 6.3 and 6.4). The main difference lies in the shape of events in CIGs modified in a similar way as the source wavelet is, but this has almost no influence on the shape of the gradient. We conclude that the behaviour of the MVA techniques analysed in the preceding chapters with the standard Ricker wavelet should not be altered by the lack of low-frequencies inherent to real data acquisition. This is a different behaviour than for FWI (Sirgue and Pratt, 2004; Virieux and

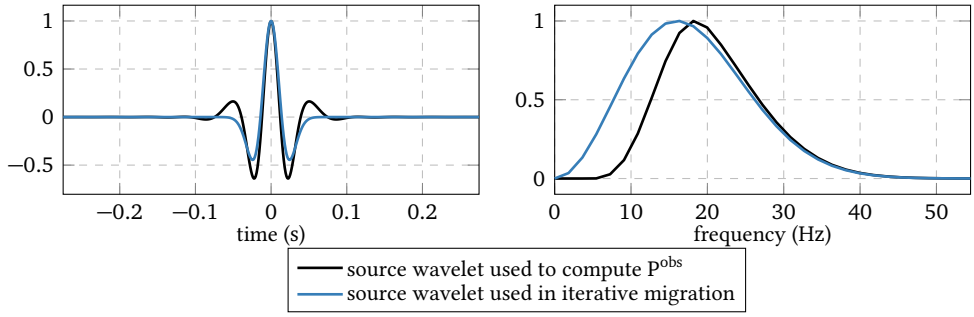


Figure 6.5. – Source wavelet used to generate observed data in test A. It is obtained by setting to zero low frequencies in the original Ricker.

Operto, 2009). To our knowledge, the importance of source wavelet estimation has not been studied in the MVA literature yet.

6.2.3. Test B – Sensibility to inaccuracies of source estimation

In test B, observed data are generated with the same reflectivity model as in the reference, but with a different source wavelet, defined as the derivative of the original Ricker wavelet (figure 6.8). Inversion is still performed with the standard Ricker.

The result of direct inversion is very similar to the reference case when applied to primaries only (figures 6.3a and 6.9a). However multiples have a larger influence than in the reference case (figures 6.3b and 6.9b): cross-talk artefacts in CIGs have larger amplitudes compared to events due to primaries and the associated positive values in the gradient spread over a larger area. This may be related to the stretching of events in CIGs, decreasing with depth (figure 6.10), so that deeper events have lower frequencies. Applying a time derivative to the source wavelet in observed data changes the frequency content of CIGs, and strengthens deeper events. Note that this is only a partial explanation as the approximate inverse should correct for the stretch.

Iterative inversion on primaries only results in CIGs very similar to the direct inversion case. The gradient is however altered by positive values around 100 m depth, which are not attenuated by smoothing (figure 6.11e). Note that the "truncated gradient" (figure 6.11f) is remarkably close to the result of direct inversion (figure 6.9a). In the case of multiples, iterative migration fails to explain both primaries and multiples correctly (figures 6.12 and 6.13). Some multiple reflections are still misinterpreted as primaries after ten iterations (figure 6.12, bottom row, between 0.45 s and 0.6 s). Actually, the errors in the shape of the source are converted into a different reflector shape compared to the reference case. Primaries and multiples interact once and twice with reflectors respectively, hence changing the shape of the reflector can compensate for the source wavelet in the case of primaries, but this result in a wrong phase for predicted multiples. At the first iteration, the reflectivity is optimised for primary reflection, then little improvement is made in the following iterations for both primaries and multiples (figure 6.13, bottom). This results in poor final data residuals and cross-talk artefacts not being attenuated in the final CIGs (figure 6.9d). This has unwanted consequences on the convergence of the adjoint problem (figure 6.14, green curve), which may not be positive definite in the case of multiples as

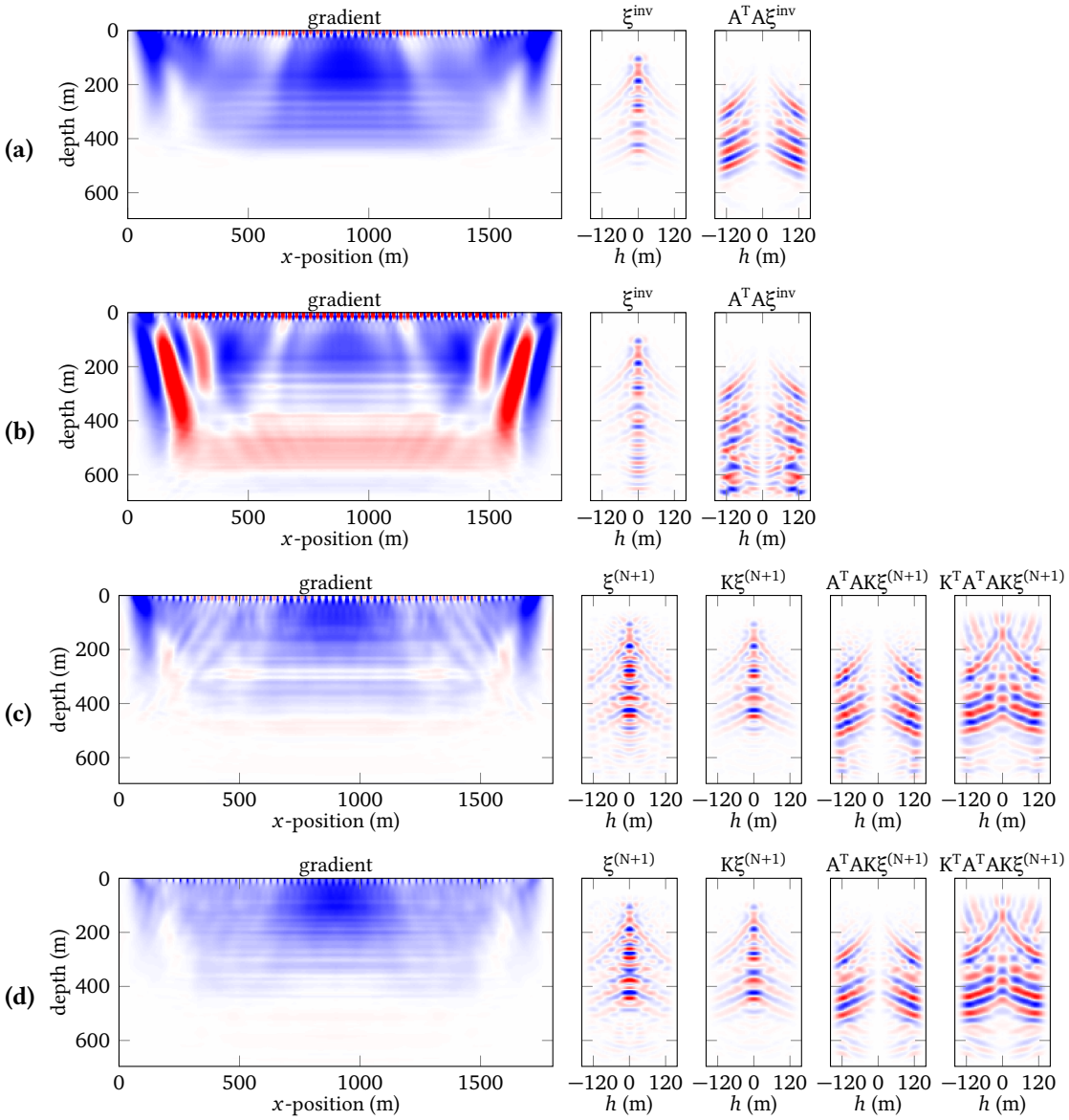


Figure 6.6. – Same as figure 6.3 when observed data are generated with the source wavelet shown in figure 6.5 (test A).

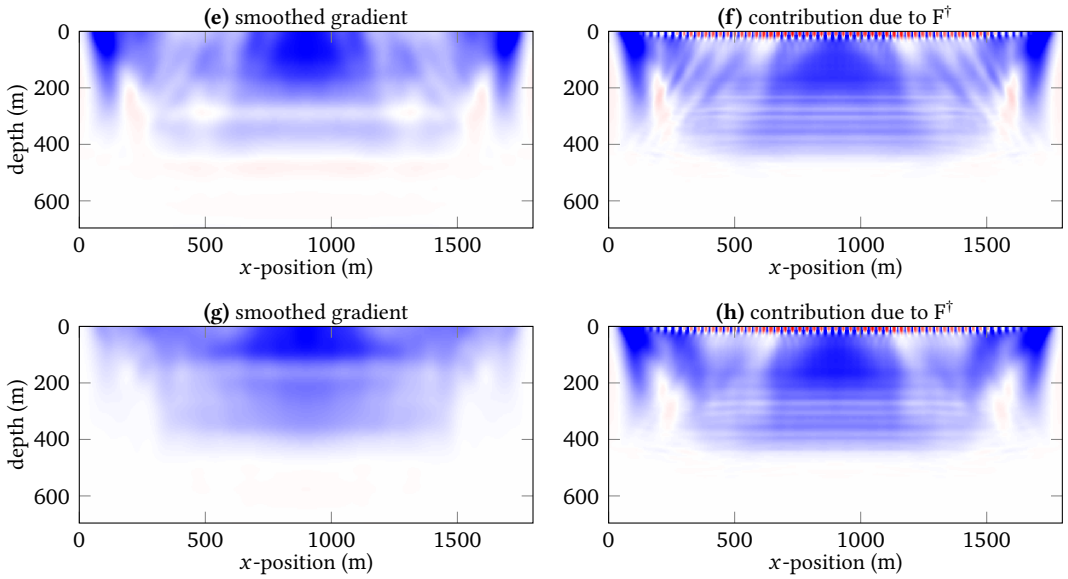


Figure 6.7. – Same as figure 6.4 for the gradients presented in figure 6.6 (test A).

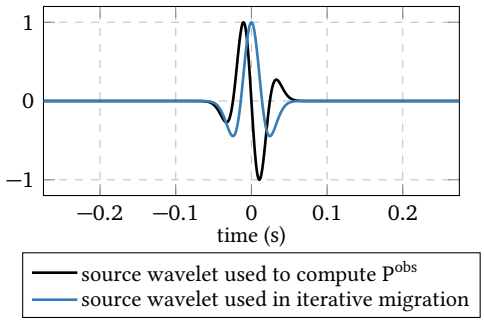


Figure 6.8. – Source wavelet used to generate observed data in test B, obtained as the time derivative of the original Ricker wavelet.

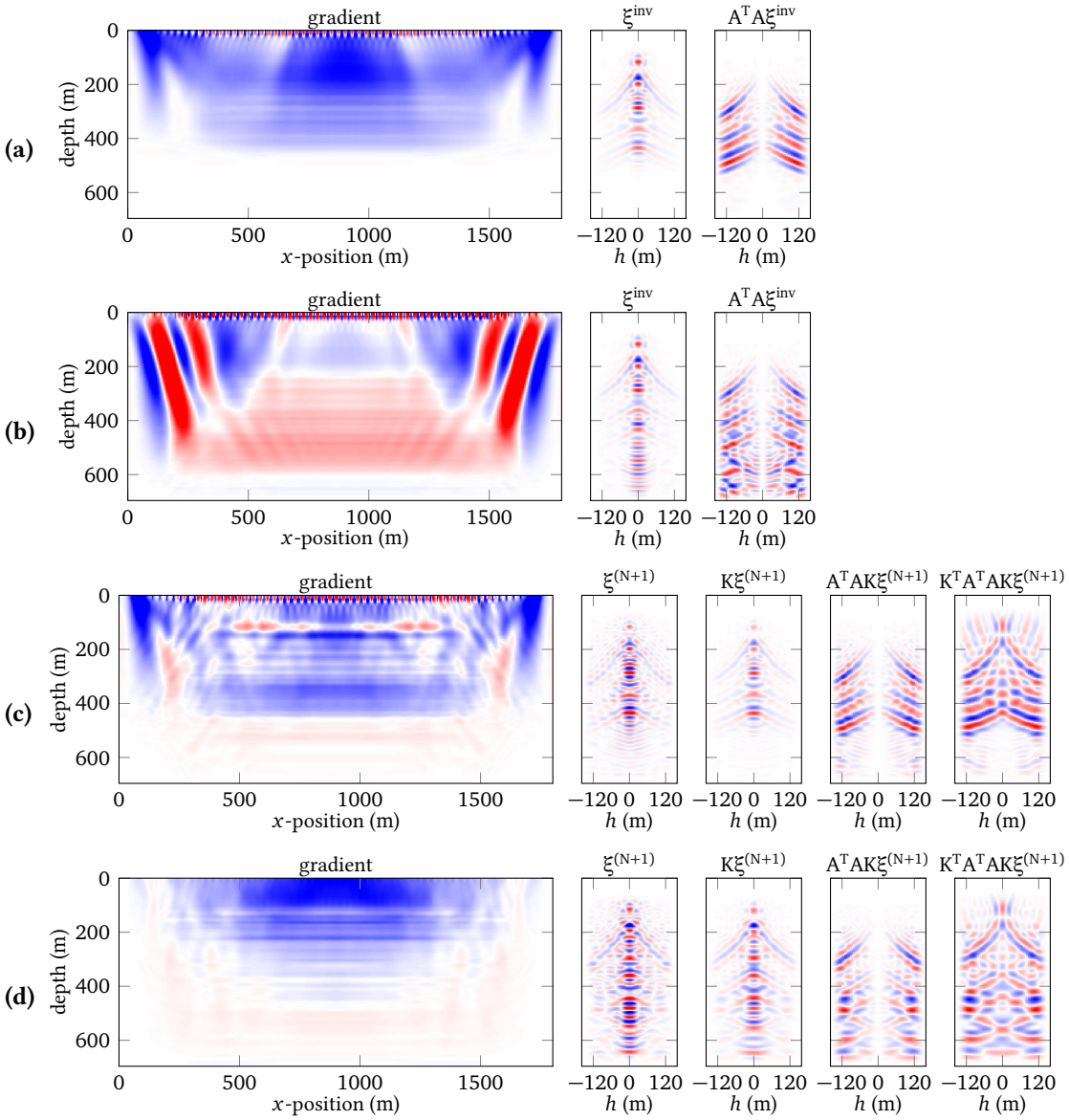


Figure 6.9. – Same as figure 6.3 when observed data are generated with the source wavelet shown in figure 6.8 (test B).

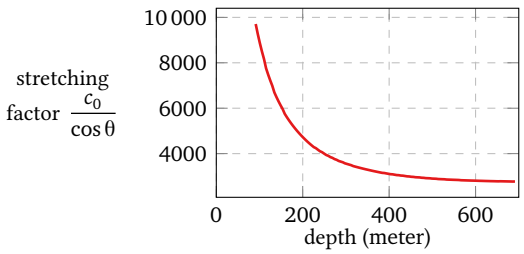


Figure 6.10. – Stretching factor $c_0/\cos\theta$, with θ the half opening angle at the image point, obtained for different depths and the maximum surface offset value $h_{\text{surface}} = 700\,\text{m}$ in the initial velocity model of figure 6.1.

discussed in section 5.3.2. A negative curvature is encountered after eight iterations and we stop the conjugate gradient iterations. Although the norm of the objective function associated with the CG-algorithm still decreases, the relative normal residuals increases from iteration 5, which is not satisfactory. A larger regularisation weight may mitigate this issue, but at the expense of a possible higher final data misfit. The gradient computed after eight iterations of resolution of the adjoint problem looks consistent (figure 6.9d) but its value and signs are actually not stable from one iteration to another. Note however that the "truncated gradient" (figure 6.11h) which does not depend on the number of adjoint iterations, shows much improvement compared to figure 6.9b and therefore stands as a reliable alternative.

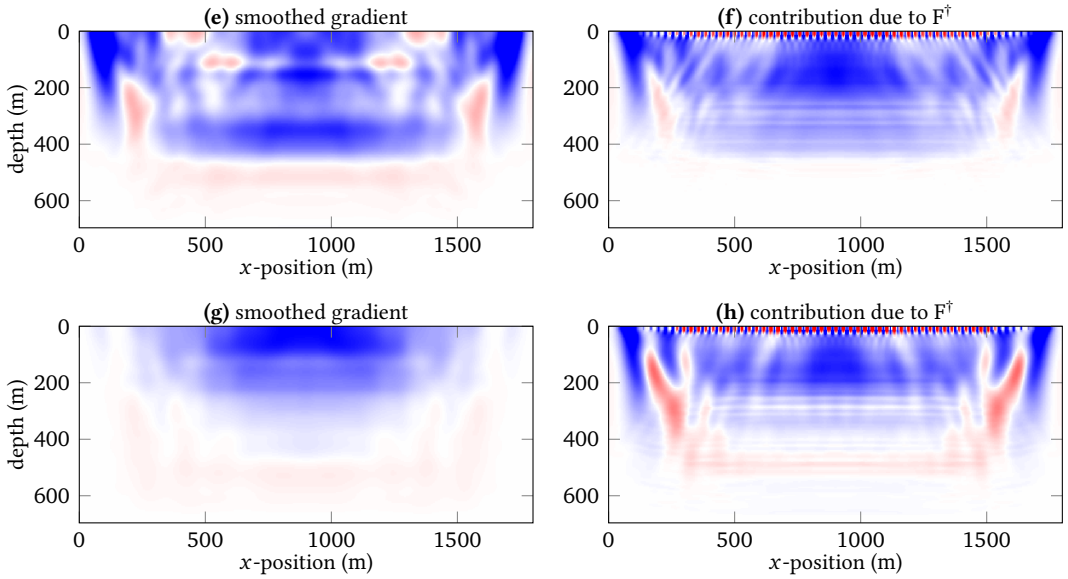


Figure 6.11. – Same as figure 6.4 for the gradients presented in figure 6.9 (test B).

This test indicates that the estimation of the source wavelet is a key point for the accuracy of iterative migration. A poor estimate prevents the attenuation of cross-talk artefacts and leads to inconsistent background velocity updates.

6.2.4. Test C – Density perturbation

We now consider a new exact model for the modelling of observed data. The exact (and initial) background velocity model are the same as those used in the reference case (figure 6.1, left) but here with a zero velocity perturbation. Instead we assume that reflections originate only from a variable density model $\rho(\mathbf{x})$, modifying the wave-equation (2.4) into

$$\frac{(i\omega)^2}{c^2}P - \rho\nabla\cdot\left(\frac{1}{\rho}\nabla P\right) = \Omega(\omega)\delta(\mathbf{x}-s). \tag{6.5}$$

For this test, we will decompose the density model $\rho(\mathbf{x})$ into $\rho(\mathbf{x}) = \rho_0(\mathbf{x}) + \delta\rho(\mathbf{x})$ and model data under a second-order Born approximation. Similar to the velocity perturbation case

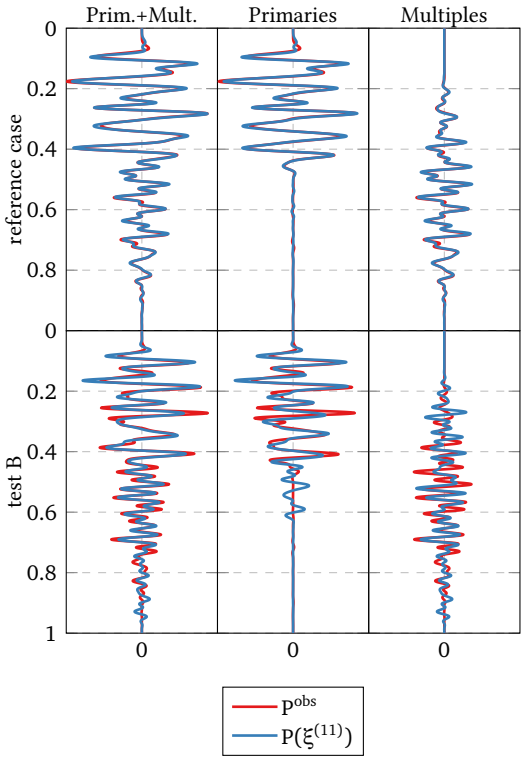


Figure 6.12. – Observed data (containing both primaries and multiples) and calculated data obtained after ten iterations of iterative migration in the reference case (top, figure 6.3d), and in test B (bottom, figure 6.9d).

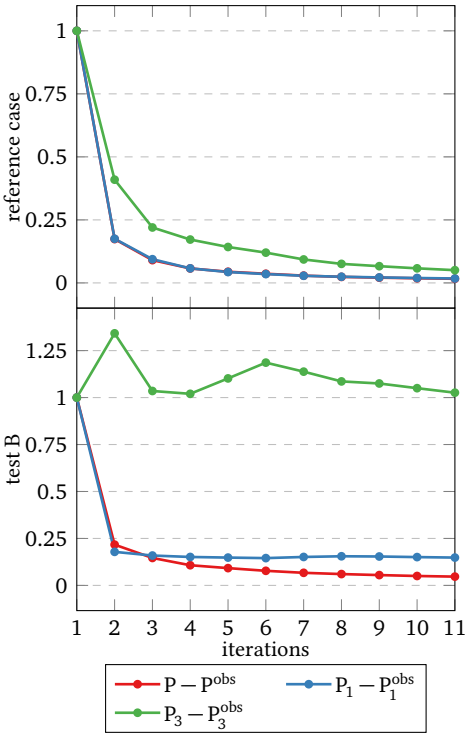


Figure 6.13. – Migration objective function (red) obtained in the reference case (top) and in test B (bottom). We also compute the data misfit of primaries alone and multiples alone (blue and green, respectively). Note that the red and blue curves in the reference case are almost superimposed.

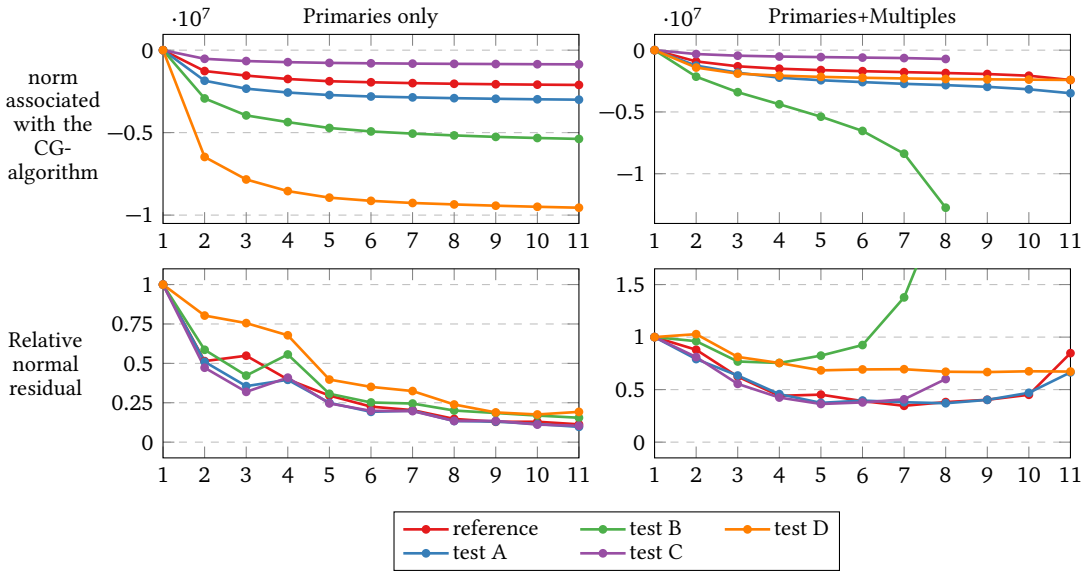


Figure 6.14. – Norm associated with the conjugate gradient algorithm (top) and relative normal residuals (bottom) obtained across iterations for the resolution of the adjoint problem in the reference case as well as in the tests A, B, C and D both in the case of primaries only (left), and in the case of primaries and multiples (right). The five case shown here correspond to different linear systems and cannot be directly compared. Note that the relative normal residuals are actually the norm of the gradient of the CG-objective and may not decrease from one iteration to the following.

(section 2.3.1), we decompose P into $P = P_0 + \delta P$ and obtain

$$\frac{(i\omega)^2}{c^2}(P_0 + \delta P) - (\rho_0 + \delta\rho)\nabla \cdot \left[\frac{1}{\rho_0 + \delta\rho}\nabla(P_0 + \delta P)\right] = \Omega(\omega)\delta(\mathbf{x} - s), \tag{6.6}$$

$$\frac{(i\omega)^2}{c^2}(P_0 + \delta P) - (\rho_0 + \delta\rho)\nabla \cdot \left[\left(\frac{1}{\rho_0} - \frac{\delta\rho}{\rho_0^2}\right)\nabla(P_0 + \delta P)\right] = \Omega(\omega)\delta(\mathbf{x} - s), \tag{6.7}$$

leading to

$$\frac{(i\omega)^2}{c^2}\delta P - \rho_0\nabla \cdot \left[\frac{1}{\rho_0}\nabla\delta P\right] = -\rho_0\nabla\left(\frac{\delta\rho}{\rho_0^2}\right) \cdot \nabla P_0 + \delta\rho\nabla\left(\frac{1}{\rho_0}\right) \cdot \nabla P_0, \tag{6.8}$$

Assuming a homogeneous background density model $\rho_0(\mathbf{x}) = \rho_0$, this expression simplifies into

$$\frac{(i\omega)^2}{c^2}\delta P - \Delta\delta P = -\frac{1}{\rho_0}\nabla\delta\rho \cdot \nabla P_0 \tag{6.9}$$

In test C, we set $\rho_0 = 1$, and we consider a laterally invariant density perturbation with the same shape as the velocity perturbation defined in the reference case (figure 6.1, right). The magnitude of the perturbation is determined such that the amplitude of reflection data at zero surface offset is the same as in the reference case (figure 6.15). Moving from a velocity to a density perturbation does not change the kinematics of primary and multiples reflections. The main difference is that amplitudes are lower for wide opening angles, corresponding to large surface offsets of early primary reflections due to the shallower reflectors (figure 6.15, right).

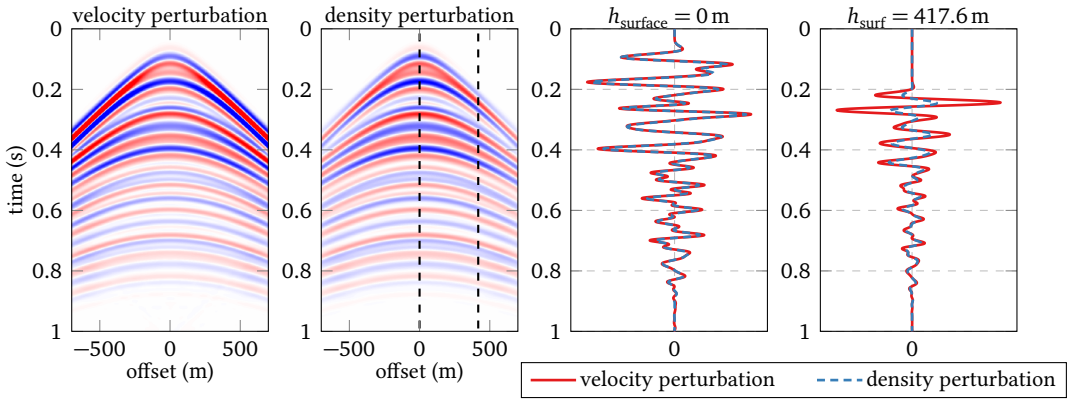


Figure 6.15. – Observed data obtained with a perturbation of the velocity model (1st column) and with a perturbation of the density model (2nd column). Both plots use the same colour scale. Traces extracted at the position indicated by the dashed lines are shown on the two most right panels.

This difference of amplitudes has a direct impact on CIGs obtained by direct inversion. Compared to the reference case (figure 6.3a), the shallower reflectors have much lower amplitudes at large value of the subsurface-offset, so that only the defocusing of the deeper reflectors is

interpreted in the velocity analysis (figure 6.16a). The gradient looks more homogeneous but this is due only to the attenuation of shallower reflectors. The same effect appears when multiples are added to observed data. Due to smaller opening angles than primaries, their amplitudes in observed data remain similar to the reference case. As a consequence they have higher influence on velocity analysis: the amplitude of positive energy below the deeper reflector is relatively stronger than the above negative values due to primaries (figure 6.16b).

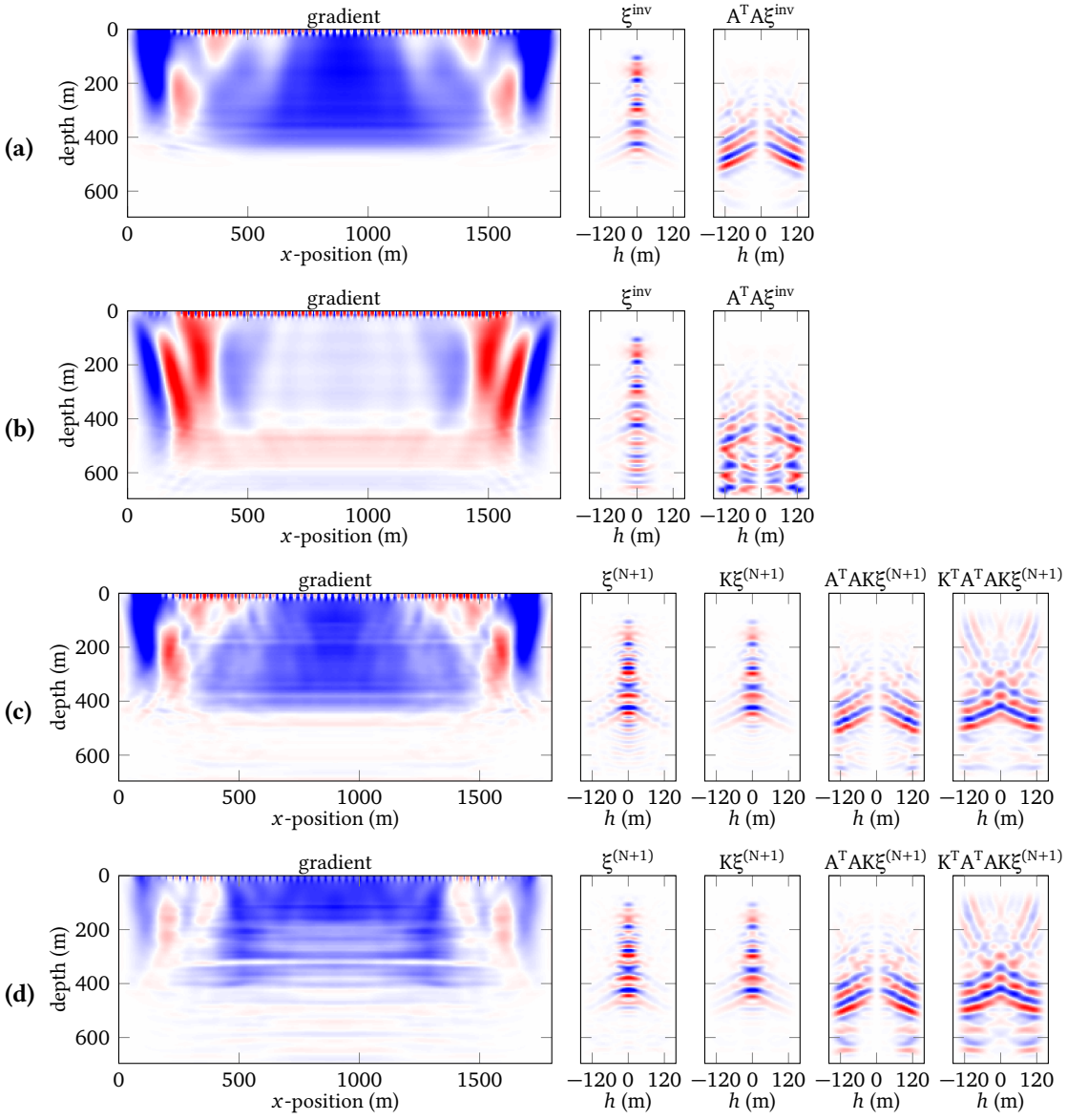


Figure 6.16. – Same as figure 6.3 when observed data are generated with a density perturbation instead of a velocity perturbation (test C).

After iterative migration and filtering, we obtain CIGs very similar to the result shown in

figure 6.16a, both in the case of primaries only and in the case of primaries and multiples (figures 6.16c and 6.16d). The final reflectivity correctly explains both primaries and multiples (figure 6.17) although reflections are modelled with a velocity perturbation for the determination of the optimal model perturbation. Compared to the reference case (figures 6.12 and 6.13, top), the final misfit is slightly higher for multiples (figure 6.18), but this does not prevent cross-talk artefacts from being attenuated. The associated gradients are very close to the one obtained in figure 6.16a, especially after smoothing (figures 6.19e and 6.19g). Again, the truncated gradients (figures 6.19f and 6.19h) are remarkably close to the velocity update obtained in the direct inversion case, meaning that the reflectivity image obtained after iterative migration allows to faithfully reproduce the kinematics of primary reflections.

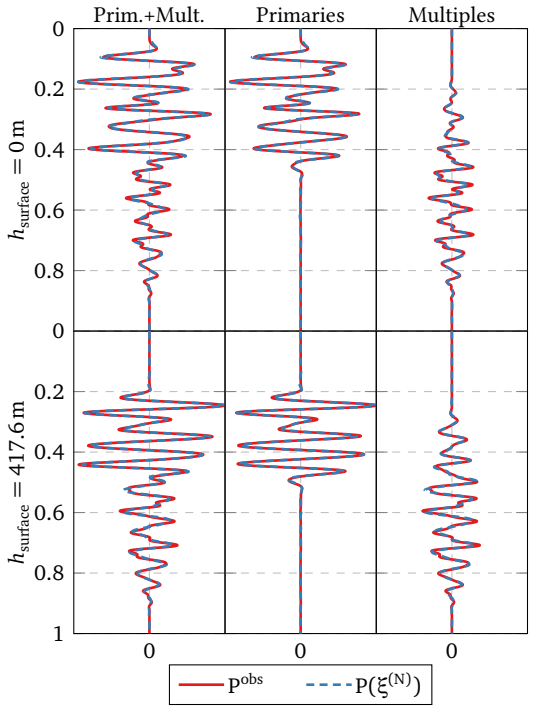


Figure 6.17. – Observed data and calculated data obtained after ten iterations of iterative migration for two values of the surface offset, $h = 0$ m (top) and $h = 417.6$ m (bottom) (test C).

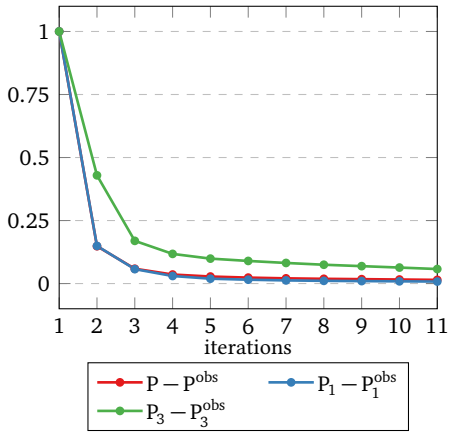


Figure 6.18. – Migration objective function (red) obtained when observed data are modelled with a density model perturbation instead of a velocity model perturbation (test C).

We conclude from test C that our approach correctly extracts the kinematic information contained in reflections originating from density perturbations. Iterative migration succeeds in finding a velocity perturbation model explaining correctly both primaries and multiples reflections. The amplitude difference, especially for wide opening angles, between reflections originating from density and velocity perturbation is compensated by a modification of the value of the reflectivity model at large values of h. The MVA gradient is still consistent, the difference being that a smaller weight is given in the background velocity update to reflections

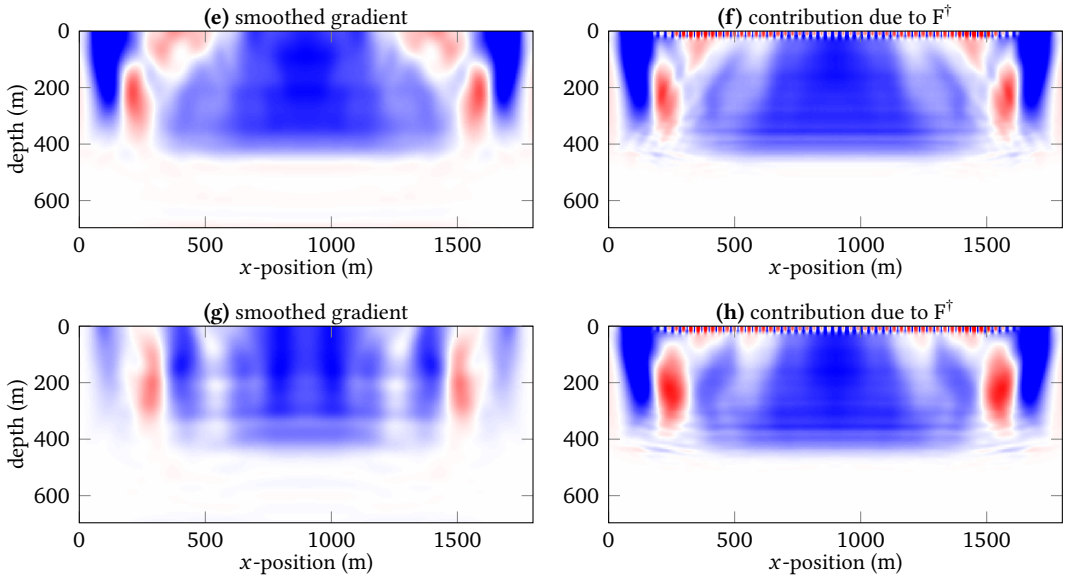


Figure 6.19. – Same as figure 6.4 for the gradients presented in figure 6.16 (test C).

originating from shallow reflectors.

6.2.5. Test D – Observed data modelled without the Born approximation

Finally we test the ability of our approach to deal with observed data modelled without the Born approximation. We use an exact velocity model different from tests A, B and C but with the same dimensions and we keep the same acquisition parameters. The velocity increases with depth, similar to the previous case, but instead of a dense reflectivity, we add two layers with homogeneous velocity (2600 m/s) between 200 m and 400 m depth and 550 m and 600 m depth; the density model is variable with interfaces located at the same positions as the velocity discontinuities (figure 6.20, red curves). Both models are laterally invariant and the velocity perturbations are not proportional to the density perturbations.

Observed data are modelled with a finite-difference variable density acoustic propagation code. In the case of primaries only, PMLs are implemented on each edge of the model while a free-surface condition is used to model multiples (figure 6.21, 1st column). The main effect compared to second-order Born approximation are the following:

- an angle dependent reflectivity coefficient at the free surface;
- ghost effects at the source and receiver positions;
- all orders of multiples are modelled, as well as internal multiples.

For comparison, we smooth the velocity and density models and define associated velocity and density perturbations (figure 6.20, blue curves) to model data under a second-order Born approximation (figure 6.21, 2nd column). To take ghosts effects into account, we use a scaled second-order time derivative of the Ricker wavelet in the case of multiples (figure 6.22).

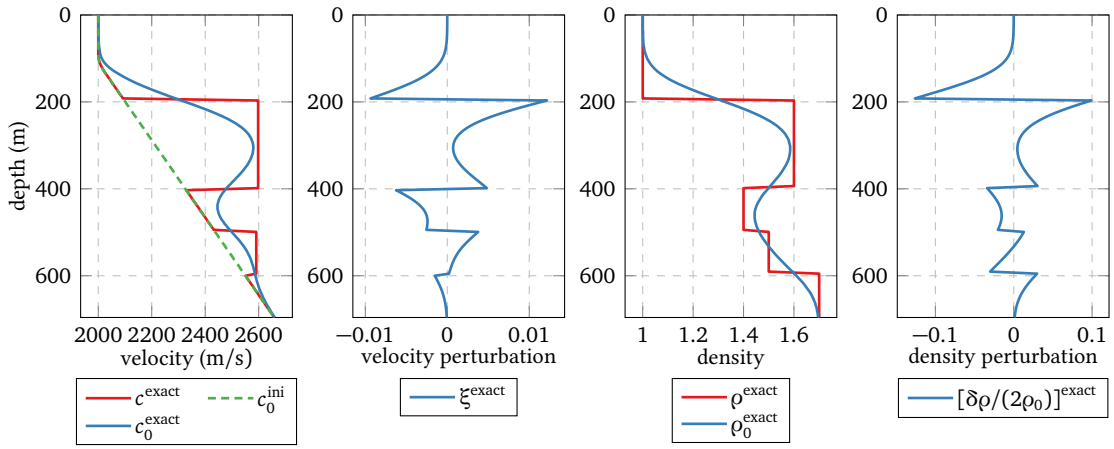


Figure 6.20. – Exact discontinuous velocity and density 1D profiles used to model observed data (red, first and third columns). The blue curves correspond to smoothed velocity and density models, and associated velocity and density perturbations used to model data with second-order Born approximation. The dashed green line is the initial velocity model used to compute the gradient (test D).

In the primaries only example, both data sets are very similar. Amplitudes are overestimated by the Born approximation, but with consistent kinematics and a similar amplitude ratio at zero and large offsets (figure 6.21, top row). In the case of multiples, the kinematics is not as well reproduced by the Born approximation as in the previous case, but remains consistent. However the amplitudes are overestimated with much larger discrepancies at large offsets than at zero offsets. The origin of these errors is the use of a second-order time derivative to model ghost reflections, valid at zero-offset only.

We now compute the background velocity updates using the observed data sets obtained without the Born approximation (figure 6.21, 1st column). For calculated data, the density model is assumed homogeneous and the initial velocity model is increasing with depth (figure 6.20, dashed green curve). Direct inversion on primaries data yields CIGs with downward curved events only, which is consistent with the initial too low velocity model (figure 6.23a). The central part of the associated gradient is smooth and homogeneous with negative values as expected. In the multiple case, we use the modified source wavelet introduced in the preceding paragraph to model ghost reflections for the inversion (figure 6.22). Cross-talk artefacts with an upward curvature are superimposed on the events related to primaries, resulting in spurious positive values in the gradient (figure 6.23b).

Iterative inversion in the primaries only case yields CIGs similar to direct inversion (figure 6.23c). The gradient is negative but exhibits oscillations around reflector positions, which can be nonetheless attenuated by smoothing (figure 6.24e). Again, the “truncated gradient” (figure 6.24f) is very close to the result shown in figure 6.23a. This contrasts with the mixed results obtained with multiples. Cross-talk artefacts are still visible on CIGs (figure 6.23d), although final data residuals are very small (figures 6.25 and 6.26). Our interpretation of this apparent contradiction is that the final reflectivity contains events with downward curvature

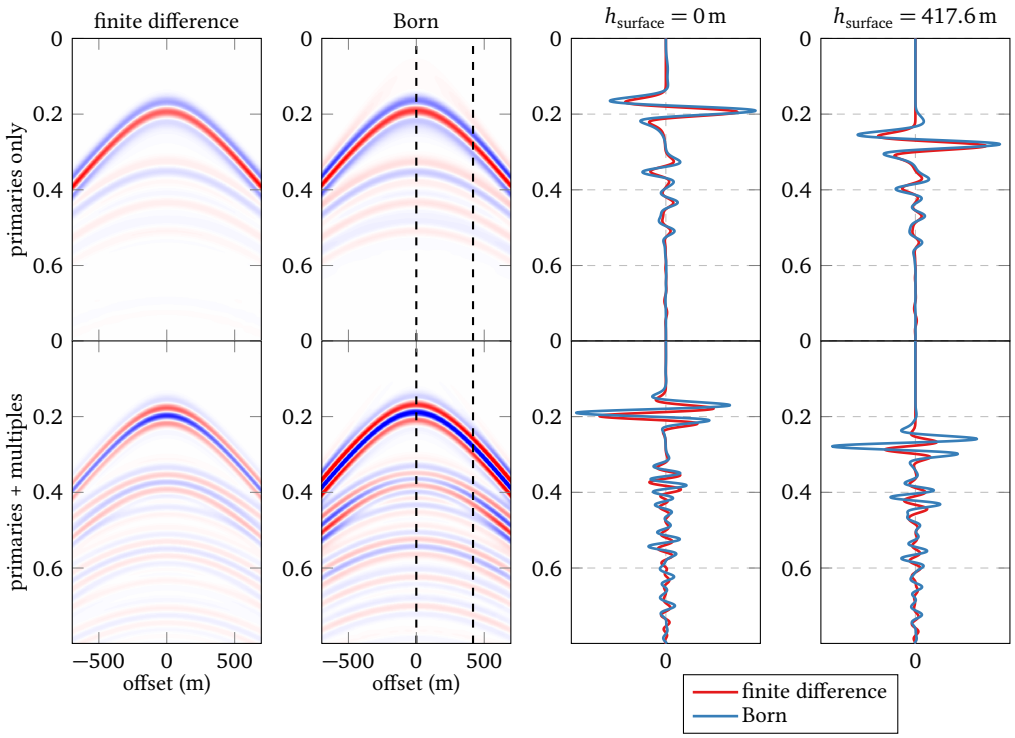


Figure 6.21. – Observed data computed with a finite-difference modelling code and a free-surface condition in the multiple case (left column, corresponding to the red 1D profiles of figure 6.20). We compare this data set with data modelled under the second-order Born approximation (2nd column, corresponding to the blue 1D profiles of figure 6.20) (test D).

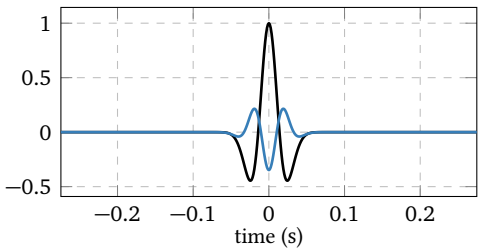


Figure 6.22. – Ricker source wavelet with maximum frequency of 40 Hz used to model observed data with a finite difference scheme (black). A second-order time derivative of this wavelet is used to take source and receiver ghost effects into account when modelling data under the second-order Born approximation (test D).

corresponding to the true reflectors which correctly explain primary reflections. However the amplitudes of multiples modelled from these reflectors do not match those in observed data, because of the inaccuracy of both the second-order Born approximation and the modelling of ghosts used here. The difference is explained by residual cross-talk artefacts, which are nonetheless smaller than in figure 6.23b. As a consequence, spurious positive values are still visible in the gradient (figure 6.23d). They are not attenuated after smoothing (figure 6.24g). Note also that the convergence of the adjoint problem is slower than in the previous cases (figure 6.14, orange curve). As in the previous example, we observe that the truncated gradient provides more consistent velocity updates and seems more insensitive to residual cross-talk artefacts in CIGs (figure 6.24h).

We conclude from test D that iterative migration has difficulty in presence of a free-surface, because source and receiver ghosts are not properly modelled. We used a second-order time derivative of the source wavelet to account for this effect, but this approximation is valid for zero reflection angles only. A better modelling tool should be investigated to deal with this issue. As for the preceding examples, the best result is obtained here with the "truncated gradient" approach.

6.3. Iterations on the background velocity model

We now consider a 2D model with lateral variations and iterate on the background velocity model. The model is 2500 m large and is made of three 1D velocity profiles similar to the previous example with two sub-vertical faults and a low-velocity layer between 400 m and 550 m depth in the central part of the model (figure 6.27, top). We also consider a variable density model with interfaces located at the same positions as in the velocity model, but with different contrast values.

We have seen in the preceding section that free-surface remains an issue as we cannot model properly ghost reflections at wide angles. Therefore we consider here a simpler case and model observed data under a second-order Born approximation. For that purpose we use smoothed versions of the velocity and density models (figure 6.27, bottom). From now on, "exact velocity model" refers to this smoothed version. Although the forward modelling code for observations and inversion both use the second-order Born approximation, we are not in the position of an "inverse crime". Observed data are indeed modelled with variable density and both velocity and density perturbations, the density perturbations being stronger. For inversion, we assume a constant density model ($\rho_0 = 1$) and only model reflections due to velocity perturbations.

Starting with an homogeneous (2000 m/s) background velocity model as initial guess, the inversion aims at minimising the following MVA objective function

$$J_1[c_0] = \frac{\left\| h \cdot z \cdot c_0^{\beta} \cdot \xi[c_0](z,x,h) \right\|_{\mathscr{E}}^2}{\left\| z \cdot c_0^{\beta} \cdot \xi[c_0](z,x,h) \right\|_{\mathscr{E}}^2}. \tag{6.10}$$

To prevent convergence towards background velocity models minimising the energy in CIGs instead of minimising only defocused energy, we consider an objective function normalised by the energy of the reflectivity image without multiplication by h (Chauris and Noble, 2001). As

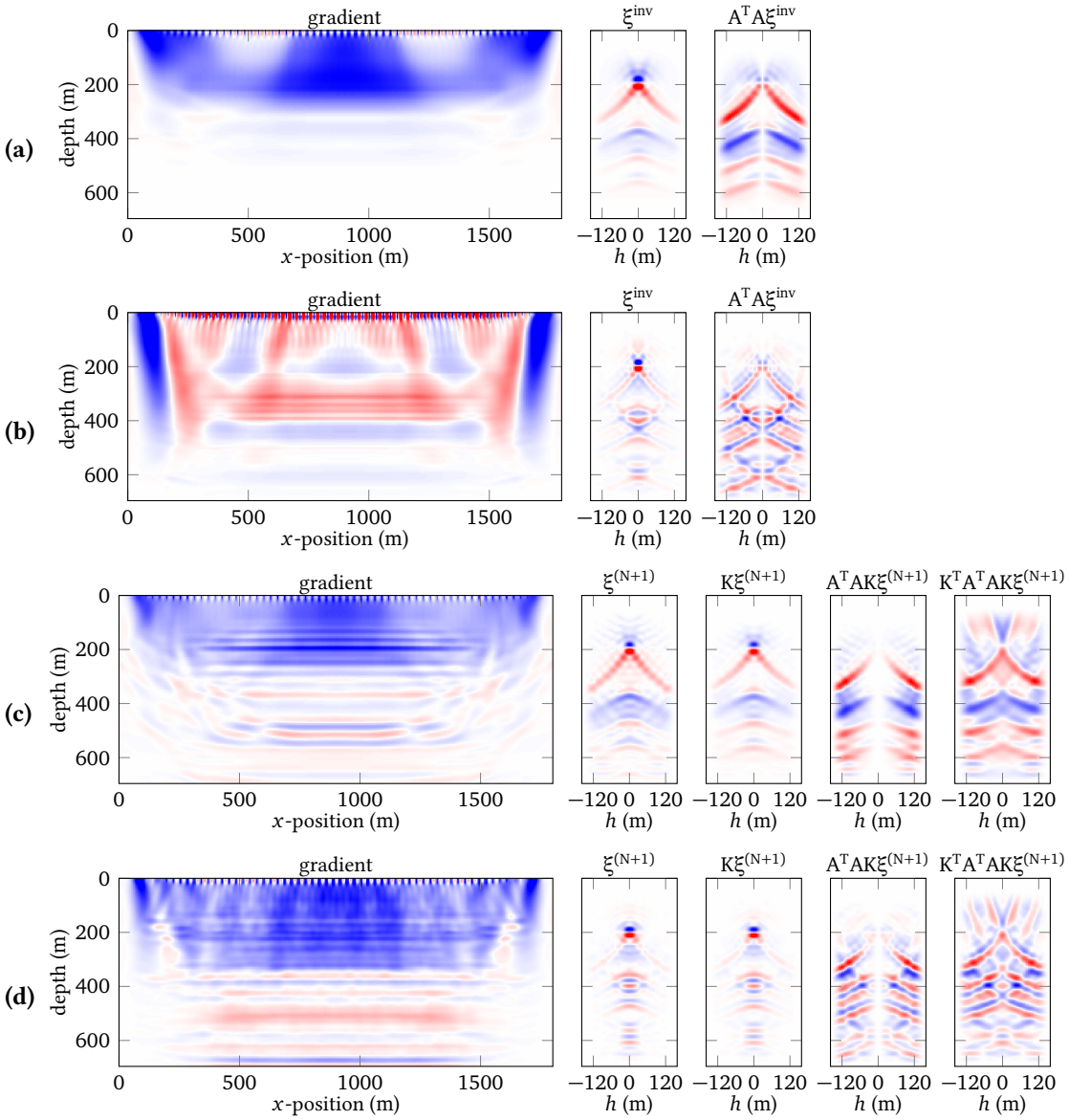


Figure 6.23. – Same as figure 6.3 when observed data are generated with a finite-difference modelling code without Born approximation (test D).

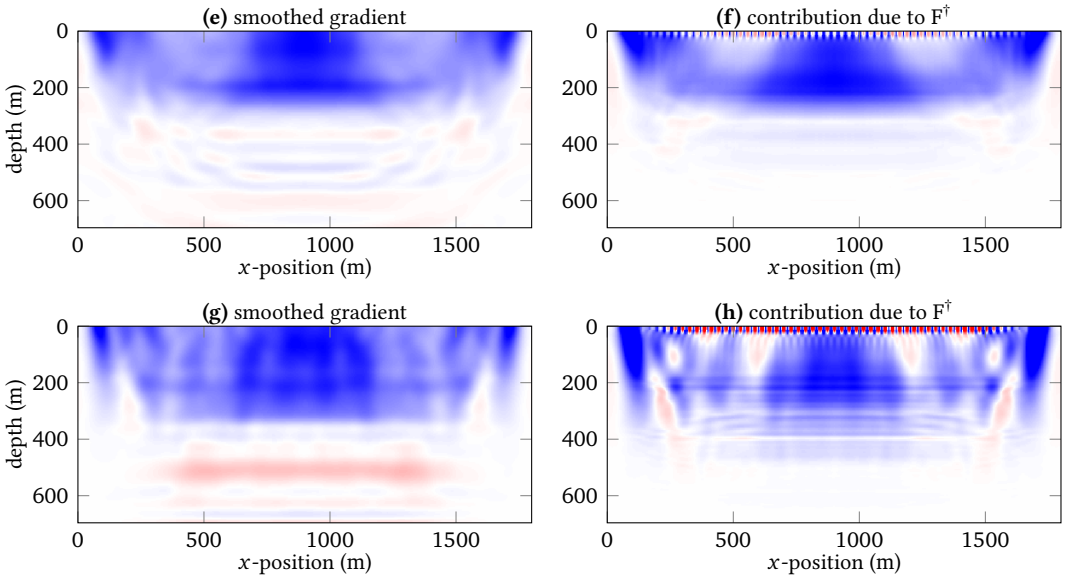


Figure 6.24. – Same as figure 6.4 for the gradients presented in figure 6.23 (test D).

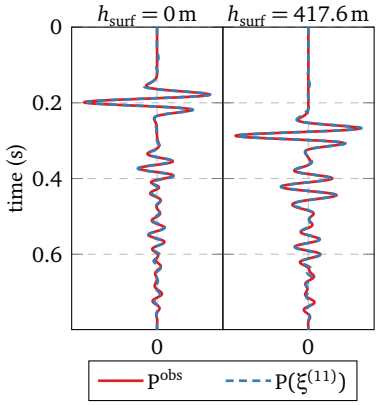


Figure 6.25. – Residuals obtained at zero offset (left) and large offset (right) after 10 iterations of iterative migration (test D).

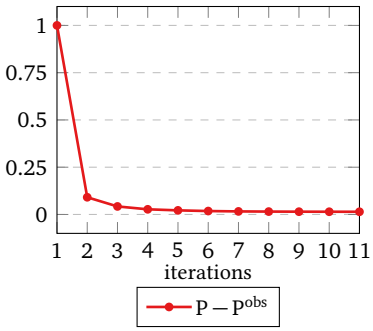


Figure 6.26. – Value of the data misfit across iterations (test D).

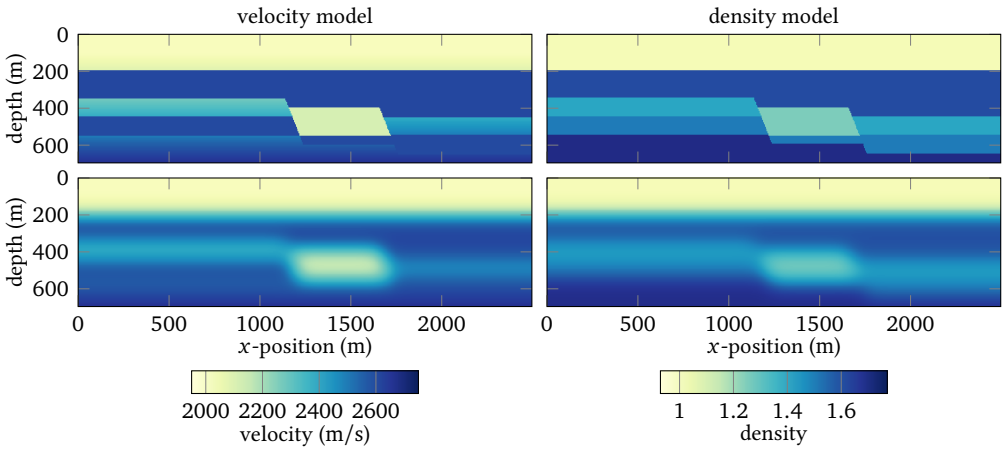


Figure 6.27. – Exact velocity and density model (top), and their respective smoothed versions (bottom) used to generate observed data with a second-order Born approximation.

the shallower interface is a strong reflector, a multiplication by z is introduced in the annihilator, so that deeper reflectors have enough influence on the velocity model reconstruction. Besides, as velocities are underestimated in the initial velocity model, reflectors are migrated to shallower depths, and should be shifted toward deeper positions across iterations, which conflicts with the z-multiplication in CIGs. Therefore a multiplication by z is also introduced in the normalisation term of the objective function (6.10).

We first use the direct inversion strategy on primaries only and perform 20 iterations of non-linear conjugate gradient with the Polak-Ribière formula (figures 6.28 and 6.29). To accelerate the convergence, we assume that the velocity in the first layer is known and set to zero the gradient below 100 m depth. We also introduce a preconditioner consisting of a multiplication by the depth z^α with $\alpha = 1$ in the applications. Finally we smooth the resulting velocity update with a gaussian blur, with a stronger smoothing in the x-direction ($\sigma_x = 140\,\text{m}$ and $\sigma_z = 90\,\text{m}$). These values are progressively decreased every four iterations to recover more detailed structure in the last iterations, as shown in table 6.2.

iterations	σ_z	σ_x
1 to 4	$\sigma_{z0} = 90\,\text{m}$	$\sigma_{x0} = 140\,\text{m}$
5 to 8	$0.75 \cdot \sigma_{z0}$	$0.75 \cdot \sigma_{x0}$
9 to 12	$0.5 \cdot \sigma_{z0}$	$0.5 \cdot \sigma_{x0}$
13 to 20	$0.25 \cdot \sigma_{z0}$	$0.25 \cdot \sigma_{x0}$

Table 6.2. – Parameters of the gaussian blur used to smooth the gradient at each iteration on the background velocity model.

The most energetic event in the initial CIGs is the continuous interface located at 2000 m

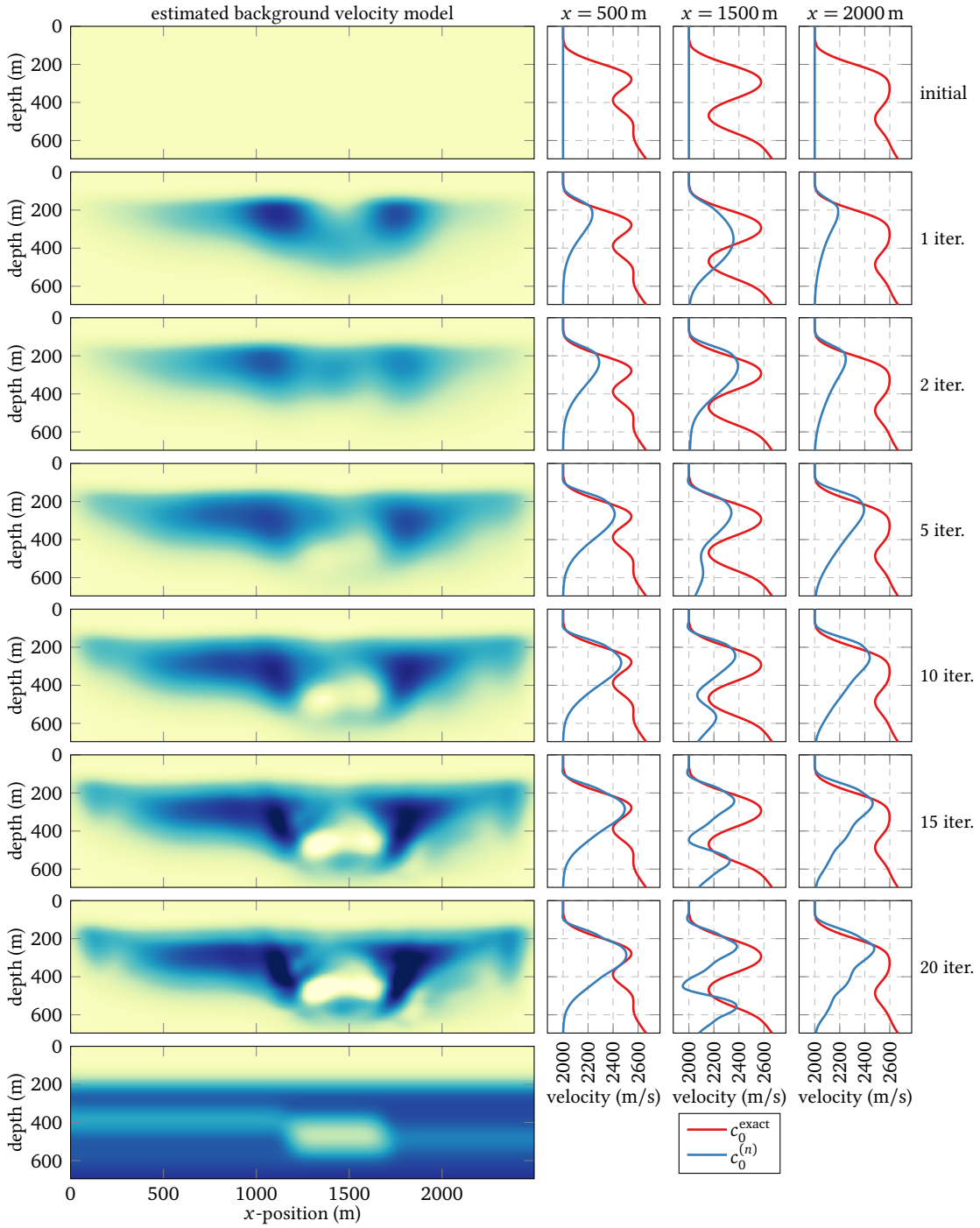


Figure 6.28. – Result of iterations on the background velocity model with the direct inversion strategy on data containing primaries only. The colour bar ranges from 1950 m/s to 2750 m/s and is similar to the one used in the left column of figure 6.27.

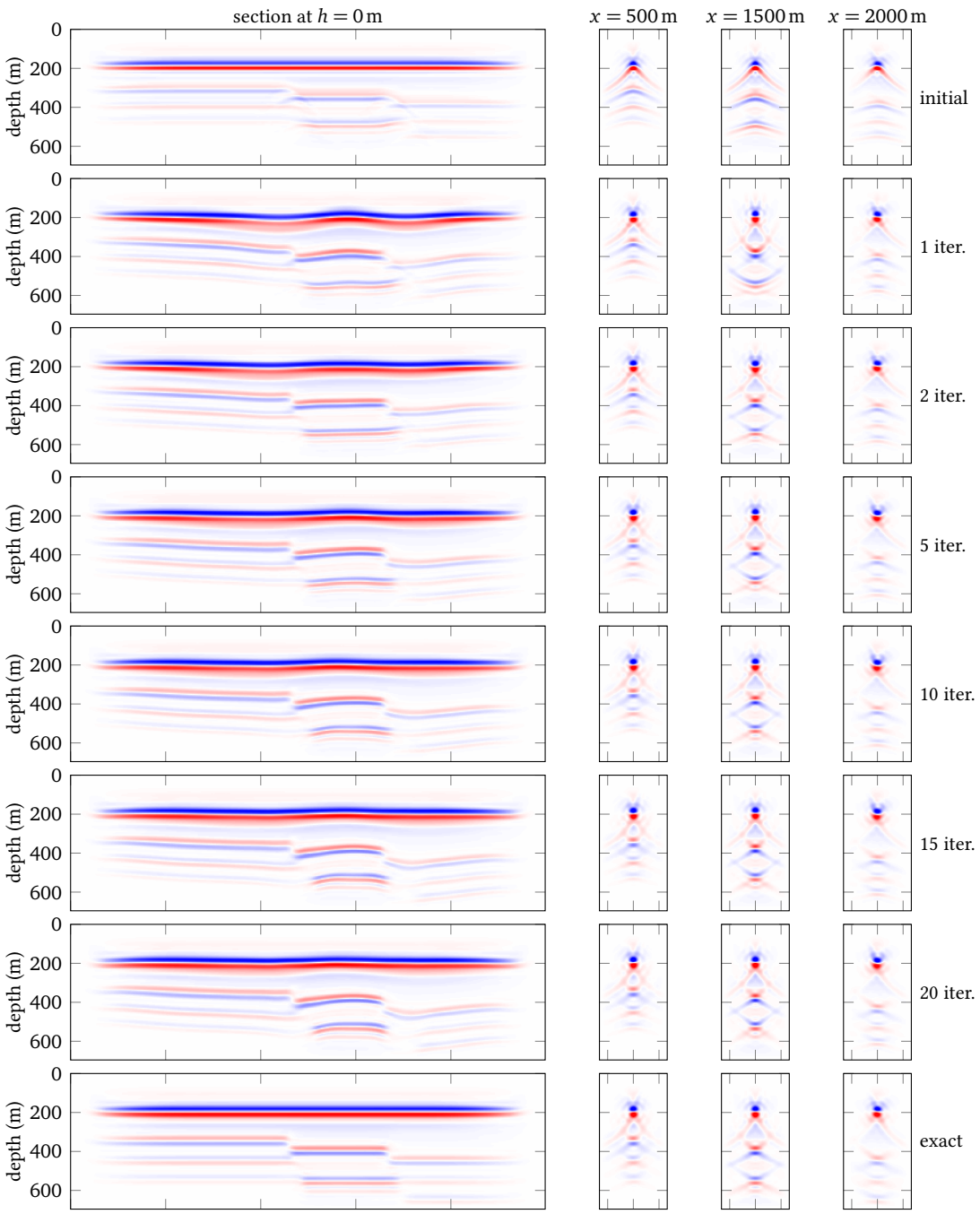


Figure 6.29. – Reflectivity image obtained by direct inversion in the background velocity models shown in figure 6.28. The bottom row displays the results obtained in the exact velocity model. The same colour scale is used for all sections at $h = 0$ (left column). All CIGs are also plotted with the same colour scale, which has been truncated for a better representation of deeper reflectors.

depth (figure 6.29, top row). The associated defocused energy drives the first velocity updates and the shallower velocity contrast is recovered after a couple of iterations (figure 6.28). Subsequent iterations address the focusing of events related to the low velocity anomaly at larger depths, progressively appearing between iterations 5 and 15. After 9 iterations, the value of the objective function has decreased to the value obtained in the exact model (figure 6.30, left). Smaller values are reached by further attenuating defocused energy but this does not necessarily result in more relevant information on the background velocity model. In particular, velocity updates after iteration 15 are mostly located around the faults to the detriment of the low velocity layers on the left and right sides of the faults (around $z = 400\,\text{m}$ and $z = 500\,\text{m}$, respectively) which are not recovered in the final model. Finally note that CIGs are well focused after 5 iterations and similar to those obtained in the exact model (figure 6.29, 4th and bottom row). Few modifications are visible in subsequent iterations compared to the evolution of the background velocity model.

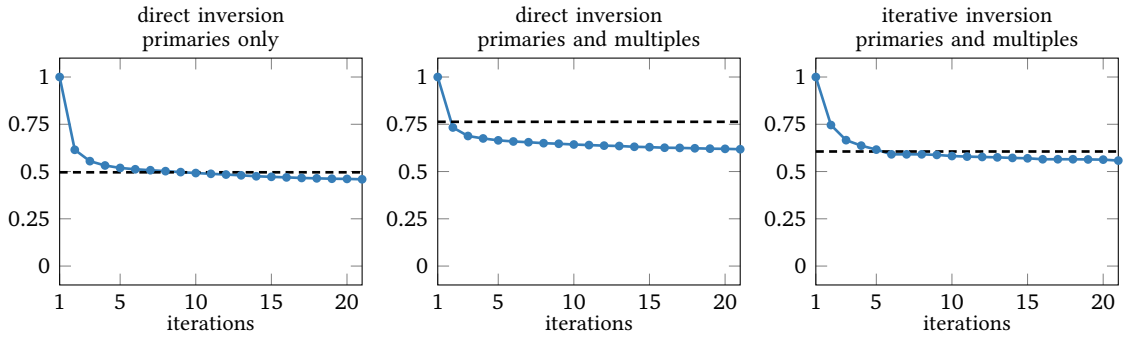


Figure 6.30. – Value of the MVA objective function (6.10) across iterations (blue, solid). The dashed black line represents the value obtained in the exact background velocity model. Three different inversions strategies are shown: direct inversion on primaries only (left, corresponding to figure 6.28), direct inversion on primaries and multiples (centre, corresponding to figure 6.31), and iterative inversion on primaries and multiples with application of the filter $\mathbf{F}^\dagger\mathbf{F}$ and the "truncated gradient" as background velocity update (right, final result shown in figure 6.33b, bottom right).

The same example is run with first-order surface multiples in observed data with identical smoothing parameters (figures 6.31 and 6.32). We use direct inversion for the definition of the extended image to evaluate the impact of multiples on the reconstruction of the velocity model. As already discussed in section 4.4.2, the associated MVA objective function is not minimal for the correct velocity model due to the presence of multiples. Here, a smaller value is reached after a single iteration (figure 6.30, middle), indicating that inversion converges to a compromise model which should focus both cross-talks artefacts and events corresponding to true reflectors. The velocity models obtained in the first iterations have a shape similar to the primaries only case (figures 6.28 and 6.31), but after a few iterations, events with an upward curvature (around $z = 400\,\text{m}$ depth at iteration 5 for example, fourth row in figure 6.32) misguide background velocity updates and prevent a proper focusing of the event corresponding to the shallowest interface. Consequently, the value of the velocity at this interface is not correctly recovered,

and neither is the low velocity anomaly.

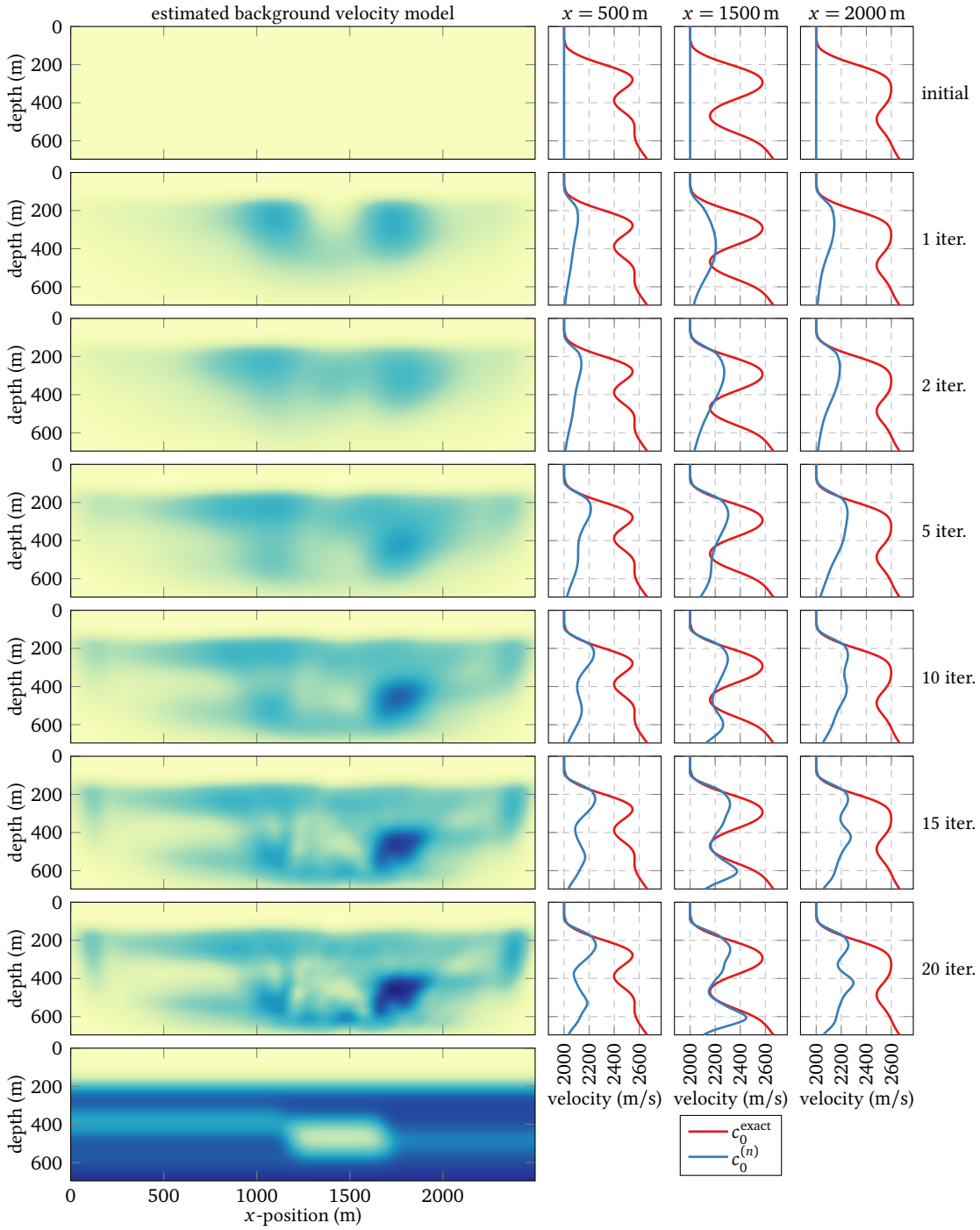


Figure 6.31. – Same as figure 6.28 with both primaries and first-order surface multiple in observed data.

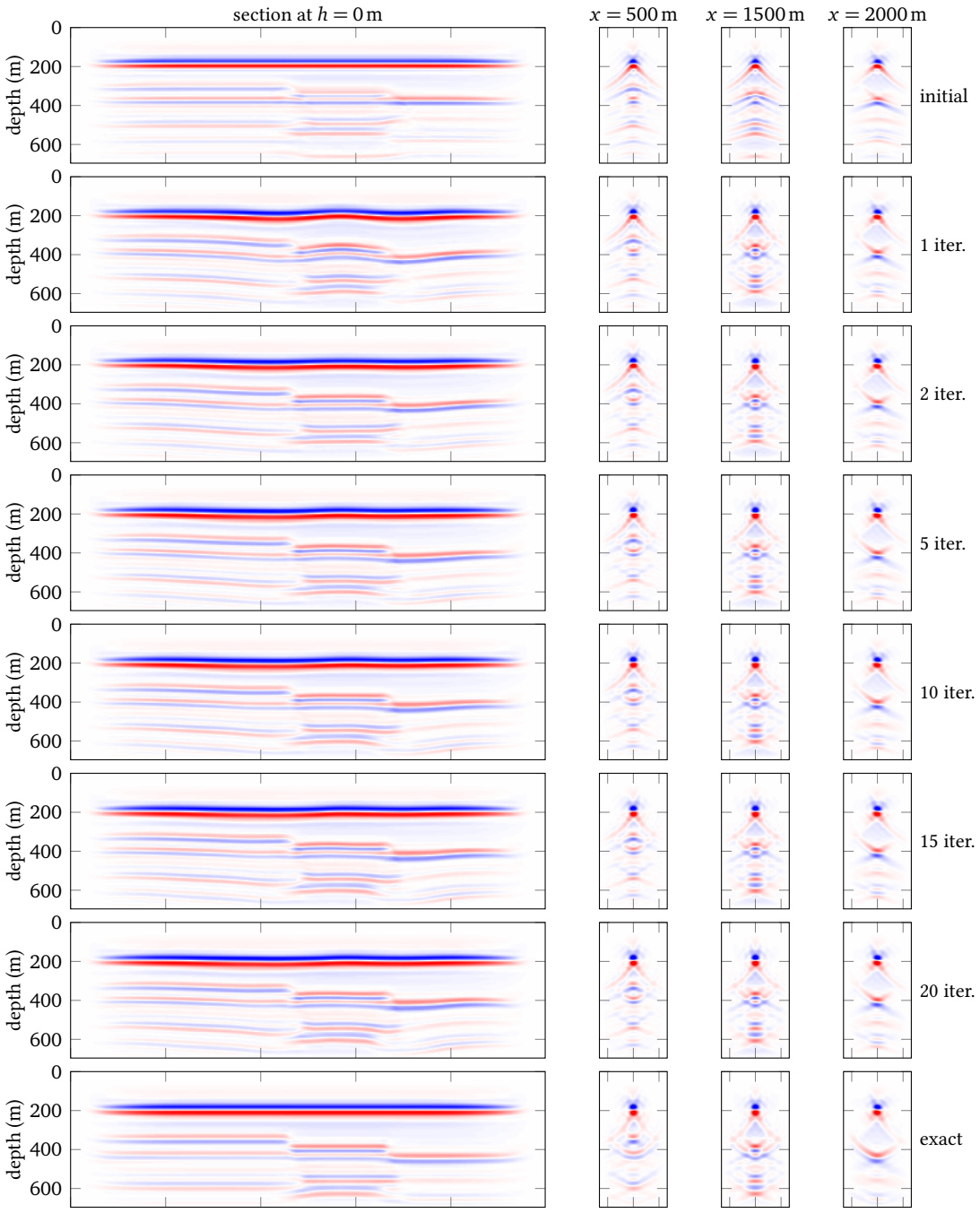


Figure 6.32. – Same as figure 6.29 in the case of multiples for the background velocity models of figure 6.31.

Finally we replace direct inversion by iterative migration, still with both primaries and first-order surface multiples in observed data. We perform 5 inner iterations on the reflectivity

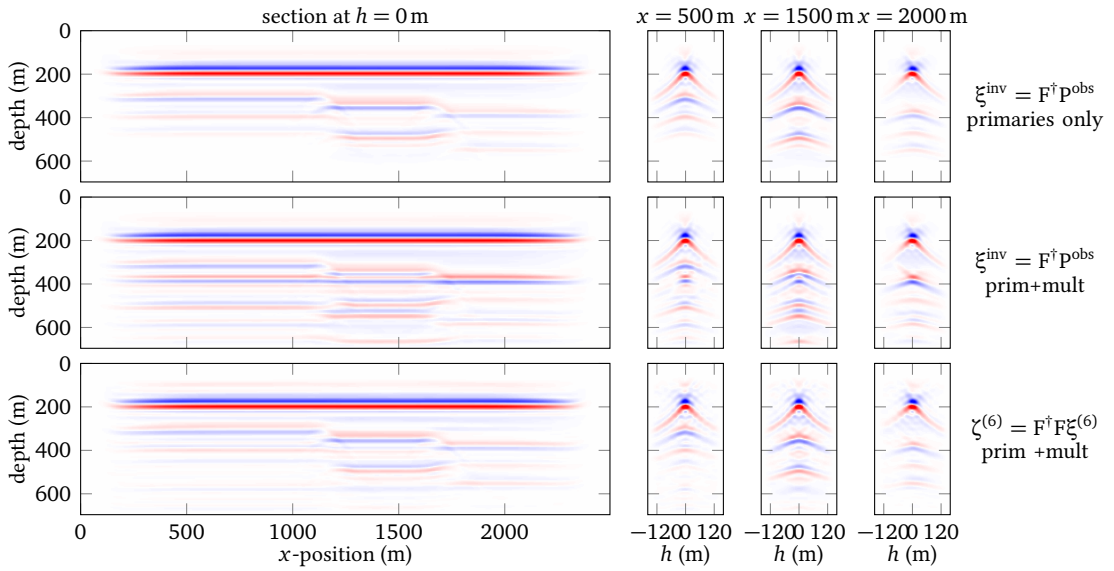
model for each outer iteration and compute a background velocity update with the "truncated gradient" strategy: primaries are remodelled from the final reflectivity model and the direct inversion strategy applied to this new data set to obtain a background velocity update. We begin by comparing the reflectivity sections obtained in the initial-velocity model (figure 6.33a). The cross-talk artefacts appearing after direct inversion the complete data set (around $z = 400\,\text{m}$) due to the multiple reflecting twice at the shallower interface are greatly attenuated after five iterations. Although the gradient obtained with direct inversion on primaries and multiples is quite similar to the primaries only case, we obtain a much closer velocity update with the "truncated strategy" (figure 6.33b, left column). Note that the largest values in the gradient are due to the faults, but consistent homogeneous updates are obtained on the left and right sides of the fault. The objective function decreases at the same pace as in the primary only case, and reaches a slightly lower value than the one obtained in the correct velocity model. Finally, the background velocity model recovered after fifteen iterations is very similar to the result of inversion obtained in the primaries only case (figure 6.33b, right column), showing that multiples have been correctly included in the process.

We conclude from this example that the truncated gradient approach, consisting of applying the direct inversion to primaries only recomputed from the final result of iterative migration, is an efficient strategy to deal with surface-related multiples, which otherwise prevent the inversion to converge to a reliable background velocity model. The inversion should in theory also be performed with the complete gradient, but this is more computationally expensive as the adjoint problem has to be solved and as more inner iterations are needed in practice to obtain consistent gradients. The truncated gradient is not formally the gradient of an objective function but efficiently deals with multiples at a reasonable computational cost.

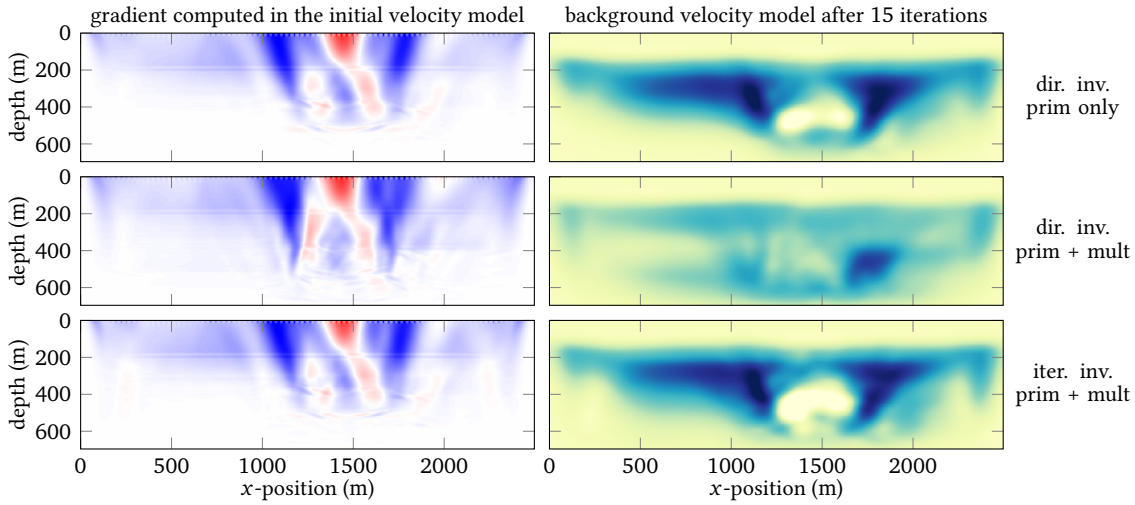
6.4. Incomplete acquisition

Finally we consider an example with an incomplete acquisition to test the ability of our strategy to benefit from the additional information contained in multiples. For this example, only the first MVA gradient is computed (we do not loop over the background velocity model). We consider a 700 m deep and 3750 m large model, discretised on a $4.8 \times 4.8\,\text{m}$ grid, with sources every six grid points from $x = 0\,\text{m}$ to $x = 1500\,\text{m}$ and from $x = 2250\,\text{m}$ to $x = 3750\,\text{m}$ (figure 6.34). Receivers are located at each grid point at the surface within $\pm 700\,\text{m}$ around each source. Unlike sources, receivers are not removed from the central part of the model. Iterative migration is performed with this acquisition setting. After 20 iterations, data are re-computed for sources located every four grid points, including the central part. Using the truncated gradient strategy, velocity analysis is performed on this new data set with the direct inversion strategy (section 4.2).

We want to compare this background velocity update with the result obtained using the direct inversion strategy on the original data set. We consider a pure 1D velocity and density model (figure 6.34) built by extending the central part of the exact model shown in figure 6.27. Hence the difference between the results of direct and iterative inversion are not due to 2D effects. The initial background velocity model is homogeneous ($2000\,\text{m/s}$), so that we expect a negative gradient.



(a) Migrated images in the initial velocity model, using direct inversion on primaries only (top), direct inversion on primaries and multiple (middle), and iterative migration and filtering on primaries and multiples (bottom).



(b) First gradient of the inversion (left), corresponding to the migrated images of (a), and background velocity model recovered after 15 iterations.

Figure 6.33. – Results of 20 iterations on the background velocity model using the iterative migration strategy with observed data containing both primaries and first-order surface multiples, and comparison with the results obtained with direct inversion. In the iterative case, the "truncated gradient" strategy is used to compute the background velocity update.

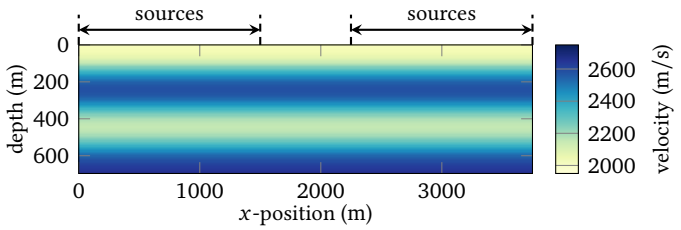


Figure 6.34. – Laterally invariant velocity model considered in section 6.4. It is constructed by extending the central part of the 2D velocity model shown in figure 6.27. The horizontal black arrows indicate sources lateral position. There are no sources between $x = 1500\,\mathrm{m}$ and $x = 2250\,\mathrm{m/s}$.

With this acquisition setting, no zero-offset trace is recorded in the central part of the model. The width of the acquisition hole (750 m) is only a little larger than the maximum surface offset (700 m). However tapers are applied to smooth the discontinuity in source and receivers positions and to attenuate edge effects. Hence sources located near the acquisition hole have little influence and the area not illuminated by primaries and direct inversion is not negligible (figure 6.35, top). Consequently there is no update of the background velocity in the central part of the model using the direct inversion strategy.

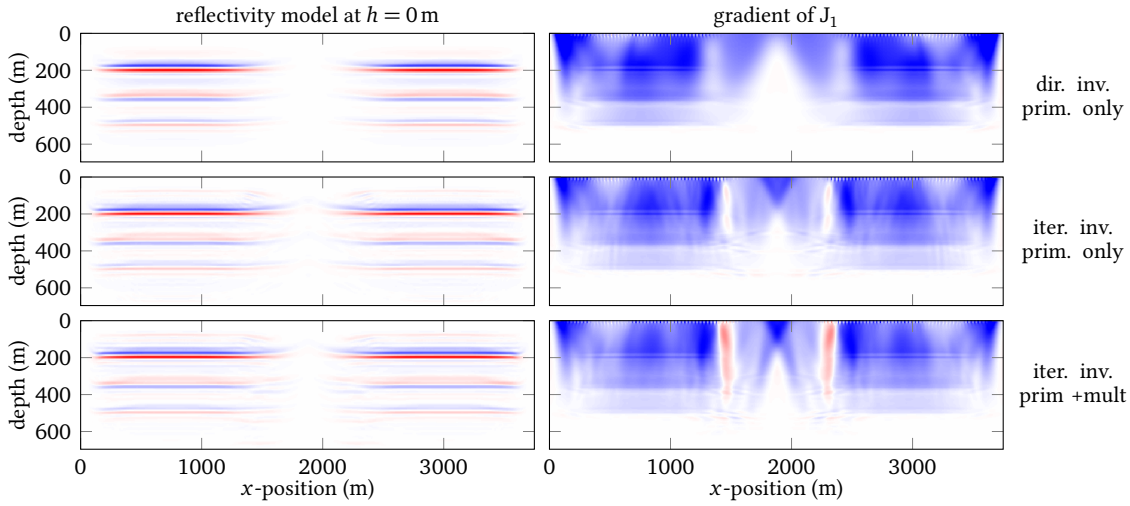


Figure 6.35. – Section at $h = 0$ of the reflectivity model (left) and MVA gradient (right) obtained in a homogeneous background velocity model (2000 m/s) with three different strategies: direct inversion on primaries only (top), and truncated gradient computed after 20 iterations on J_0 with primaries only in both observed and calculated data (middle), and with primaries and multiples in both observed and calculated data (bottom). In the iterative case, the reflectivity section displayed here is the result of application of the filter $F^{\dagger}F$ to the final reflectivity model computed by iterative migration.

Twenty iterations are performed on the reflectivity section in the primary only case, as well as in the case of first-order surface multiples (figures 6.35 and 6.36). In the primary only case, iterative migration provides a reflectivity model with a smaller hole in the central part of the model compared to direct inversion (figure 6.35, middle). The area with non-zero background velocity update is also reduced. In the case of multiples, we obtain a very similar reflectivity image (figure 6.35, bottom). Although the multiples illuminate the central part of the model, we do not succeed in recovering a continuous reflector. Some cross-talk artefact are even still visible (around $z = 400\,\text{m}$ and $x = 1600\,\text{m}$ for example). As a consequence the MVA gradient does not exhibit improvement compared to the primary only case and there is even a stronger artefact with a wrong sign around $x = 1600\,\text{m}$ and $x = 2200\,\text{m}$. A better result might be obtained by performing much more iterations, but the associated computational cost would not be affordable.

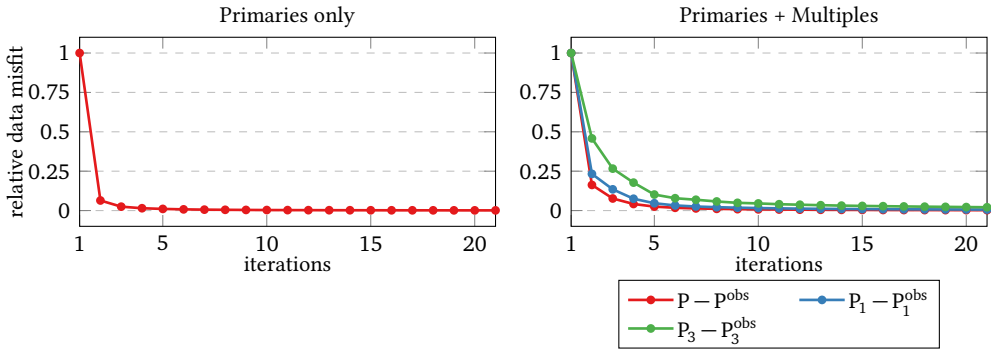


Figure 6.36. – Relative data misfit across iterations when performing iterative migration in a homogeneous velocity model (2000 m/s) with primary only (left, corresponding to figure 6.35, middle) and primaries and first-order surface multiples (left, corresponding to figure 6.35, right).

We conclude from this example, that iterative migration allows to extract more information on the edges of the model compared to direct inversion, even in the case of primaries only. However, the additional information contained in first-order surface multiples is difficult to recover as it is located on the edges of the acquisition where strong edge effects alter the gradient. Also the tapers applied to the sources and receivers on the edges of the acquisition reduce the weight given to multiples containing additional information.

6.5. Conclusion

In this chapter, we have tested the behaviour of iterative velocity analysis in the case of direct inversion and iterative migration when observed data are computed with a forward modelling code different from the one used during inversion. We have shown that our approach is robust against the lack of low frequencies in observed data and to mild errors in amplitude predictions. However, a reliable estimation of the source wavelet is required for iterative migration to obtain extended reflectivity images able to explain both primaries and surface-related multiples. In

the perspective of real data applications, we have shown that ghost reflections are an issue and that they should be properly modelled or removed. Finally we have compared the background velocity update obtained as the complete gradient of J_1 with an alternative strategy introduced in chapter 5 and referred to as *truncated gradient* in this chapter. It consists of applying the direct inversion strategy to primary reflection data re-computed from the result of iterative migration. We have shown that this approach yields consistent background velocity updates and is more robust than the complete gradient computation with also a lower computational cost.

Then this strategy has been applied on a 2D synthetic example consisting of a model with lateral velocity and density variations, observed data being modelled under a second-order Born approximation with both velocity and density perturbations. After fifteen iterations on the background velocity model, we obtained a result similar to the one obtained with the direct inversion strategy and observed data containing only primary reflections. This example shows the efficiency of our approach applied to data containing both primaries and first-order surface multiples and yielding results similar to those obtained without multiples in observed data.

Finally, we have shown on an example with an incomplete acquisition that iterative migration allows to recover additional information about the reflectivity model and the associated background velocity update. However, it has not succeed in extracting the information contained in first-order surface multiples. Further investigation is needed to achieve better results, in particular by extending the method to higher-order multiples.

Chapter 7.

Conclusions and Perspectives

Contents

Résumé du chapitre 7

Je me suis intéressé dans cette thèse à la détermination des grandes longueurs d'onde du modèle de vitesse de propagation des ondes P par les méthodes d'analyse de vitesse par migration. J'ai en particulier étudié l'orientation profondeur de l'optimisation par semblance différentielle (DSO). La qualité d'un macro-modèle de vitesse est estimée en analysant la répartition d'énergie dans un modèle de réflectivité, « étendu » avec l'offset en profondeur, obtenu par migration des données observées.

J'ai montré que remplacer l'étape de migration classique par une migration à amplitude préservée, ou inversion, pour le calcul des CIGs a une influence déterminante sur le succès de la méthode. En particulier l'inversion permet de s'affranchir des artefacts de migration qui ne satisfont pas le critère cinématique de la DSO. J'ai introduit et comparé dans ce manuscrit deux stratégies : une inversion « directe » où le modèle de réflectivité est obtenu par application d'un pseudo-inverse aux données observées et une stratégie itérative où le modèle de réflectivité est défini en tant que solution d'un nouveau problème inverse. Avec une légère modification de la fonction coût de la DSO, la stratégie utilisant l'inversion directe produit des remises à jour du macro-modèle cohérentes et lisses, et les « artefacts du gradient » sont fortement atténués. En revanche elle n'est définie que pour le cas de réflexions primaires. Pour pouvoir considérer également des réflexions multiples dans la DSO, j'ai étudié le cas où le modèle de réflectivité est déterminé par migration itérative, d'abord dans le cas de primaires seuls où il peut être comparé à la stratégie utilisant l'inversion directe. J'ai montré sur des exemples numériques que la migration itérative atténue efficacement les artefacts de migration et fournit des modèles de réflectivité proches de ceux obtenus par inversion directe. Par ailleurs, la convergence de la migration itérative peut être accélérée en utilisant un préconditionneur construit à partir du pseudo-inverse. La principale difficulté du cas itératif est l'instabilité du calcul de la remise à jour du macro-modèle, qui est définie comme le gradient de la boucle externe d'un problème d'optimisation à deux niveaux. En pratique, cette instabilité signifie que de faibles différences entre deux modèles de réflectivité peuvent conduire à de grandes différences sur le gradient. Elle est liée à l'apparition, au cours de la migration itérative, d'évènements aux grandes valeurs de l'offset en profondeur du modèle de réflectivité. Cette énergie, non porteuse d'information cinématique, est amplifiée par l'application de l'annihilateur et perturbe la migration itérative. L'instabilité du calcul du gradient peut être atténuée en régularisant la boucle interne. La difficulté est alors de déterminer le poids donné à la régularisation : ce choix est délicat, coûteux numériquement, et doit a priori être renouvelé à chaque remise à jour du macro-modèle. J'ai étudié au chapitre 5 une autre régularisation, portant sur la boucle externe et qui consiste à appliquer au résultat de la migration itérative un opérateur atténuant l'énergie non porteuse d'information cinématique et responsable des instabilités. Cette approche ne requiert pas de régularisation spécifique de la boucle interne et fournit des remises à jour du macro-modèle similaires à celles données par l'approche directe.

J'ai étudié l'extension de l'approche itérative aux cas des réflexions multiples de surface du premier ordre, modélisées ici avec une approximation de Born du second ordre. La migration itérative devient alors un problème d'optimisation non linéaire. J'ai illustré sur des exemples numériques l'atténuation au cours des itérations des artefacts dans les CIGs causés par l'interprétation des multiples comme réflexions primaires. Dans le cas de multiples, la régularisation

de la boucle interne demeure essentielle pour obtenir des remises à jour du macro-modèle satisfaisantes, même avec la régularisation sur la boucle externe introduite dans le cas des primaires. Pour avoir un schéma plus flexible et moins coûteux numériquement, j'ai finalement proposé une stratégie alternative où le modèle de réflectivité étendu obtenu par migration itérative est utilisé pour générer un nouveau jeu de données ne comprenant que des réflexions primaires, cohérents cinématiquement avec ceux contenus dans les données observées. La remise à jour du macro-modèle est alors obtenue en appliquant la stratégie d'inversion directe à ce nouveau jeu de données. Cette remise à jour n'est pas le gradient d'une fonction objective mais elle demeure cohérente et robuste d'une part et moins coûteuse numériquement que le calcul du gradient complet d'autre part.

Dans cette étude, l'introduction des multiples dans la stratégie de DSO a été faite dans l'optique de s'affranchir des artefacts liés aux multiples, sans réellement chercher à exploiter l'information supplémentaire qu'ils contiennent. Cela demanderait probablement de considérer tous les ordres de multiples de surface ainsi que les multiples internes. Des approximations de Born d'ordres plus élevés ne constituent probablement pas une solution satisfaisante. En particulier j'ai illustré les difficultés de cette approche lorsque les effets de surface libre sont introduits dans la modélisation des données observées. Dans l'optique d'une prise en compte de tous les ordres de multiples et des multiples internes, la construction de CIGs fonction de l'offset en profondeur avec les méthodes d'imagerie de Marchenko a été étudié par Díaz *et al.* (2016) et Díaz (2016). Cette approche semble prometteuse, mais il reste à définir une stratégie efficace pour le calcul de la remise à jour du macro-modèle.

En vue de possibles applications à des données réelles, j'ai testé dans le chapitre 6 la robustesse de la DSO définie avec la migration itérative sur des jeux de données obtenus avec une physique différente de celle utilisée pour l'inversion. J'ai en particulier illustré la sensibilité de la migration itérative vis-à-vis de l'estimation de l'ondelette de source dans le cas de multiples. Il serait intéressant de chercher à introduire d'avantage de physique dans la méthodologie de la DSO. Par exemple, une méthodologie unifiée prenant en compte simultanément les ondes réfléchies (primaires et multiples) et les ondes transmises suivant la stratégie de (Lameloise et Chauris, 2016) reste à définir. Par ailleurs, la méthodologie conduisant à la formule d'inversion de perturbations du modèle de vitesse, proposée dans la section 4.2, pourrait être appliquée au cas où les réflexions sont causées par des perturbations du modèle de densité. Plus généralement, la physique de propagation des ondes utilisée pourrait également prendre en compte une densité variable ainsi que les effets de propagation élastique, d'atténuation et d'anisotropie.

Enfin la stratégie d'optimisation utilisée pour l'inversion du macro-modèle n'a pas été étudiée en détail dans cette étude. Huang et Symes (2015) utilisent la méthode de projection des variables (VPM) et une fonction objective différente où le terme de DSO est assimilé à une régularisation de la fonction coût comparant données observées et données calculées. Par ailleurs, l'exploitation de l'information contenue dans le Hessien de la fonction coût de DSO n'a été que peu explorée dans la littérature, en particulier en raison du coût numérique élevé de son évaluation complète. C'est pourtant un aspect essentiel dans la remise à jour du macro-modèle. Le coût de calcul de la DSO et la taille mémoire qu'elle demande restent par ailleurs le principal obstacle à son extension au cas 3D.

7.1. Conclusions

In this thesis, I have investigated a Migration Velocity Analysis (MVA) technique for the resolution of the seismic inverse problem under the constant-density acoustic wave-equation approximation. This technique is defined in the image domain and relies on the Born approximation and a separation of scales of the velocity model into a rapidly varying reflectivity model and a smooth macro-model. It aims at assessing the quality of an estimated background velocity model using the redundancy of seismic data. In the depth-oriented formulation, physical reflectivity images parametrised by spatial coordinates are extended with an additional variable h called subsurface-offset. After migration of seismic data to this extended domain, inaccuracies in the initial macro-model result in defocused energy at non-zero values of h. Using the Differential Semblance Optimisation (DSO) strategy, the macro-model is iteratively corrected by minimising an objective function penalising defocused energy in Common Image Gathers (CIGs) until the extended reflectivity model becomes physical.

In this thesis, I have addressed the issue of spurious events appearing in CIGs which prevent the adoption of MVA techniques as standard seismic data processing tools. These artefacts are due to limited extension of acquisition geometries on one hand (Lameloise *et al.*, 2014; Mulder, 2014), and to multiple reflections misinterpreted as primary events on the other hand (Mulder and ten Kroode, 2002; Li and Symes, 2007), and do not properly focus for the correct velocity model. As a consequence, the MVA objective function is not minimum for the correct velocity model and its gradient does not provide a consistent background velocity update. Here I have investigated the use of inversion instead of migration to deal with both issues and obtain CIGs free of artefacts as well as consistent MVA gradients.

7.1.1. Inversion Velocity Analysis

The original formulation of MVA assumes primary reflection data only, and defines a reflectivity image by application of the adjoint of the extended Born modelling operator to observed data. In the context of high-frequency approximation of the wave-equation, Lameloise *et al.* (2014) introduced an extended quantitative migration compensating for uneven illumination and geometrical spreading. Combined with the horizontal contraction technique (Fei and Williamson, 2010; Shen and Symes, 2015), this approach yields CIGs free of migration artefacts and smooth gradients. However ray theory is limited in the presence of complex geology, for which wave-equation based approaches are more appropriate. In this study, an approximate inverse using only wave-equation operators has been presented (section 4.2), with a formulation similar to other proposals (Hou and Symes, 2015, 2017). All these formulas are actually inverse operators in an asymptotic sense. Their derivation rely on ray theory and high-frequency approximations, but their final expressions are free of ray quantities. This strategy results in improved background velocity updates compared to standard migration. Moreover the introduction of a power of the background velocity in the annihilator has been shown to remove oscillations located around the reflector position from the MVA gradient (section 4.2) yielding results similar to the horizontal contraction approach (Fei and Williamson, 2010; Shen and Symes, 2015). However the macro-model update derived with inversion is the gradient of an objective function, contrary to the case of horizontal contraction. Eventually, note that the

advantage of this strategy is the recovery of true-amplitude images and CIGs free of migration artefacts.

As inversion formulas proposed so far are designed for primary reflections only, I have studied in this thesis iterative migration as a replacement for direct inversion. The final formulation has been constructed progressively in chapters 2 to 5. The analysis has been performed in the case of primaries only, using the result of direct inversion as a reference, and then applied to the case of multiples. I also studied an equivalent 1D MVA problem in sections 4.2.5 and 5.2 which may not be representative for all the features of 2D subsurface-oriented DSO, but is a useful analysis tool as it exhibits a behaviour very similar to the 2D case with a much lower computational cost. The main conclusions of each chapter are summarised in table 7.1. I underline here the key aspects of the method.

In iterative MVA, the extended reflectivity image is not obtained via an inverse formula but as the solution of an inverse problem. For a given background velocity model, the model perturbation best explaining observed data in a least-squares sense is determined through an optimisation procedure. If calculated data are computed under the first-order Born approximation (primaries only), the associated inverse problem is linear and can be solved with a standard conjugate gradient algorithm. I have shown that this technique efficiently attenuates migration artefacts in CIGs and yields results similar to direct inversion in the primary-only case. Compared to the result of direct inversion, we obtain after several iterations of migration a reflectivity image with a better deconvolution of the source wavelet, but this has little impact on the final data misfit.

The derivation of the gradient of the outer objective function has been analysed in chapter 3. An approximate value of this gradient is computed using the final result of two iterative problems: iterative migration on one hand, and a linear adjoint problem, similar to the direct problem but with a different source term on the other hand. One would expect the gradient to converge to a stable value at the same pace as iterative migration. Besides, as direct and iterative inversion provide similar reflectivity images, we may presume that the associated background velocity updates would be similar. However, numerical examples in chapters 3 to 5 have shown that this is not the case in practice. The value of the reflectivity for large h has been shown to have little impact on data residuals but a major influence on the velocity analysis, after being amplified by the annihilator. Similar observations have been made by Huang (2016). This issue has been identified by comparing the value of both objective function J_0 and J_1 across inner iterations. The first one reaches convergence after a few iterations, while the second does not stabilise because of small modification at large values of h (figure 3.8). Another practical difficulty is the slow convergence speed of the adjoint problem. As a consequence, the value of the gradient obtained with successive values of the adjoint variable is not stable. This is an undesirable behaviour as we would like the gradient to converge to a stable value after a given number of iterations performed for the resolution of both problems. I have shown that sufficient regularisation on the reflectivity model helps mitigating these issues, but this solution is not fully satisfactory. In particular, the determination of regularisation parameters remains a tedious task and we still observe residual oscillations around reflector positions not present in the gradient obtained by direct inversion. For a more stable procedure, I have proposed to modify the usual MVA objective function by applying a "filter" depending on the background velocity model to CIGs before measuring defocused energy. This filter does not change the shape

	main realisations and conclusions	limitations
chapter 2	• an extended reflectivity model is defined by iterative migration to minimise the misfit between observed data and calculated data • migration and cross-talk artefacts are greatly attenuated with iterations, improving the shape of the MVA cost function;	• a strategy should be defined for the coupling with velocity analysis;
chapter 3	• two methods for the computation of the MVA gradient after iterative migration are compared; • the selected method assumes that iterative migration reaches convergence. An adjoint variable is determined as the solution of a linear problem and the gradient is computed from the last iterates of the direct and adjoint problem;	• iterative MVA is computationally expensive as two iterative systems have to be resolved for the gradient computation; • sufficiently strong regularisation is needed to obtain gradients free of spurious oscillations and stable across inner-iterations;
chapter 4	• an approximate inverse of the extended Born modelling operator is defined for primaries only, leading to a new IVA strategy; • introducing this approximate inverse as a preconditioner greatly accelerates the convergence speed of the direct problem;	• the adjoint problem converges much slower than the direct problem and the associated sequence of MVA gradients are not stable; • the MVA gradients obtained after direct and iterative migration are quite different despite the similarity of the associated reflectivity images;
chapter 5	• iterative MVA is analysed on a pure 1D case. It exhibits a behaviour similar to the 2D case; • a filter $F^{\dagger}F$ is introduced in the definition of the MVA objective function to attenuate unwanted energy at large values of h; • with this modified MVA objective function, the adjoint problem converges much faster and the associated sequence of gradients is more stable with a shape similar to the gradient of direct inversion;	• the robustness of the approach regarding non-Born data should be investigated (beyond the inverse crime);
chapter 6	• the direct and iterative inversion strategy are applied to a series of synthetic data computed with a modelling engine different from the one used during inversion; • direct inversion in the primaries only case is robust to inaccuracies of the source wavelet and to amplitude errors; • direct and iterative inversion strategies are robust with respect to density perturbations;	• in the multiple case, iterative migration fails to explain both primaries and multiples correctly if the shape of the source wavelet is inaccurately estimated; • application to real data.

Table 7.1. – Summary of the main conclusions of each chapter.

of defocused events in CIGs but attenuates spurious oscillations responsible for the instabilities described above. I have shown that this filter changes the source term of the adjoint problem which is now easier to solve. The value of the gradient is stable with iterations and similar to the one obtained by direct inversion. The additional cost represented by the application of the filter is largely compensated by the improved convergence of the adjoint problem, allowing to reduce the number of iterations.

7.1.2. Multiple reflections

I have studied the extension of this iterative procedure to the case of first-order surface-related multiples. It consists of using a second-order Born approximation for forward modelling in the inner inverse problem. This changes the gradient formula in both the inner and outer minimisation problem. Iterative migration is not a linear inverse problem any more. Regarding the computation of MVA gradient, the adjoint problem is still linear but may not be positive-definite if final data residuals are not small enough.

I have shown that cross-talk artefacts due to multiples are attenuated across iterations, leading to MVA gradients consistent with the primary-only case. However the computation of the migration objective function and of its derivatives is much more expensive compared to the linear case (table 3.1). Moreover iterative migration becomes a non-linear optimisation procedure. To reduce the computational expense of the method, we have proposed to use the approximate inverse as a preconditioner, which greatly improves the convergence of the direct and adjoint problem, even in the case of multiples. Finally we have proposed an alternative strategy where iterative migration is used to retrieve an extended reflectivity image free of cross-talk artefacts, allowing to re-compute primary reflections and to apply the direct inversion procedure which has been shown to be robust and computationally efficient. Although the background velocity update defined by this strategy is not the gradient of an objective function, the procedure is lees computationally expensive as the iterative resolution of the adjoint problem is not necessary. Furthermore, we can reconstruct data for source positions missing in the original acquisition. Hence eventual additional information in the reflectivity image due to extra-information contained in multiples may be incorporated in the re-computation of primaries.

In this study, multiples have been introduced with the objective of retrieving gradients similar to those obtained with multiple-free data. The models and acquisition settings used in the examples were not specifically designed to use the additional information contained in multiples. We have shown that iterative migration succeeds in removing the imprint of multiples on the reflectivity, leading to gradients similar to those obtained in the case of primaries only. However, the example of section 6.4 shows that extracting additional information from multiples is a much more complex task, especially when considering only first-order surface-related multiples. The areas investigated by this kind of multiples is actually small and located on the edges of the acquisition where edge effects alter MVA gradient obtained in the primaries only case. Extending MVA techniques to all order of multiples and internal multiples may provide more favourable examples with wider illumination of the subsurface (see next section).

7.2. Perspectives

7.2.1. Using all orders of multiples

Extending MVA to all orders of multiples could be performed with the approach investigated by Díaz (2016), based on the resolution of the *Marchenko* equation. Using observed data and Green's function $G_0(s, \mathbf{x}, \omega)$ computed in an estimated macro-model $c_0(\mathbf{x})$, the iterative resolution of the Marchenko equations allows to retrieve the complete Green's functions $G(s, \mathbf{x}, \omega)$ including all multiple reflections. This step is discussed in Wapenaar *et al.* (2014) for the case of internal multiples and extended to surface multiples by Singh *et al.* (2015). The accuracy of the macro-model can then be evaluated in two different ways. First, one possibility is to decompose the complete Green's function into a downgoing G^- and an upgoing wavefield G^+ to construct extended CIGs with a deconvolution imaging condition and measure defocused energy in a standard way (Díaz *et al.*, 2016). An alternative consists of directly estimating the velocity model as G is reconstructed everywhere within the subsurface and is solution of the wave-equation

$$\frac{(\mathrm{i}\omega)^2}{c^2(\mathbf{x})} G(s, \mathbf{x}, \omega) - \Delta G(s, \mathbf{x}, \omega) = \delta(x - s), \tag{7.1}$$

the value of the velocity model $c(\mathbf{x})$ can be recovered from the knowledge of the complete Green's function following

$$c^2(\mathbf{x}) = \frac{\left\langle (\mathrm{i}\omega)^2 G \,\middle|\, G \right\rangle}{\left\langle \Delta G \,\middle|\, G \right\rangle}. \tag{7.2}$$

This model can be decomposed into $c(\mathbf{x}) = c_0(\mathbf{x}) + \delta c(\mathbf{x})$. For the determination of $G(s, \mathbf{x}, \omega)$, one needs to provide a smooth background model. If the kinematics is not correct, a standard MVA procedure can be defined using the values of P^{obs} and $c_0(\mathbf{x})$. Both the strategy of Díaz (2016) and our approach rely on an iterative procedure depending on the value of the macro-model $c_0(\mathbf{x})$. In Díaz (2016) the purpose is to retrieve the complete Green's function G, whereas here the unknown of the iterative process is an extended reflectivity model best explaining observed data. In both cases, the difficulty is to compute the velocity update due to the dependence of the iterative process to the macro-model $c_0(\mathbf{x})$. Here we derive a strategy using the adjoint-state method and the assumption that the output of the iterative migration is defined as the minimiser of a cost function. Further investigations are required to define a similar strategy in the case of Díaz (2016). It is possible that a regularisation term (equivalent of the filter K) has to be introduced with Marchenko-based MVA process. Note that the major advantage of the Marchenko approach is that both internal and surface multiples are consistently integrated in the imaging procedure, yielding CIGs free of cross-talk artefacts (Díaz, 2016).

7.2.2. Introducing more physics in MVA techniques

In section 4.2, an inversion formula allowing to retrieve an extended velocity perturbation explaining observed data has been presented. This formula was tested in section 6.2.4 on reflection data originating from a density perturbation and modelled under the Born approximation. Alternatively, one could also derive an inversion formula for density perturbation data, using a strategy similar to the one used in section 4.2. The difference is that here, gradients of Green's

function are involved instead of a their product with a second-order time derivative, which requires the derivation of new weighting operators different from the original inverse formula. This may enable an other inversion strategy for multi-component observed data. Using the surface recording of both horizontal and vertical displacements, one could try to invert for a velocity or a density perturbation, or for a velocity and impedance perturbation, as these two parameters exhibit less coupling (Zhou *et al.*, 2015).

In this study we have used an explicit scale separation between a smooth background velocity model and a model perturbation. Following a strategy proposed by (Zhou, 2016) in the framework of Full Waveform Inversion (FWI), one could also investigate a more natural scale separation, using velocity and variable density to parametrise the model, the velocity model controlling the kinematics of wave propagation and density accounting for the reflective property of the subsurface instead of the velocity model perturbation.

In the constant-density acoustic approximation, Lameloise and Chauris (2016) show how transmitted waves can be included in MVA techniques by constructing extended images in a very similar manner to the usual procedure for reflections. They propose a strategy in which these two kind of events are used successively. Transmitted waves are used in a first step to update the shallow part of the velocity model; then primary reflections are inverted to reach the deeper part of the model. In our approach, primaries and multiples are naturally inverted together. Inverting reflection and transmission data simultaneously could potentially better constrain the inversion. As pointed out by Lameloise (2015), an issue is that transmitted events are more energetic than reflections, requiring to introduce weights to balance the amplitudes of both kinds of events. Moreover, CIGs constructed with transmitted events do not have the same physical interpretation as CIGs built with reflection data; in particular they are not linked to a specific interface. Hence, the inclusion of reflection and transmitted data in a unified framework needs further investigation.

7.2.3. Importance of the Hessian for iterations over the velocity model

A more thorough study should be led on the choice of optimisation strategies for the external loop aiming at determining the background velocity model. In particular the effect of the Hessian of the objective function should be better taken into account. The gradients obtained by direct inversion in the case of primaries (section 4.2) are remarkably smooth, but their shape, and in particular the importance of side effect vary greatly with the number of surface and subsurface-offsets, and the reflectors' depth. We may expect a proper introduction of the Hessian to attenuate these variations. Most applications on synthetic and real data "only" use the l-BFGS strategy to take second-order effects in to account. An interesting strategy is proposed by Liu *et al.* (2014b) and Shen and Symes (2015) who compute an estimate of the diagonal of the Hessian matrix as the result of its application to a unit vector. An extension is the Truncated Newton strategy, already studied in the framework of FWI (Métivier *et al.*, 2013, 2014). Contrary to the l-BFGS algorithm, these techniques require an efficient way of computing the product of the Hessian of the objective function with a vector of the $\mathscr{M}$ space. One could use a second-order adjoint-state technique similar to the one presented in Métivier *et al.* (2013, 2014) and in section 3.2.3. An example of derivation of the matrix-vector product is presented in appendix E in the case of direct inversion. Note however that this approach is quite

expensive. Moreover the positive-definitiveness of the Hessian matrix should be investigated.

7.2.4. Inversion strategy

We have considered here a nested optimisation procedure with two separate objective functions, the inner inverse problem solving for the reflectivity for a given macro-model, and the outer objective function solving for the macro-model. Another possibility is to use a single objective function (Fleury and Perrone, 2012; Huang and Symes, 2015),

$$J_2[c_0,\xi] = \frac{1}{2}\left\|F[c_0]\xi - P^{\text{obs}}\right\|^2_{\mathscr{D}_{\text{obs}}} + \frac{\lambda}{2}\left\|A\xi\right\|^2_{\mathscr{E}}, \tag{7.3}$$

where the first term is related to data misfit and the second measures defocused energy in CIGs. Note that the DSO strategy was originally introduced as a regularisation for the FWI objective function (Symes and Kern, 1994).

An efficient procedure for the minimisation of this objective function is the Variable Projection Method (VPM) (van Leeuwen and Mulder, 2009; Rickett, 2013; Huang and Symes, 2015). It consists of inverting first for the reflectivity ξ (following $\xi = F^{\dagger}[c_0]P^{\text{obs}}$ in the direct inversion case) and to replace in the definition of the objective function. The first step is similar to what have been proposed in this study, the second is different because we considered in this thesis only the second term in equation (7.3) for the derivation of a velocity model.

A difficulty remains the choice of the weight λ. Furthermore, in the case of iterative migration with multiples, the expression replacing ξ in equation (7.3) should properly take cross-talk effects into account, otherwise the second term of the objective function may mislead the background velocity update.

7.2.5. Extension to 3D

In 3D, the observed data space has five dimensions (s_x, s_y, r_x, r_y, t), requiring two extension parameters for the model space, for example horizontal subsurface offsets h_x and h_y (table 4.1). A new inversion formula taking into account these new dimensions should be defined with a strategy similar to the 2D case (section 4.2). The major issue of the extension is its computational cost, which disqualifies the iterative determination of an extended reflectivity image with a large number of iterations. Preconditioning as proposed in this thesis or in Hou and Symes (2016a) is an essential element for the extension to 3D. In addition to 3D wavefield propagation, a cross-correlation should be performed for each couple (h_x, h_y). New strategies such as those mentioned in section 1.3.4 should be further investigated for a more affordable extension of MVA to 3D. Another difficulty is that 3D acquisition do not provide dense source and receiver coverage in all directions, which may have undesirable effects on reflectivity image and associated gradients.

7.2.6. Application to real data

As a prelude to real data applications, we have applied in chapter 6 the proposed MVA strategy to observed data computed with a forward modelling code different from the one used during inversion. We showed that an incorrect estimation of the source wavelet was an issue for the

iterative case in the presence of multiples, but that direct inversion was insensitive to this kind of imperfection. More importantly, ghosts effects due to the free-surface were shown to be an issue for iterative migration as they modify the amplitude and phase of reflection data, especially at wide angles. Ghost should be removed from observed data or properly included in the forward modelling to cope with this issue.

More generally, propagation effects that are not accounted for in the forward propagation step may lead to inconsistent CIGs and velocity updates. For example the impact of attenuation and elastic effects on the amplitude may be an issue for regular MVA strategies. Converted waves are also not predicted under the Born approximation.

Finally the irregularity of the acquisition geometry should be taken care of. Direct inversion formulas are indeed derived with the assumption of dense source and receiver coverage. Iterative migration may be an interesting alternative in this case as it deals more effectively with irregular acquisition (Nemeth *et al.*, 1999).

Appendix A.

Modelling of free-surface reflection under the Born approximation

In this appendix related to section 2.3.2, we justify on a 1D case the expression for the operator M_s defined in equation (2.12) accounting for the reflection at the free-surface. Intuitively, we would like to use the Born formula and define a source term made of primaries recorded at the surface multiplied by a constant reflection coefficient $\mathrm{R} = (-1)$, as the free surface is known to act as a mirror (Schuster, 2007). Numerical comparison with a finite-difference modelling with a free-surface condition suggests that an additional coefficient and a time derivative $2\mathrm{i}\omega/c_0$ have to be applied as well. The purpose of this appendix is to explain the reason for this additional term. Actually, the origin of the reflection at the free surface can be described by a constant velocity perturbation above the surface with value (-1). However, in practice, the model considered for simulation does not extend above the free surface and we consider a reflectivity localised at the surface. This requires a modification of the Born formula.

Let us consider a pure 1D case with a single source and a single receiver both buried and localised at depth z_s. The surface is at a depth $z_0 < z_s$. In the asymptotic approximation, the 1D Green's function reads

$$\mathrm{G}_0(z_s, z, \omega) = \frac{c_0(z)}{2\mathrm{i}\omega} \mathrm{e}^{-\mathrm{i}\omega\tau(z_s,z)}, \tag{A.1}$$

where $\tau(s,z)$ is the traveltime between the source and the depth z the surface. First we consider a constant reflectivity model above the surface $\xi(z) = \mathrm{R} \cdot \mathrm{H}(z_0 - z)$ where H stands for the Heaviside function. Using equation (2.10), the upgoing incident wave reads

$$\mathrm{P}_{\mathrm{up}}(z_s, z, \omega) = \frac{c_0(z)}{2\mathrm{i}\omega} \mathrm{e}^{-\mathrm{i}\omega\tau(z_s,z)} \Omega(\omega), \tag{A.2}$$

and the downgoing reflected wave recorded at the source position reads

$$\mathrm{P}_{\mathrm{down}}(z_s, z_r = z_s, \omega) = \mathrm{R}\Omega(\omega) \int_{z=z_0}^{-\infty} \mathrm{e}^{2\mathrm{i}\omega\tau(z_s,z)} \, \mathrm{d}z. \tag{A.3}$$

We linearise the value of the traveltime around the surface,

$$\tau(z_s, z) \approx \tau_0 + \frac{1}{c_0(z)}(z_s - z_0), \tag{A.4}$$

where $\tau_0 = \tau(z_s, z_0)$. This leads to

$$\mathrm{P}_{\mathrm{down}}(z_s, z_r = z_s, \omega) = \mathrm{R}\frac{c_0(z)}{2\mathrm{i}\omega} \mathrm{e}^{-2\mathrm{i}\omega\tau_0} \Omega. \tag{A.5}$$

The downgoing wavefield has the same shape as the upgoing wavefield, except for the multiplication by the reflection coefficient $R = (-1)$. The surface acts as a mirror on a upgoing wavefields.

In practice we consider a reflectivity localised at z_0: $\xi(z) = R \cdot \delta(z - z_0)$. Following the same steps, this leads to

$$P_{\text{down}}(z_s, z_r = z_s, \omega) = R e^{-2i\omega\tau_0} \Omega(\omega). \tag{A.6}$$

Compared to, equation (A.5), a coefficient $\frac{c_0}{2i\omega}$ is missing in the result. To model the surface reflection properly, the source term of the Born modelling equation (2.10b) has to be multiplied by $\frac{c_0}{2i\omega}$.

Appendix B.

Non-linear local optimisation

In this appendix, we review standard non-linear optimisation methods. For an extensive review, the reader is referred to Nocedal and Wright (2006). Non-linear optimisation is needed in chapter 2 to determine a reflectivity model minimising the data misfit when multiples are included in the calculation of reconstructed data. The objective function of DSO is also minimised with non-linear optimisation to determine a background velocity model minimising defocused energy in CIGs.

B.1. General form of the algorithm

Without loss of generality, we use in this appendix the notations of iterative migration, that is we want to find a vector $\xi \in \mathscr{E}$ minimising a scalar objective function $J_0(\xi)$. With gradient-based methods, the reflectivity model at iteration $(n+1)$ is updated from the reflectivity model at iteration (n) following

$$\xi^{(n+1)} = \xi^{(n)} + \alpha^{(n)} d^{(n)}. \tag{B.1}$$

where $d \in \mathscr{E}$ is called descent direction and the positive scalar α is called step size. The process is initialised with an initial guess $\xi^{(1)} = \xi^{\text{ini}}$, and is stopped when a termination criterion is satisfied, for example when the value of the objective function or the norm of its gradient goes below a given threshold. In the numeric examples shown in this thesis, a maximum number of iterations N is set, so that the final result is $\xi^{(N+1)}$.

In local optimisation methods, the value of the objective function should decrease at each iteration, that is $J_0(\xi^{(n+1)}) < J_0(\xi^{(n)})$. Then d and α have to verify some conditions:

- to ensure that there exists a step length α allowing the decrease of J_0, d should be a *descent direction*, meaning that it should verify

$$\left\langle d^{(n)} \,\middle|\, g^{(n)} \right\rangle_{\mathscr{E}} \leq 0, \tag{B.2}$$

  where g is the gradient of J_0,

$$g^{(n)} = \frac{\partial J_0}{\partial \xi}(\xi^{(n)}). \tag{B.3}$$

  The most obvious choice for d is the opposite of the gradient. Alternative strategies providing faster convergence are presented in section B.3.

- If condition (B.2) is verified, very small values of α ensure that the value of J_0 at iteration $(n+1)$ is smaller than its value at iteration (n). However to avoid performing numerous small steps, we would like to find a value for $\alpha^{(n)}$ close to the exact minimiser of J_0 along $\xi^{(n)} + \alpha^{(n)} d^{(n)}$. An optimal α value $\alpha^{(n)}$ would lead to

$$\left\langle d^{(n)} \,\middle|\, g^{(n+1)} \right\rangle_{\mathscr{E}} = 0. \tag{B.4}$$

  Determining a good value for $\alpha^{(n)}$ is actually a one-dimensional optimisation problem. We elaborate on *linesearch* strategies addressing this problem in the next section.

B.2. Linesearch

The linesearch procedure aims at finding the minimum of J_0 along the direction $\xi^{(n)} + \alpha^{(n)} d^{(n)}$. In practice a scalar objective function $\psi : \mathbb{R} \mapsto \mathbb{R}$ is defined for fixed values of $\xi^{(n)}$ and $d^{(n)}$ as

$$\psi(\alpha) = J_0\big(\xi^{(n)} + \alpha d^{(n)}\big), \tag{B.5a}$$

and its gradient reads

$$\psi'(\alpha) = \left\langle d^{(n)} \,\middle|\, \frac{\partial J_0}{\partial \xi}\big(\xi^{(n)} + \alpha d^{(n)}\big) \right\rangle_{\mathscr{E}}. \tag{B.5b}$$

For linear problems, ψ is quadratic and the exact minimiser can easily be determined. For non-linear problems, the shape of ψ is more complex and there is not necessarily an analytic formula for the minimiser of ψ. Then an inexact linesearch procedure is used to determine a "good" minimiser of ψ. The objective is to find a value for α ensuring the following conditions called *Wolfe conditions* (figure B.1):

- sufficient decrease condition (also called Armijo condition)

$$\psi(\alpha) \leq \psi(0) + c_1 \alpha \psi'(0), \tag{B.6a}$$

  The scalar coefficient c_1 is positive, hence small values always satisfy this condition.

- curvature condition

$$\psi'(\alpha) \geq c_2 \psi'(0), \tag{B.6b}$$

  with c_2 a positive scalar coefficient. This condition excludes too small values of α which satisfy the first conditions. In the case of the *strong* Wolfe conditions, a more restrictive curvature condition is chosen

$$\big|\psi'(\alpha)\big| \leq c_2 \big|\psi'(0)\big|. \tag{B.6c}$$

  The two positive coefficients c_1 and c_2 should verify $0 < c_1 < c_2 < 1$ to ensure that there exists a step length satisfying both conditions.

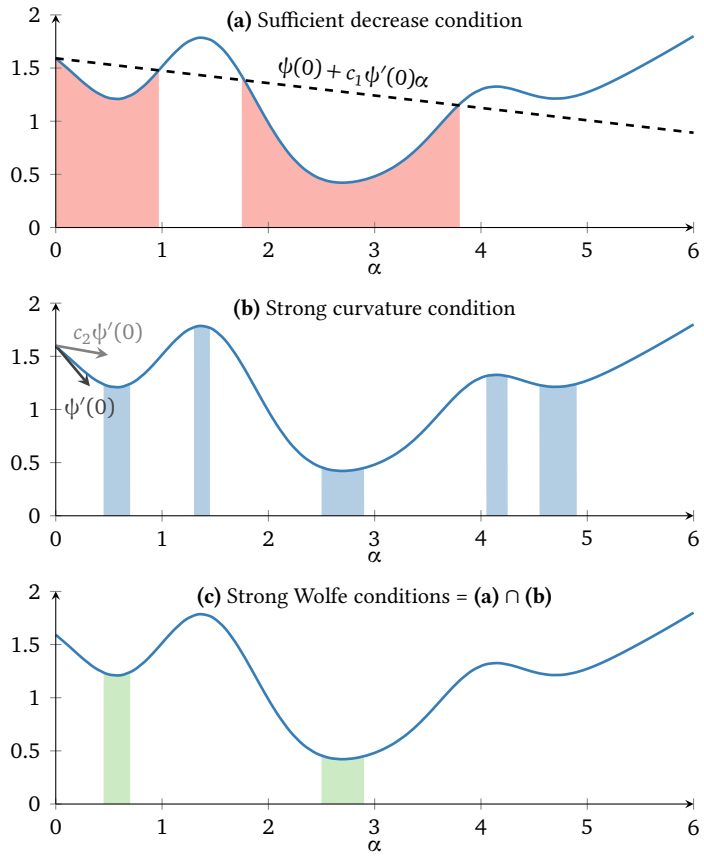


Figure B.1. – The strong Wolfe conditions for a relatively complex scalar function ψ (blue, solid): (a) sufficient decrease (or Armijo) condition; (b) strong curvature condition; (c) combination of the sufficient decrease condition and the strong curvature condition. The coloured areas show the range of acceptable step length values.

The values of the coefficient c_1 and c_2 and the choice of the regular or strong curvature conditions depend on the descent direction strategy (detailed in the next section. For non-linear conjugate gradients methods, the strong Wolfe conditions are required and recommended parameters are $c_1 = 1\times10^{-4}$ and $c_2 = 0.1$. For quasi-Newton methods, the regular curvature condition is sufficient with $c_2 = 0.9$ (Nocedal and Wright, 2006).

Algorithms ensuring the strong Wolfe conditions are relatively expensive because the value and the gradient of the objective function should be calculated at each trial step size. Therefore the choice of successive trial steps should be efficient. Efficient strategies usually consist of two steps. First an interval containing an acceptable step size is determined. Then this interval is reduced iteratively until an acceptable step size is found. For our implementation, we have chosen the strategy of Moré and Thuente (1994) as recommended by Nocedal and Wright (2006, p. 162), using polynomial interpolations to compute the successive step size.

B.3. Descent direction

We review classic strategies used to determine a descent direction.

Steepest Descent
The steepest descent is the simplest strategy. The descent direction is simply the opposite of the gradient:

$$d^{(n)} = -g^{(n)}. \tag{B.7}$$

It obviously satisfies the descent condition (B.2). It is a natural choice because it is the direction providing the maximum decrease of J_0 in the neighbourhood of $\xi^{(n)}$. However other strategies provide faster decrease of the objective function.

Non-linear conjugate-gradient
The conjugate-gradient algorithm originally designed for linear problems can be extended to non-linear optimisation. The descent direction is based on the steepest descent but an additional term takes into account the value of the gradient obtained in the previous iterations,

$$d^{(n)} = -g^{(n)} + \beta^{(n)} d^{(n-1)}. \tag{B.8}$$

Many formulas have been proposed for the definition of β (Hager and Zhang, 2006). The most popular are

$$\beta_{\text{FR}}^{(n)} = \frac{\left\langle g^{(n)} \,\middle|\, g^{(n)} \right\rangle_{\mathscr{E}}}{\left\langle g^{(n-1)} \,\middle|\, g^{(n-1)} \right\rangle_{\mathscr{E}}} \quad \text{(Fletcher and Reeves, 1964)} \tag{B.9a}$$

and

$$\beta_{\text{PR}}^{(n)} = \frac{\left\langle g^{(n)} \,\middle|\, g^{(n)} - g^{(n-1)} \right\rangle_{\mathscr{E}}}{\left\langle g^{(n-1)} \,\middle|\, g^{(n-1)} \right\rangle_{\mathscr{E}}} \quad \text{(Polak and Ribiere, 1969).} \tag{B.9b}$$

The Fletcher-Reeves formula is actually the one implemented in the linear version of the algorithm. In the linear case, successive gradients are orthogonal and the Polak-Ribière formula is equivalent to the Fletcher-Reeves one.

The descent condition (B.2) reads

$$\left\langle d^{(n)} \,\middle|\, g^{(n)} \right\rangle_{\mathscr{E}} = -\left\| g^{(n)} \right\|_{\mathscr{E}}^2 + \beta^{(n)} \left\langle d^{(n-1)} \,\middle|\, g^{(n)} \right\rangle_{\mathscr{E}}. \tag{B.10}$$

Then an exact linesearch automatically ensures that $d^{(n)}$ is a descent direction (equation B.4). More generally a step length ensuring the strong Wolfe conditions is sufficient (Nocedal and Wright, 2006, p. 122). The Polak-Ribière formula is the one providing the faster convergence in most cases (Nocedal and Wright, 2006, p. 131). It is the one used in this study.

Newton method
Newton strategies are based on a second-order Taylor expansion of the objective function

$$J_0(\xi^{(n)} + \alpha^{(n)} d^{(n)}) = J_0(\xi^{(n)}) + \alpha^{(n)} \left\langle d^{(n)} \,\middle|\, g^{(n)} \right\rangle_{\mathscr{E}} + \frac{1}{2} \left(\alpha^{(n)}\right)^2 \left\langle d^{(n)} \,\middle|\, H^{(n)} d^{(n)} \right\rangle_{\mathscr{E}}, \tag{B.11}$$

where we have noted H the Hessian matrix of J_0,

$$H^{(n)}(\xi^{(n)}) = \frac{\partial^2 J_0}{\partial \xi^2}(\xi^{(n)}). \tag{B.12}$$

Deriving equation (B.11) with respect to $d^{(n)}$ leads to the *Newton equation*

$$H^{(n)} d^{(n)} = -g^{(n)}. \tag{B.13}$$

The Newton direction $d^{(n)}$ is the solution of this equation, which reads, provided that the Hessian matrix is invertible

$$d^{(n)} = -\left(H^{(n)}\right)^{-1} g^{(n)}. \tag{B.14}$$

The descent condition equation (B.2) reads

$$\langle d^{(n)} \,|\, g^{(n)} \rangle_{\mathscr{E}} = -\left\langle g^{(n)} \,\middle|\, \left(H^{(n)}\right)^{-1} g^{(n)} \right\rangle, \tag{B.15}$$

therefore if H is positive definite, the Newton direction is a descent direction. If it is not positive definite, it may not be invertible and may not define a descent direction.

Quasi-Newton methods

The use of a Newton descent requires the computation of the Hessian matrix and its inversion, which are both computer expensive procedures. *Quasi-Newton* methods are designed to alleviate this difficulty. They use an approximation of the inverse of the Hessian $B^{(n)} \approx \left(H^{(n)}\right)^{-1}$, and the quasi-Newton direction reads

$$d^{(n)} = -B^{(n)} g^{(n)}. \tag{B.16}$$

The steepest descent can actually be interpreted as a quasi-Newton method with the Hessian approximated with the identity operator. A popular choice for the definition of B uses the BFGS formula (named after Broyden, Fletcher, Goldfarb and Shanno) (Nocedal and Wright, 2006, p. 136). Starting with the identity, the approximation of H^{-1} is improved iteratively at each iteration

$$B^{(n+1)} = \left(I - \rho^{(n)} s^{(n)} y^{(n)^T}\right) H^{(n)} \left(I - \rho^{(n)} y^{(n)} s^{(n)^T}\right) + \rho^{(n)} s^{(n)} s^{(n)^T}, \tag{B.17}$$

with $\rho^{(k)} = \frac{1}{\langle y^{(n)} \,|\, s^{(n)} \rangle_{\mathscr{E}}}$. A difficulty is that the huge matrix $B^{(n)}$ has to be kept in memory. This difficulty can be overcome with the *limited-memory* BFGS (l-BFGS) technique (Nocedal, 1980), which computes the value of $B^{(n)}$ from the values of $\xi^{(n)}$ and $g^{(n)}$ at the m preceding iterations (where m typically ranges between 3 and 20). With this method, there is no need to compute nor to store $H^{(n)}$.

An alternative strategy called the truncated Newton method (Nocedal and Wright, 2006, p. 168) aims at resolving the linear system (B.13) iteratively with the classical linear conjugate gradient algorithm. The method needs only an efficient way of computing the application of the Hessian to a vector of $\mathscr{E}$ (matrix-vector product) and there is no storage requirements. See Métivier *et al.* (2013, 2014) for applications of the truncated Newton method in the framework of Full Waveform Inversion.

Appendix C.

Computation of the adjoint of the approximate inverse of the extended Born modelling operator

In section 4.2, we have proposed a formula for an approximate inverse of the extended Born modelling operator $\mathrm{F} : \mathscr{E} \mapsto \mathscr{D}_{\text{obs}}$. In section 4.3, this approximate inverse, noted $\mathrm{F}^\dagger : \mathscr{D}_{\text{obs}} \mapsto \mathscr{E}$, is used as a preconditioner to improve the convergence rate of the iterative migration process. The derivation involves the adjoint of the approximate inverse, noted here $(\mathrm{F}^\dagger)^{\mathrm{T}} : \mathscr{E} \mapsto \mathscr{D}_{\text{obs}}$. In this appendix, we derive an expression for $(\mathrm{F}^\dagger)^{\mathrm{T}}$, using the adjoint-state technique. We begin by reviewing the implementation of the approximate inverse $\mathrm{F}^\dagger$.

C.1. Equations for the forward map

The application of the approximate inverse $\mathrm{F}^\dagger$ to a vector P of $\mathscr{D}_{\text{obs}}$ produces a vector ξ^{inv} of $\mathscr{E}$. The implementation is very similar to the case of the adjoint F^{T} described in section 2.5.1. It consists of computing modified source and receiver wavefields, noted here $\overline{\mathrm{P}_0}$ and $\overline{\lambda_1}$ respectively, by analogy with the notations of section 2.5.1. The application of the approximate inverse to P is obtained by solving the following equations,

$$\begin{cases} \mathscr{L}_0 \overline{\mathrm{P}_0} = \overline{\mathrm{S}}, & \text{(C.1a)} \\ \mathscr{L}_0^* \overline{\lambda_1} = \mathrm{W_r P}, & \text{(C.1b)} \\ \xi_0 = \overline{\mathrm{Q}}[\overline{\mathrm{P}_0}, \overline{\lambda_1}], & \text{(C.1c)} \\ \xi_{\text{inv}} = \mathrm{W_{mod}} \xi_0, & \text{(C.1d)} \end{cases}$$

where we defined

- the modified extended cross-correlation operator $\overline{\mathrm{Q}}[u, v] : \mathscr{D} \times \mathscr{D} \mapsto \mathscr{E}$, defined for $(u, v) \in \mathscr{D}$ as

$$\left[\overline{\mathrm{Q}}[u, v]\right](\mathbf{x}, h) = \int_s \int_\omega \frac{1}{(\mathrm{i}\omega)} u^*(s, \mathbf{x} - h, \omega) v(s, \mathbf{x} + h, \omega) \, \mathrm{d}\omega \, \mathrm{d}s. \qquad \text{(C.2)}$$

  It is similar to the usual operator $\mathrm{Q}[u, v]$ (equation 2.38), except that a first order integration replaces the second order derivative;

- the weighting operator $\mathrm{W}_{\mathrm{mod}} : \mathscr{E} \mapsto \mathscr{E}$, which consists of a derivative with respect to z and a multiplication by a power α of c_0,

$$\Big[\mathrm{W}_{\mathrm{mod}}\xi(\mathbf{x},h)\Big](\mathbf{x},h) = -\frac{k}{c_0^\alpha(\mathbf{x})}\frac{\partial}{\partial z}\Big[\xi(\mathbf{x},h)\Big], \tag{C.3}$$

  If ξ is defined as a velocity perturbation $\xi = \delta c/(2c_0)$, then $k = 8$ and $\alpha = 0$; if ξ is defined as a squared slowness perturbation $\xi = 2\delta c/c_0^3$, then $k = 32$ and $\alpha = 2$.

  The vertical derivative in this weight can be interpreted under the high-frequency approximation as a multiplication by the cosine of the half-opening angle and the cosine of the dip angle at the image point (figure C.1),

$$\frac{\mathrm{i}\omega}{c_0(\mathbf{x})}\cos\theta\cos\varphi; \tag{C.4}$$

- the weighting operator $\mathrm{W}_{\mathrm{r}} : \mathscr{D}_{\mathrm{obs}} \mapsto \mathscr{D}$, which in practice constructs a dipole source term around the depth z_r of the receivers. Its practical implementation is

$$\Big[\mathrm{W}_{\mathrm{r}}\mathrm{P}(s,r,\omega)\Big]\big(s,\mathbf{x}=(x,z),\omega\big) = \begin{cases} -\dfrac{1}{2\Delta z}\mathrm{P}(s,r,\omega) & \text{if } z = z_r - 1, \\ +\dfrac{1}{2\Delta z}\mathrm{P}(s,r,\omega) & \text{if } z = z_r + 1, \\ 0 & \text{elsewhere,} \end{cases} \tag{C.5}$$

  with Δz the size of a grid point in the z direction and z_r the position of the receivers. It can be interpreted as a multiplication by the cosine of the take-off angle at the receiver position (figure C.1),

$$\frac{\mathrm{i}\omega}{c_0(r)}\cos\beta_r; \tag{C.6}$$

- Finally, $\overline{\mathrm{S}}$ is a modified source term, which involves a deconvolved version $\widetilde{\Omega}$ of the source wavelet Ω and an other weighting operator W_{s} defined in the same way as W_{r} for the source depth z_s

$$\overline{\mathrm{S}} = \mathrm{W}_{\mathrm{s}}\Big[\delta(\mathbf{x}-s)\widetilde{\Omega}\Big]. \tag{C.7}$$

  W_{s} can be interpreted as a multiplication by the cosine of the take-off angle at the source position (figure C.1),

$$\frac{\mathrm{i}\omega}{c_0(s)}\cos\beta_s. \tag{C.8}$$

C.2. Equations for the adjoint map

We want to compute the adjoint $(\mathrm{F}^\dagger)^{\mathrm{T}}$ of the operator $\mathrm{F}^\dagger$. It is applied to a vector χ of $\mathscr{E}$ and computes a vector $\mathrm{Q} = (\mathrm{F}^\dagger)^{\mathrm{T}}\chi$ in $\mathscr{D}_{\mathrm{obs}}$. We use the adjoint-state method and define a scalar objective function

$$\Gamma(\chi,\mathrm{P}) = \langle\chi\,|\,\mathrm{F}^\dagger\mathrm{P}\rangle_{\mathscr{E}}, \tag{C.9}$$

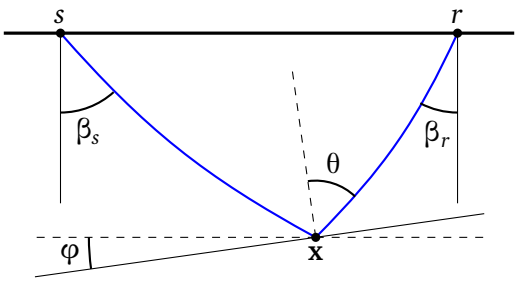


Figure C.1. – Angles involved in the interpretation of the weights defining the inverse operator $F^\dagger$.

such that

$$\frac{\partial \Gamma}{\partial P}(\chi, P) = (F^\dagger)^T \chi. \tag{C.10}$$

The state equations are given by equations (C.1). We note $\xi_{inv} = F^\dagger P$ and define the Lagrangian with the adjoint states η, η_0, $\overline{\mu_0}$ and $\overline{\nu_1}$,

$$\begin{aligned}\overline{\Gamma} = \langle \chi \,|\, \xi_{inv} \rangle_{\mathscr{E}} &- \langle \eta \,|\, \xi_{inv} - W_{mod} \xi_0 \rangle_{\mathscr{E}} \\ &- \langle \eta_0 \,|\, \xi_0 - \overline{Q}[\overline{P_0}, \overline{\lambda_1}] \rangle_{\mathscr{E}} \\ &- \langle \overline{\mu_0} \,|\, \mathscr{L}_0 \overline{P_0} - \overline{S} \rangle_{\mathscr{D}} \\ &- \langle \overline{\nu_1} \,|\, \mathscr{L}_0^* \overline{\lambda_1} - W_r P \rangle_{\mathscr{D}}\end{aligned} \tag{C.11}$$

Deriving the Lagrangian with respect to the state variables $(\xi_{inv}, \xi_0, \overline{P_0}, \overline{\lambda_1})$ gives the adjoint equations

$$\begin{cases} \eta = \chi, & \text{(C.12a)} \\ \eta_0 = W_{mod}^* \eta, & \text{(C.12b)} \\ \mathscr{L}_0^* \overline{\mu_0} = \overline{K^+}[\overline{\lambda_1}, \eta_0], & \text{(C.12c)} \\ \mathscr{L}_0^* \overline{\nu_1} = \overline{K^-}[\overline{P_0}, \eta_0], & \text{(C.12d)} \end{cases}$$

and the derivatives of $\overline{\Gamma}$ with respect to P gives the application of $(F^\dagger)^T$ to χ,

$$(F^\dagger)^T \chi = \frac{\partial \overline{\Gamma}}{\partial P} = W_r^* \overline{\nu}_1. \tag{C.13}$$

These equations involve the adjoint of the weighting operators defined in the previous section,

- the operators $\overline{K^-}[P, \xi]$ and $\overline{K^+}[P, \xi] : \mathscr{D} \times \mathscr{E} \mapsto \mathscr{D}$, defined for $P \in \mathscr{D}$ and $\xi \in \mathscr{E}$ as

$$\overline{K^-}[P, \xi](s, \mathbf{x}, \omega) = \int_h \frac{1}{(i\omega)} P(s, \mathbf{x} - 2h, \omega) \xi(\mathbf{x} - h, h) \, dh, \tag{C.14a}$$

$$\overline{K^+}[P, \xi](s, \mathbf{x}, \omega) = \int_h \frac{1}{(i\omega)} P(s, \mathbf{x} + 2h, \omega) \xi(\mathbf{x} + h, h) \, dh. \tag{C.14b}$$

Note that the integration operator is causal in equation (C.14a) and anti-causal in equation (C.14b) (Claerbout, 2014). They are similar to operators K^- and K^+ defined in equations (2.19) and (2.35), except that the time integration replaces the second order time derivative;

- the adjoint $\mathrm{W}_{\mathrm{mod}}^* : \mathscr{E} \mapsto \mathscr{E}$ of $\mathrm{W}_{\mathrm{mod}}$, defined as

$$\Big[\mathrm{W}_{\mathrm{mod}}^* \xi(\mathbf{x},h)\Big](\mathbf{x},h) = k\frac{\partial}{\partial z}\Big[\frac{1}{c_0^\alpha(\mathbf{x})}\xi(\mathbf{x},h)\Big], \tag{C.15}$$

- and the adjoint $\mathrm{W}_{\mathrm{r}}^* : \mathscr{D} \mapsto \mathscr{D}_{\mathrm{obs}}$ of W_{r}, defined as

$$\Big[\mathrm{W}_{\mathrm{r}}^* \mathrm{P}(s,\mathbf{x},\omega)\Big](s,r,\omega) = \frac{1}{2\Delta z}\Big[\mathrm{P}(s,(x,z_r+1),\omega) - \mathrm{P}(s,(x,z_r-1),\omega)\Big]. \tag{C.16}$$

Note that only the value of $\overline{\nu_1}$ is required and $\overline{\mu_0}$ does not need to be computed. If we remove the weighting operators, that is $\mathrm{W}_{\mathrm{mod}} = \mathrm{I}_{\mathscr{E}}$, $\mathrm{W}_{\mathrm{r}}^* = \mathrm{M}$, $\overline{\mathrm{K}^-} = \mathrm{K}^-$ and $\widetilde{\Omega} = \Omega$, then equations (C.12) are equivalent to the usual first-order extended Born modelling. It means that in terms of implementation, $(\mathrm{F}^\dagger)^{\mathrm{T}}$ and F have the same structure.

Appendix D.

Analytic expression for the MVA gradient after direct inversion in 1D

This appendix is related to section 5.2.2 and details the derivation of the analytic expression of the background velocity update obtained after direct inversion in the 1D case (equations 5.13b, 5.15c and 5.15d) (we do not detail the case of the adjoint (equations 5.13a, 5.15a and 5.15b), which is very similar). The purpose is to obtain an analytic expression providing results similar to those obtained with the adjoint-state method and wave-equations operators (section 4.2.10). Note that the derivation does not assume a homogeneous background velocity model.

We recall the expression of the objective function,

$$J_1^{\text{inv}}[c_0] = \frac{1}{2}\int_z a^2(z) c_0^{2\beta}(z) \xi_{\text{inv}}^2(z)\,\mathrm{d}z, \tag{D.1}$$

where ξ_{inv} is obtained by application of the approximate inverse $F^\dagger$ to observed data (equation 5.7c),

$$\xi_{\text{inv}}(z) = \left[F^\dagger P^{\text{obs}}\right](z) = -4\frac{\partial}{\partial z}\int_\omega \frac{1}{(\mathrm{i}\omega)^*}\widetilde{\Omega}^*(\omega)\left[\frac{\partial}{\partial z_s} G_0^*(s,z,\omega)\right]^2 P(\omega)\,\mathrm{d}\omega, \tag{D.2}$$

In section 5.2.2, we have derived the following expression for the gradient G_{inv} of J_1^{inv} with respect to c_0,

$$G_{\text{inv}}(y) = \beta c_0^{2\beta-1}(y) a^2(y) \xi_{\text{inv}}^2(y) + \int_z Q_{\text{inv}}(y,z)\eta_{\text{inv}}(z)\,\mathrm{d}z \tag{D.3}$$

where η_{inv} is obtained as the partial derivative of J_1 with respect to ξ_{inv},

$$\eta_{\text{inv}}(z) = c_0^{2\beta}(z) a^2(z) \xi_{\text{inv}}(z), \tag{D.4}$$

and $Q_{\text{inv}} : \mathscr{E} \mapsto \mathscr{M}$ is defined as

$$Q_{\text{inv}}(y,z) = \frac{\partial\left[F^\dagger P^{\text{obs}}\right](z)}{\partial c_0(y)}. \tag{D.5}$$

The purpose of this appendix is the derivation of the expression of Q_{inv} given in equation (5.15c).

D.1. Derivation of Q_{inv}

The derivation starts from equation (D.2), hence we need an expression for the derivatives of the Green's function $G_0(s,z,\omega)$ with respect to s and z, and with respect to the value of c_0 at depth y. These expressions are given by,

$$\begin{cases} \dfrac{\partial G_0(s,z,\omega)}{\partial c_0(y)} = \dfrac{2(i\omega)^2}{c_0^3(y)} G_0(s,y,\omega) G_0(y,z,\omega), & \text{(D.6a)} \\ \dfrac{\partial G_0(s,z,\omega)}{\partial z_s} = -\dfrac{(i\omega)}{c_0(s)} G_0(s,z,\omega), & \text{(D.6b)} \\ \dfrac{\partial G_0(s,z,\omega)}{\partial z} = \dfrac{(i\omega)}{c_0(z)} G_0(s,z,\omega). & \text{(D.6c)} \end{cases}$$

The derivative with respect to $c_0(y)$ (D.6a) is given by the first-order Born approximation, reviewed in section 2.3.1: the perturbation $\delta G_0(s,z,\omega)$ of a reference Green's function $G_0(s,z,\omega)$ due to a perturbation δc of the reference background velocity model c_0 obeys (equation 2.6b)

$$\left(\frac{(i\omega)^2}{c_0^2} - \Delta\right) \delta G_0(s,z,\omega) = \frac{2\delta c(z)}{c_0^3(z)} (i\omega)^2 G_0(s,z,\omega), \tag{D.7}$$

and $\delta G_0(s,z,\omega)$ can be expressed as (equation 2.10b),

$$\delta G_0(s,z,\omega) = \int_x (i\omega)^2 G_0(s,x,\omega) \frac{2\delta c(x)}{c_0^3(x)} G_0(y,z,\omega)\, dx. \tag{D.8}$$

Equation (D.6a) is obtained as the derivative of this expression with respect to δc.

We now detail the important steps in the derivation of Q_{inv}. Starting from the expression of $F^\dagger P^{\text{obs}}$ given in equation (D.2), we compute the following quantity,

$$\frac{\partial}{\partial c_0(y)} \left[\left\langle \eta_{\text{inv}} \,\middle|\, F^\dagger P^{\text{obs}} \right\rangle_{\mathcal{E}} \right] \tag{D.9a}$$

$$= -\frac{\partial}{\partial c_0(y)} \left\{ 4 \left\langle \frac{\partial}{\partial z} \eta_{\text{inv}} \,\middle|\, \int_\omega \frac{1}{i\omega} \widetilde{\Omega}^*(\omega) \left[\frac{\partial}{\partial z_s} G_0^*(s,z,\omega) \right]^2 P^{\text{obs}}(\omega)\, d\omega \right\rangle_{\mathcal{E}} \right\}, \tag{D.9b}$$

where we have explicitly separated the term $\partial \eta_{\text{inv}} / \partial z$ as is done in the calculation with the adjoint-state method (equation 4.27g). Then, using equation (D.6b),

$$= -4 \left\langle \frac{\partial}{\partial z} \eta_{\text{inv}} \,\middle|\, \frac{\partial}{\partial c_0(y)} \int_\omega \frac{(i\omega)}{c_0^2(s)} \widetilde{\Omega}^*(\omega) \left[G_0^*(s,z,\omega) \right]^2 P^{\text{obs}}(\omega)\, d\omega \right\rangle_{\mathcal{E}} \tag{D.9c}$$

$$= 16 \left\langle \eta_{\text{inv}} \,\middle|\, \frac{\partial}{\partial z} \int_\omega \frac{(i\omega)^3}{c_0^2(s) c_0^3(y)} \widetilde{\Omega}^*(\omega) G_0^*(s,z,\omega) G_0^*(s,y,\omega) G_0^*(y,z,\omega) P^{\text{obs}}(\omega)\, d\omega \right\rangle_{\mathcal{E}}, \tag{D.9d}$$

where we have used equation (D.6a). Eventually we need to compute the derivative with respect to z of the product of Green's functions inside the integral, whose value depends on the relative position of y and z,

$$\frac{\partial}{\partial z} \left[G_0^*(s,z,\omega) G_0^*(y,z,\omega) \right] = \begin{cases} 2i\omega / c_0(z) & \text{if } y \le z, \\ 0 & \text{if } y \ge z, \end{cases} \tag{D.10}$$

leading to the expression given in equation (5.15c),

$$Q_{\text{inv}}(y,z) = \begin{cases} \dfrac{32}{c_0^2(s)c_0^3(y)c_0(z)} \displaystyle\int_\omega (\mathrm{i}\omega)^4 \widetilde{\Omega}^*(\omega) G_0^*(s,z) G_0^*(s,y) G_0^*(y,z)\,\mathrm{d}\omega & \text{if } y < z, \\ 0 & \text{if } y \geq z, \end{cases} \tag{D.11}$$

where the dependence of the Green's function to ω have been omitted.

D.2. High-frequency expression

For the numerical applications in section 5.2, we have assumed an homogeneous background velocity model ($c_0(z) = c_0$) and replaced the value of the Green's function G_0 in equations (D.2) and (D.11) by its asymptotic expression in a 1D homogeneous medium ($c_0(z) = c_0$),

$$G_0(s,z,\omega) = \frac{c_0}{2\mathrm{i}\omega} \mathrm{e}^{-\mathrm{i}\omega\tau(s,z)}. \tag{D.12}$$

leading to the expressions given in equations (5.9c) and (5.15d)

$$\xi_{\text{inv}}(z) = \frac{2}{c_0} \int_\omega \widetilde{\Omega}^*(\omega) \mathrm{e}^{2\mathrm{i}\omega\tau(s,z)} P(\omega)\,\mathrm{d}\omega, \tag{D.13}$$

$$Q_{\text{inv}}(y,z) = \begin{cases} -\dfrac{4}{c_0^3} \displaystyle\int_\omega (\mathrm{i}\omega) \widetilde{\Omega}^*(\omega) \mathrm{e}^{(\mathrm{i}\omega)\left(\tau(s,z)+\tau(s,y)+\tau(y,z)\right)} P^{\text{obs}}(\omega)\,\mathrm{d}\omega & \text{if } y < z, \\ 0 & \text{otherwise.} \end{cases} \tag{D.14}$$

We check the validity of equation (D.11) on a numerical example by comparing the gradient obtained with this expression with the gradient obtained by finite differences and slightly perturbed model (figure D.1). The latter consists of computing the value of J_1 (equation D.1) for velocity models with small velocity perturbations at position y $c(y) = c_0(y) \pm \delta c_0$. The computation is repeated for each depth y. Note that in this approach, ξ_{inv} is not obtained through the asymptotic expression (D.13), but by solving wave-equations as described in appendix C.1.

It is actually important to note that equation (D.14) has been obtained by replacing the value of the Green's functions by their high-frequency approximation (equation D.12) after the calculation of the derivative with respect to c_0. One would obtain a different expression by applying the derivative with respect to c_0 to a high-frequency approximation of $F^\dagger P^{\text{obs}}$ such as equation (D.13). In particular, operator Q_{inv} would not be upper triangular any more.

To illustrate this point, we consider the derivative of the Green's function $G(s,z,\omega)$ with respect to $c_0(y)$. Applying the derivative first leads to equation (D.6a) and then replacing the Green's function gives

$$\frac{\partial G_0(s,z,\omega)}{\partial c_0(y)} = \frac{2}{c_0} \mathrm{e}^{-(\mathrm{i}\omega)(\tau(s,y)+\tau(y,z))}. \tag{D.15}$$

On the other hand, computing the derivative of the high-frequency approximation (D.6a) with respect to $c_0(y)$ leads to

$$\frac{\partial G_0(s,z,\omega)}{\partial c_0(y)} = \begin{cases} \dfrac{2}{c_0} \mathrm{e}^{-(\mathrm{i}\omega)(\tau(s,y)+\tau(y,z))} & \text{if } y < z, \\ 0 & \text{if } y \geq z, \end{cases} \tag{D.16}$$

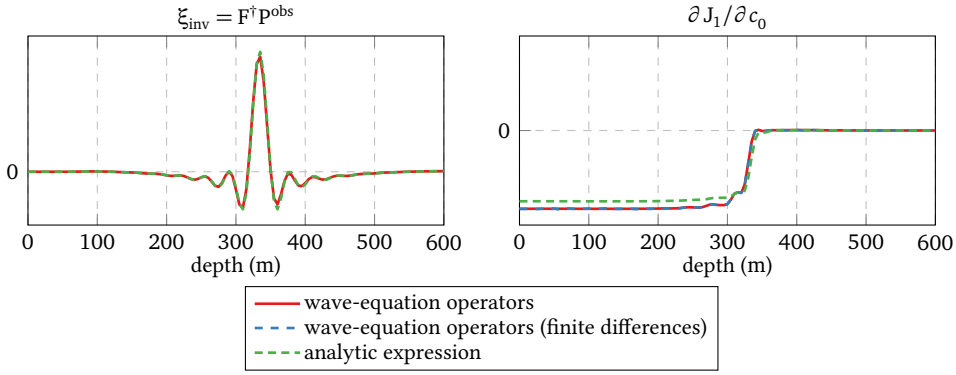


Figure D.1. – Comparison of the reflectivity (left) and background velocity update (right) obtained with the analytic expressions (equations D.13 and D.14) (green dashed line) with the results obtained by solving wave-equations with a finite differences code (red) as described in section 4.2.11. A third value for the gradient (blue) is obtained by computing the value of ξ_{inv} with wave-equations propagation for slightly perturbed velocity model. For this example, we took $\beta = 3/2$ and use the same reflectivity and background velocity model as in section 5.2, that is homogeneous exact and initial velocity models (3000 m/s and 2500 m/s) and an exact reflectivity model made of a single reflector at 400 m depth.

as the derivative of the traveltime $\tau(s,z)$ with respect to $c_0(y)$ reads

$$\frac{\partial \tau(s,z)}{\partial c_0(y)} = \begin{cases} -\dfrac{1}{c_0^2(y)} & \text{if } y < z \\ 0 & \text{if } y \geq z. \end{cases} \tag{D.17}$$

The main difference is that the high-frequency approximation of the Green's function between s and z is insensitive to a perturbation below z, whereas the general Green's function may be affected by a perturbation of the velocity model below z as a diffracting point for example (figure D.2, right).

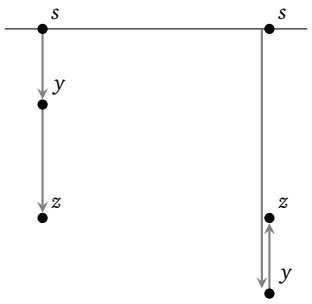


Figure D.2. – Perturbation of the Green's function $G_0(s,z,\omega)$ giving the response in z to a source in s, when the perturbation is located above z (left) and below z (right).

Appendix E.

Product of the Hessian of the MVA objective function with a vector of the $\mathscr{M}$-space

In this appendix we consider the case of non-linear iterations on the background velocity model c_0 to determine a minimum of the MVA objective function $J_1[c_0]$. In the following calculations, we consider only the case of direct inversion of primary reflections, the reflectivity image ξ being defined by the application of operator $F^\dagger$ to observed data P^{obs}. We keep a very general notation for the expression of J_1,

$$J_1[c_0] = f_{J_1}[c_0, \xi] \qquad \text{with} \qquad \xi = F^\dagger P^{obs}, \tag{E.1}$$

so that further expressions are applicable to the standard definition $J_1[c_0] = \|A\xi\|^2_{\mathscr{E}}$, to a normalised version $J_1[c_0] = \|A\xi\|^2_{\mathscr{E}} / \|\xi\|^2_{\mathscr{E}}$ and to similar versions with the introduction of a power of the background velocity model c_0^β as proposed in section 4.2.

In Newton optimisation methods (appendix B.3), the descent direction $d \in \mathscr{M}$ is defined as the solution of the following linear system, called Newton equation,

$$\left[\frac{\partial^2 J_1}{\partial c_0^2}\left[c_0^{(n)}\right]\right] d = -\frac{\partial J_1}{\partial c_0}\left[c_0^{(n)}\right], \tag{E.2}$$

where $c_0^{(n)}$ is the value of the background velocity model at the nth non-linear iteration. Most application of MVA do not solve this equation and use a quasi-Newton such as l-BFGS. In Liu *et al.* (2014b) and Shen and Symes (2015), d is defined as the opposite of the gradient element-wise divided by the product of the Hessian with the unit vector of the $\mathscr{M}$-space. One could also use the truncated-Newton method, already studied in the framework of Full Waveform Inversion by Métivier *et al.* (2013). It consists of iteratively solving equation (E.2) with the linear conjugate-gradient algorithm, without explicitly computing the full Hessian matrix. Then several products of the Hessian matrix with a vector $V \in \mathscr{M}$ are required. An efficient procedure to compute this product should then be defined.

We propose here a method based on the adjoint-state method (Plessix, 2006), very similar to the one used by Métivier *et al.* (2013) for FWI and to the method presented in section 3.2.3 to compute the Hessian-vector product in the case of the migration objective function c_0. Given a vector V of $\mathscr{M}$, we define the scalar objective function $\Gamma[c_0, V]$ as

$$\Gamma[c_0, V] = \left\langle \frac{\partial J_1}{\partial c_0} \middle| V \right\rangle_{\mathscr{M}}, \tag{E.3}$$

such that the derivative of Γ with respect to c_0 gives the desired matrix-vector product,

$$\frac{\partial \Gamma}{\partial c_0}[c_0, \mathrm{V}] = \left[\frac{\partial^2 \mathrm{J}_1}{\partial c_0^2}[c_0]\right]\mathrm{V}. \tag{E.4}$$

Noting G the gradient of J_1 with respect to c_0, we define the Lagrangian by defining constraints on the state equations solved to determine the gradient with the adjoint variables $\gamma \in \mathscr{M}$, $(\Xi_0, \Xi, \mathrm{E}, \mathrm{E}_0) \in \mathscr{E}$ and $(\Pi_0, \mathrm{L}_1, \mathrm{N}_1, \mathrm{M}_0) \in \mathscr{D}$,

$$\begin{aligned}
\overline{\Gamma} = \left\langle \mathrm{G} \middle| \mathrm{V} \right\rangle_{\mathscr{M}} &- \left\langle \gamma \middle| \mathrm{G} - \frac{\partial f_{\mathrm{J}_1}}{\partial c_0}[c_0, \xi] - \mathrm{C}[\mu_0, \mathrm{P}_0] - \mathrm{C}[\nu_1, \lambda_1] - f_{\mathrm{mod}}[c_0, \eta, \xi_0] \right\rangle_{\mathscr{M}} \\
&- \left\langle \Pi_0 \middle| \mathscr{L}_0 \mathrm{P}_0 - \overline{\mathrm{S}} \right\rangle_{\mathscr{D}} \\
&- \left\langle \mathrm{L}_1 \middle| \mathscr{L}_0^* \lambda_1 - \mathrm{W}_{\mathrm{r}} \mathrm{M}^{\mathrm{T}} \mathrm{P}^{\mathrm{obs}} \right\rangle_{\mathscr{D}} \\
&- \left\langle \Xi_0 \middle| \xi_0 - \overline{\mathrm{Q}}[\mathrm{P}_0, \lambda_1] \right\rangle_{\mathscr{D}} \\
&- \left\langle \Xi \middle| \xi - \mathrm{W}_{\mathrm{mod}} \xi_0 \right\rangle_{\mathscr{E}} \\
&- \left\langle \mathrm{E} \middle| \eta - \frac{\partial f_{\mathrm{J}_1}}{\partial \xi}[c_0, \xi] \right\rangle_{\mathscr{E}} \\
&- \left\langle \mathrm{E}_0 \middle| \eta_0 - \mathrm{W}_{\mathrm{mod}}^* \eta \right\rangle_{\mathscr{E}} \\
&- \left\langle \mathrm{N}_1 \middle| \mathscr{L}_0 \nu_1 - \overline{\mathrm{K}^-}[\mathrm{P}_0, \eta_0] \right\rangle_{\mathscr{D}} \\
&- \left\langle \mathrm{M}_0 \middle| \mathscr{L}_0^* \mu_0 - \overline{\mathrm{K}^+}[\lambda_1, \eta_0] \right\rangle_{\mathscr{D}},
\end{aligned} \tag{E.5}$$

where C, $\overline{\mathrm{Q}}$, $\overline{\mathrm{K}^-}$, $\overline{\mathrm{K}^+}$, $\mathrm{W}_{\mathrm{mod}}$, $\mathrm{W}_{\mathrm{mod}}^*$, W_{r} and $\overline{\mathrm{S}}$ are defined in equations (3.19), (C.2), (C.3), (C.5), (C.7), (C.14a), (C.14b) and (C.15), and $f_{\mathrm{mod}} : (\mathscr{M} \times \mathscr{E} \times \mathscr{E}) \mapsto \mathscr{M}$ is defined as

$$f_{\mathrm{mod}}[c_0, \eta, \xi_0](\mathbf{x}) = k\alpha c_0^{\alpha-1}(\mathbf{x}) \int_h \eta(\mathbf{x}, h) \frac{\partial \xi_0}{\partial z}(\mathbf{x}, h)\, \mathrm{d}h, \tag{E.6}$$

with the value of k and α defined in equation (C.3). Deriving the Lagrangian with respect to the state variable μ_0, ν_1, η_0, η, ξ, ξ_0, λ_1, P_0 yields the following sets of adjoint equations

$$\begin{cases}
\mathscr{L}_0 \mathrm{M}_0 = \mathrm{B}[\mathrm{P}_0, \mathrm{V}] & \text{(E.7a)} \\
\mathscr{L}_0^* \mathrm{N}_1 = \mathrm{B}[\lambda_1, \mathrm{V}] & \text{(E.7b)} \\
\mathrm{E}_0 = \overline{\mathrm{Q}}[\mathrm{P}_0, \mathrm{N}_1] + \overline{\mathrm{Q}}[\mathrm{M}_0, \lambda_1] & \text{(E.7c)} \\
\mathrm{E} = \mathrm{W}_{\mathrm{mod}} \mathrm{E}_0 + \left[\dfrac{\partial f_{\mathrm{mod}}}{\partial \eta}[c_0, \eta, \xi_0]\right]\mathrm{V} & \text{(E.7d)}
\end{cases}$$

$$\begin{cases}
\Xi = \left[\dfrac{\partial^2 f_{\mathrm{J}_1}}{\partial \xi^2}[c_0, \xi]\right]\mathrm{E} + \left[\dfrac{\partial^2 f_{\mathrm{J}_1}}{\partial c_0 \partial \xi}[c_0, \xi]\right]\mathrm{V} & \text{(E.8a)} \\
\Xi_0 = \mathrm{W}_{\mathrm{mod}}^* \mathrm{E}_0 + \left[\dfrac{\partial f_{\mathrm{mod}}}{\partial \xi_0}[c_0, \eta, \xi_0]\right]\mathrm{V} & \text{(E.8b)} \\
\mathscr{L}_0 \mathrm{L}_1 = \mathrm{B}[\nu_1, \mathrm{V}] + \overline{\mathrm{K}^-}[\mathrm{P}_0, \Xi_0] + \overline{\mathrm{K}^-}[\mathrm{M}_0, \eta_0] & \text{(E.8c)} \\
\mathscr{L}_0^* \Pi_0 = \mathrm{B}[\mu_0, \mathrm{V}] + \overline{\mathrm{K}^+}[\lambda_1, \Xi_0] + \overline{\mathrm{K}^+}[\mathrm{N}_1, \eta_0], & \text{(E.8d)}
\end{cases}$$

where operator $\mathrm{B} : \mathscr{D} \times \mathscr{M} \mapsto \mathscr{D}$ is defined for $u \in \mathscr{D}$ and $\mathrm{V} \in \mathscr{M}$ as

$$\mathrm{B}[u, \mathrm{V}](s, \mathbf{x}, \omega) = 2(\mathrm{i}\omega)^2 \frac{\mathrm{V}(\mathbf{x})}{c_0^3(\mathbf{x})} u(s, \mathbf{x}, \omega). \tag{E.9}$$

Eventually the desired matrix-vector product is obtained as the gradient of Γ with respect to c_0 and reads

$$\begin{aligned}\left[\frac{\partial^2 \mathrm{J}_1}{\partial c_0^2}[c_0]\right]\mathrm{V} =& \frac{\partial \Gamma}{\partial c_0} \\ =& \mathrm{C}[\Pi_0, \mathrm{P}_0] + \mathrm{C}[\mathrm{L}_1, \lambda_1] + \mathrm{C}[\mathrm{N}_1, \nu_1] + \mathrm{C}[\mathrm{M}_0, \mu_0] \\ & - \frac{3\mathrm{V}}{c_0}\big(\mathrm{C}[\mu_0, \mathrm{P}_0] + \mathrm{C}[\nu_1, \lambda_1]\big) \\ & + \left[\frac{\partial^2 f_{\mathrm{J}_1}}{\partial c_0^2}[c_0, \xi]\right]\mathrm{V} + \left[\frac{\partial f_{\mathrm{mod}}}{\partial c_0}[c_0, \eta, \xi_0]\right]\mathrm{V} + \left[\frac{\partial^2 f_{\mathrm{J}_1}}{\partial \xi \partial c_0}[c_0, \xi]\right]\mathrm{E}.\end{aligned} \tag{E.10}$$

We have divided the adjoint equations into two sets. In the first group of equation, a vector $\mathrm{E} \in \mathscr{E}$ is determined through computations very similar to those used to determine ξ. The second set of equations looks like the one used to determine the gradient $\partial \mathrm{J}_1 / \partial c_0$. Although only four "new" wavefields have to be computed $(\Pi_0, \mathrm{L}_1, \mathrm{M}_0, \mathrm{N}_1)$, the approach is quite expensive because 4D wavefields arrays are too large to be kept in memory and need to be recomputed, first to determine the value of E, then to compute the remaining contributions of the matrix-vector product.

compute $\xi = \mathrm{F}^\dagger \mathrm{P}^{\mathrm{obs}}$ and evaluate $\mathrm{J}_1[c_0]$	2	P_0, λ_1
given ξ, compute $\dfrac{\partial \mathrm{J}_1}{\partial c_0}[c_0]$	4	P_0, λ_1, ν_1, μ_0
given ξ and $\dfrac{\partial \mathrm{J}_1}{\partial c_0}[c_0]$, compute $\left[\dfrac{\partial \mathrm{J}_1}{\partial c_0}[c_0]\right]\mathrm{V}$	12	$2\big(\mathrm{P}_0, \lambda_1, \mathrm{M}_0, \mathrm{N}_1\big)$, ν_1, μ_0, L_1, Π_0

Table E.1. – Number of wave-equations to be solved to compute the value of the MVA objective function in the direct inversion case, its gradient and the product of its Hessian with a vector of $\mathrm{V} \in \mathscr{M}$. The number indicated here should be multiplied by the number of source positions considered in the acquisition.

References

Nota Bene: in the electronic version of the document, the title of the references is clickable and refers to the editor webpage (according to their DOI when available) where they can be downloaded.

Abma, R., N. Kabir, K. H. Matson, S. Michell, S. A. Shaw and B. McLain, 2005. Comparisons of Adaptive Subtraction Methods for Multiple Attenuation. The Leading Edge, **24**(3), 277–280 (cit. on p. 38).

Aki, K. and P. G. Richards, 2002. Quantitative Seismology. 2nd ed. University Science Books (cit. on p. 7).

Alkhalifah, T., 2005. Tau-Migration and Velocity Analysis: Application to Data from the Red Sea. Geophysical Prospecting, **53**(5), 643–653 (cit. on pp. 23, 111).

Alkhalifah, T. and Z. Wu, 2016a. Migration Velocity Analysis Using Pre-Stack Wave Fields: Migration Velocity Analysis. Geophysical Prospecting, online (cit. on p. 111).

– 2016b. The Natural Combination of Full and Image-Based Waveform Inversion. Geophysical Prospecting, **64**(1), 19–30 (cit. on p. 18).

Anandakrishnan, S., D. D. Blankenship, R. B. Alley and P. L. Stoffa, 1998. Influence of Subglacial Geology on the Position of a West Antarctic Ice Stream from Seismic Observations. Nature, **394**(6688), 62–65 (cit. on p. 8).

Asnaashari, A., R. Brossier, S. Garambois, F. Audebert, P. Thore and J. Virieux, 2013. Regularized Seismic Full Waveform Inversion with Prior Model Information. Geophysics, **78**(2), R25–R36 (cit. on p. 9).

Batany, Y.-M., 2016. Signal Separation in Convolutive Mixtures: Contributions to Blind Separation of Sparse Sources and Adaptive Subtraction of Seismic Multiples. École nationale supérieure des Mines de Paris (cit. on p. 38).

Batany, Y.-M., L. T. Duarte, D. Donno, J. M. T. Romano and H. Chauris, 2016. Adaptive Multiple Subtraction: Unification and Comparison of Matching Filters Based on the ℓq-Norm and Statistical Independence. Geophysics, **81**(1), V43–V54 (cit. on p. 38).

Behura, J., K. Wapenaar and R. Snieder, 2014. Autofocus Imaging: Image Reconstruction Based on Inverse Scattering Theory. Geophysics, **79**(3), A19–A26 (cit. on p. 42).

Bell, B. M. and J. V. Burke, 2008. Algorithmic Differentiation of Implicit Functions and Optimal Values. Ed. by C. H. Bischof, H. M. Bücker, P. Hovland, U. Naumann and J. Utke. Red. by T. J. Barth, M. Griebel, D. E. Keyes, R. M. Nieminen, D. Roose and T. Schlick. Vol. 64. Springer Berlin Heidelberg, 67–77 (cit. on p. 83).

Bérenger, J.-P., 1994. A Perfectly Matched Layer for the Absorption of Electromagnetic Waves. Journal of Computational Physics, **114**(2), 185–200 (cit. on pp. 55, 125).

Berkhout, A. J., 2012. Combining Full Wavefield Migration and Full Waveform Inversion, a Glance into the Future of Seismic Imaging. Geophysics, **77**(2), S43–S50 (cit. on p. 42).

– 2014a. Review Paper: An Outlook on the Future of Seismic Imaging, Part I: Forward and Reverse Modelling. Geophysical Prospecting, **62**(5), 911–930 (cit. on p. 36).

– 2014b. Review Paper: An Outlook on the Future of Seismic Imaging, Part II: Full-Wavefield Migration. Geophysical Prospecting, **62**(5), 931–949 (cit. on p. 42).

– 2014c. Review Paper: An Outlook on the Future of Seismic Imaging, Part III: Joint Migration Inversion. Geophysical Prospecting, **62**(5), 950–971 (cit. on p. 43).

Berkhout, A. J. and D. J. Verschuur, 1994. Multiple Technology: Part 2, Migration of Multiple Reflections. 1994 SEG Annual Meeting, 1497–1500 (cit. on p. 41).

Berkhout, A. J. and D. J. Verschuur, 1997. Estimation of Multiple Scattering by Iterative Inversion, Part I: Theoretical Considerations. Geophysics, **62**(5), 1586–1595 (cit. on pp. 38, 39).
– 2003. Transformation of Multiples into Primary Reflections. 2003 SEG Annual Meeting, 1925–1928 (cit. on p. 41).
– 2006. Imaging of Multiple Reflections. Geophysics, **71**(4), SI209–SI220 (cit. on p. 41).
Beylkin, G., 1985. Imaging of Discontinuities in the Inverse Scattering Problem by Inversion of a Causal Generalized Radon Transform. Journal of Mathematical Physics, **26**(1), 99 (cit. on pp. 15, 132).
Billette, F. and S. Brandsberg-Dahl, 2005. The 2004 BP Velocity Benchmark. 67th EAGE Conference & Exhibition, B035 (cit. on p. 125).
Billette, F., 1998. Estimation de Macro Modèles de Vitesse En Sismique Réflexion Par Stéréotomographie. École nationale supérieure des Mines de Paris (cit. on p. 11).
Billette, F. and G. Lambaré, 1998. Velocity Macro-Model Estimation from Seismic Reflection Data by Stereotomography. Geophysical Journal International, **135**(2), 671–690 (cit. on p. 17).
Biondi, B. and A. Almomin, 2014. Simultaneous Inversion of Full Data Bandwidth by Tomographic Full-Waveform Inversion. Geophysics, **79**(3), WA129–WA140 (cit. on p. 111).
Biondi, B. and W. W. Symes, 2004. Angle-domain Common-image Gathers for Migration Velocity Analysis by Wavefield-continuation Imaging. Geophysics, **69**(5), 1283–1298 (cit. on pp. 21, 112).
Bishop, T., K. Bube, R. Cutler, R. Langan, P. Love, J. Resnick, R. Shuey, D. Spindler and H. Wyld, 1985. Tomographic Determination of Velocity and Depth in Laterally Varying Media. Geophysics, **50**(6), 903–923 (cit. on pp. 11, 16).
Blazek, K. D., C. C. Stolk and W. W. Symes, 2013. A Mathematical Framework for Inverse Wave Problems in Heterogeneous Media. Inverse Problems, **29**(6), 065001 (cit. on p. 52).
Bleistein, N., J. W. Stockwell and J. K. Cohen, 2001. Mathematics of Multidimensional Seismic Imaging, Migration, and Inversion. Red. by J. E. Marsden, L. Sirovich and S. Wiggins. Vol. 13. Springer New York (cit. on p. 14).
Bremmer, H., 1951. The W.K.B. Approximation as the First Term of a Geometric-Optical Series. Communications on Pure and Applied Mathematics, **4**(1), 105–115 (cit. on p. 36).
Brossier, R., S. Operto and J. Virieux, 2015. Velocity Model Building from Seismic Reflection Data by Full-Waveform Inversion. Geophysical Prospecting, **63**(2), 354–367 (cit. on p. 17).
Brougois, A., M. Bourget, P. Lailly, M. Poulet, P. Ricarte and R. Versteeg, 1990. Marmousi, Model and Data. EAEG Workshop-Practical Aspects of Seismic Data Inversion (cit. on p. 11).
Brown, M. P. and A. Guitton, 2005. Least-Squares Joint Imaging of Multiples and Primaries. Geophysics, **70**(5), S79–S89 (cit. on p. 42).
Bunks, C., F. Saleck, S. Zaleski and G. Chavent, 1995. Multiscale Seismic Waveform Inversion. Geophysics, **60**(5), 1457–1473 (cit. on pp. 10, 13).
Červený, V., 2005. Seismic Ray Theory. Revised edition. Charles University (cit. on pp. 9, 16).
Chauris, H. and E. Cocher, 2014. Iterative Migration for Velocity Analysis in the Presence of Surface-Related Multiples. Workshop on using multiples as signal for imaging, 2014 SEG Annual Meeting (cit. on p. 47).
– 2017. From Migration to Inversion Velocity Analysis. Geophysics, **82**(3), S207–S223 (cit. on pp. 46, 47, 106, 108, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137).
Chauris, H. and C.-A. Lameloise, 2014. Removing Spurious Oscillations in the Gradient of the Differential Semblance Optimization Functional. 76th EAGE Conference & Exhibition, Th P02 08 (cit. on pp. 25, 111).
Chauris, H., C.-A. Lameloise and E. Cocher, 2015. Inversion Velocity Analysis - The Importance of Regularisation. 77th EAGE Conference & Exhibition, WS05–A02 (cit. on pp. 24, 25, 47, 85, 112).
Chauris, H., C.-A. Lameloise and D. Donno, 2013. Migration Velocity Analysis with Reflected and Transmitted Waves. 75th EAGE Conference & Exhibition, We P01 01 (cit. on pp. 28, 111).

Chauris, H. and M. Noble, 2001. Two-Dimensional Velocity Macro Model Estimation from Seismic Reflection Data by Local Differential Semblance Optimization: Applications to Synthetic and Real Data Sets. Geophysical Journal International, **144**(1), 14–26 (cit. on pp. 20, 21, 23, 111, 229).

Chauris, H., M. S. Noble, G. Lambaré and P. Podvin, 2002. Migration Velocity Analysis from Locally Coherent Events in 2-D Laterally Heterogeneous Media, Part I: Theoretical Aspects. Geophysics, **67**(4), 1202–1212 (cit. on pp. 17, 20, 21).

Chauris, H. and R.-É. Plessix, 2013. Differential Waveform Inversion - A Way to Cope with Multiples? 75th EAGE Conference & Exhibition, F06 (cit. on p. 18).

Chavent, G. and C. Jacewitz, 1995. Determination of Background Velocities by Multiple Migration Fitting. Geophysics, **60**(2), 476–490 (cit. on pp. 18, 20).

Chavent, G., 2009. Nonlinear Least Squares for Inverse Problems. Springer Verlag (cit. on p. 10).

Chi, B., L. Dong and Y. Liu, 2014. Full Waveform Inversion Method Using Envelope Objective Function without Low Frequency Data. Journal of Applied Geophysics, **109**, 36–46 (cit. on p. 14).

Claerbout, J. F., 1971. Toward a Unified Theory of Reflector Mapping. Geophysics, **36**(3), 467–481 (cit. on pp. 14, 110).

– 1985. Imaging the Earth's Interior. Blackwell scientific publications Oxford (cit. on pp. 10, 11).

– 2014. Geophysical Image Estimation by Example (cit. on pp. 46, 266).

Cocher, E. and H. Chauris, 2014. Iterative Migration to Remove the Imprint of Multiples on the Reflectivity. SEG/KOC Workshop: Seismic multiples – Are they signal or noise?, Kuwait City, Kuwait (cit. on p. 47).

Cocher, E., H. Chauris and C.-A. Lameloise, 2015. Imaging with Surface-Related Multiples in the Subsurface-Offset Domain. 77th EAGE Conference & Exhibition, We N101 11 (cit. on pp. 47, 111).

Cocher, E., H. Chauris and R.-É. Plessix, 2017a. Seismic Iterative Migration Velocity Analysis: Two Strategies to Update the Velocity Model. Computational Geoscience. Submitted (cit. on pp. 47, 112, 133).

– 2017b. Towards a Stable Iterative Migration Velocity Analysis Scheme. Geophysical Journal International. In preparation (cit. on p. 47).

Colson, B., P. Marcotte and G. Savard, 2007. An Overview of Bilevel Optimization. Annals of Operations Research, **153**(1), 235–256 (cit. on p. 80).

De Bruin, C. G. M., C. P. A. Wapenaar and A. J. Berkhout, 1990. Angle-dependent Reflectivity by Means of Prestack Migration. Geophysics, **55**(9), 1223–1234 (cit. on p. 110).

De Hoop, M. V., 1996. Generalization of the Bremmer Coupling Series. Journal of Mathematical Physics, **37**(7), 3246–3282 (cit. on p. 36).

Devaney, A. J., 1982. A Filtered Backpropagation Algorithm for Diffraction Tomography. Ultrasonic Imaging, **4**(4), 336–350 (cit. on p. 13).

Díaz, E., 2016. Extended Imaging and Tomography under Two-Way Operators. Colorado School of Mines (cit. on pp. 245, 250).

Díaz, E., P. C. Sava and S. Singh, 2016. Extended Imaging, Deconvolution, and Two-Way Wavefields: A Comparison. 2016 SEG Annual Meeting, 4487–4492 (cit. on pp. 43, 245, 250).

Domenico, S. N., 1984. Rock Lithology and Porosity Determination from Shear and Compressional Wave Velocity. Geophysics, **49**(8), 1188–1195 (cit. on p. 12).

Donno, D., H. Chauris and M. Noble, 2010. Curvelet-Based Multiple Prediction. Geophysics, **75**(6), WB255–WB263 (cit. on p. 38).

Donno, D., H. Chauris and H. Calandra, 2013. Estimating the Background Velocity Model with the Normalized Integration Method. 75th EAGE Conference & Exhibition, Tu 07 04 (cit. on p. 14).

Dragoset, B., D. J. Verschuur, I. Moore and R. Bisley, 2010. A Perspective on 3D Surface-Related Multiple Elimination. Geophysics, **75**(5), 75A245–75A261 (cit. on p. 36).

Duquet, B. and K. J. Marfurt, 1999. Filtering Coherent Noise during Prestack Depth Migration. Geophysics, **64**(4), 1054–1066 (cit. on p. 38).

Duveneck, E., 2013. A Pragmatic Approach for Computing Full-Volume RTM Reflection Angle/Azimuth Gathers. 75th EAGE Conference & Exhibition, Tu 11 01 (cit. on p. 111).

Farra, V. and R. Madariaga, 1988. Non-Linear Reflection Tomography. Geophysical Journal International, **95**(1), 135–147 (cit. on p. 16).

Faye, J. and J. Jeannot, 1986. Prestack Migration Velocities from Focusing Depth Analysis. 1986 SEG Annual Meeting, 438–440 (cit. on p. 110).

Fei, W. and P. Williamson, 2010. On the Gradient Artifacts in Migration Velocity Analysis Based on Differential Semblance Optimization. 2010 SEG Annual Meeting, 4071–4076 (cit. on pp. 4, 25, 26, 108, 111, 112, 120, 128, 129, 139, 246).

Fichtner, A., 2011. Full Seismic Waveform Modelling and Inversion. Springer (cit. on p. 12).

Fletcher, R. and C. M. Reeves, 1964. Function Minimization by Conjugate Gradients. The Computer Journal, **7**(2), 149–154 (cit. on p. 260).

Fleury, C. and F. Perrone, 2012. Bi-Objective Optimization for the Inversion of Seismic Reflection Data: Combined FWI and MVA. 2012 SEG Annual Meeting, 1–6 (cit. on p. 252).

Foster, D. and C. Mosher, 1992. Suppression of Multiple Reflections Using the Radon Transform. Geophysics, **57**(3), 386–395 (cit. on p. 37).

Friedlander, M. P. and M. Schmidt, 2012. Hybrid Deterministic-Stochastic Methods for Data Fitting. SIAM Journal on Scientific Computing, **34**(3), A1380–A1405 (cit. on p. 88).

Fu, L. and W. W. Symes, 2015. Reducing the Cost of Extended Waveform Inversion by Multiscale Adaptive Methods. 2015 SEG Annual Meeting, 1127–1131 (cit. on p. 23).

Galuzzi, B., A. Tognarelli, E. Stucchi and A. Mazzotti, 2016. Stochastic FWI on Wide-Angle Land Data with Different Order of Approximation of the 2D Acoustic Wave Equation. 78th EAGE Conference and Exhibition, We SRS2 02 (cit. on p. 13).

Gauthier, O., J. Virieux and A. Tarantola, 1986. Two-dimensional Nonlinear Inversion of Seismic Waveforms: Numerical Results. Geophysics, **51**(7), 1387–1403 (cit. on p. 13).

Grau, G. and P. Lailly, 1993. Geophysical Exploration in Areas of Complex Geology, II Sequential Migration-Aided Reflection Tomography: An Approach to Imaging Complex Structures. Journal of Applied Geophysics, **30**(1), 75–87 (cit. on p. 16).

Guillaume, P., M. Reinier, G. Lambaré, A. Cavalié, M. Adamsen and B. Bruun, 2013. Dip-Constrained Non-Linear Slope Tomography - an Application to Shallow Channel Characterization. 75th EAGE Conference & Exhibition, Th 04 01 (cit. on p. 17).

Guitton, A., 2002. Shot-Profile Migration of Multiple Reflections. 2002 SEG Annual Meeting, 1296–1299 (cit. on p. 41).

Guitton, A. and W. W. Symes, 2003. Robust Inversion of Seismic Data Using the Huber Norm. Geophysics, **68**(4), 1310–1319 (cit. on pp. 149, 180).

Guitton, A. and D. J. Verschuur, 2004. Adaptive Subtraction of Multiples Using the L1-Norm. Geophysical Prospecting, **52**(1), 27–38 (cit. on p. 38).

Hager, W. W. and H. Zhang, 2006. A Survey of Nonlinear Conjugate Gradient Methods. Pacific journal of Optimization, **2**(1), 35–58 (cit. on p. 260).

Hampson, D., 1986. Inverse Velocity Stacking for Multiple Elimination. 1986 SEG Annual Meeting, 422–424 (cit. on p. 37).

Hansen, P. C., 1999. The L-Curve and Its Use in the Numerical Treatment of Inverse Problems. Technical University of Denmark (cit. on pp. 103, 180).

Holland, J. H., 1975. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. Vol. viii. U Michigan Press (cit. on p. 9).

Hou, J. and W. W. Symes, 2015. An Approximate Inverse to the Extended Born Modeling Operator. Geophysics, **80**(6), R331–R349 (cit. on pp. 5, 25, 44, 46, 106, 108, 109, 111–114, 132, 133, 136, 139, 144, 154, 210, 246).

– 2016a. Accelerating Extended Least-Squares Migration with Weighted Conjugate Gradient Iteration. Geophysics, **81**(4), S165–S179 (cit. on pp. 109, 132, 139, 140, 252).
– 2016b. Inversion Velocity Analysis via Approximate Born Inversion. 2016 SEG Annual Meeting, 5274–5279 (cit. on pp. 44, 71).
– 2017. An Alternative Formula for Approximate Extended Born Inversion. Geophysics, **82**(1), S1–S8 (cit. on pp. 5, 25, 246).
Huang, Y., 2016. Born Waveform Inversion in Shot Coordinate Domain. Rice University (cit. on pp. 26, 44, 46, 80, 82, 89, 98, 104, 188, 247).
Huang, Y. and W. W. Symes, 2015. Born Waveform Inversion via Variable Projection and Shot Record Model Extension. 2015 SEG Annual Meeting, 1326–1331 (cit. on pp. 20, 245, 252).
Ingber, L., 1989. Very Fast Simulated Re-Annealing. Mathematical and Computer Modelling, **12**(8), 967–973 (cit. on p. 9).
Jacobs, J., F. Delprat-Jannaud, A. Ehinger and P. Lailly, 1992. Sequential Migration-Aided Reflection Tomography: A Tool for Imaging Complex Structures. 1992 SEG Annual Meeting, 1054–1057 (cit. on p. 16).
Jakubowicz, H., 1998. Wave Equation Prediction and Removal of Interbed Multiples, 1527–1530 (cit. on p. 39).
Jannane, M., W. Beydoun, E. Crase, D. Cao, Z. Koren, E. Landa, M. Mendes, A. Pica, M. Noble, G. Roeth, S. Singh, R. Snieder, A. Tarantola, D. Trezeguet and M. Xie, 1989. Wavelengths of Earth Structures That Can Be Resolved from Seismic Reflection Data. Geophysics, **54**(7), 906–910 (cit. on p. 10).
Jiang, Z., J. Sheng, J. Yu, G. T. Schuster and B. E. Hornby, 2007. Migration Methods for Imaging Different-Order Multiples. Geophysical Prospecting, **55**(1), 1–19 (cit. on p. 41).
Jiang, Z., J. Yu, G. T. Schuster and B. E. Hornby, 2005. Migration of Multiples. The Leading Edge, **24**(3), 315–318 (cit. on p. 41).
Joncour, F., G. Lambaré and J. Svay-Lucas, 2011. Preserved-Amplitude Angle Domain Migration by Shot-Receiver Wavefield Continuation. Geophysical Prospecting, **59**(2), 256–268 (cit. on p. 133).
Kabir, M. M. N. and K. J. Marfurt, 1999. Toward True Amplitude Multiple Removal. The Leading Edge, **18**(1), 66–73 (cit. on p. 37).
Kabir, M. N. and D. J. Verschuur, 1995. Restoration of Missing Offsets by Parabolic Radon Transform. Geophysical Prospecting, **43**(3), 347–368 (cit. on p. 39).
Kennedy, J. and R. Eberhart, 1995. Particle Swarm Optimization. IEEE International Conference on Neural Networks, 1995. **4**, 1942–1948 (cit. on p. 10).
Kirkpatrick, S., C. D. Gelatt and M. P. Vecchi, 1983. Optimization by Simulated Annealing. Science, **220**(4598), 671–680 (cit. on p. 9).
Kiyashchenko, D., R.-É. Plessix, B. Kashtan and V. Troyan, 2007. A Modified Imaging Principle for True-Amplitude Wave-Equation Migration. Geophysical Journal International, **168**(3), 1093–1104 (cit. on p. 15).
Komatitsch, D. and R. Martin, 2007. An Unsplit Convolutional Perfectly Matched Layer Improved at Grazing Incidence for the Seismic Wave Equation. Geophysics, **72**(5), SM155–SM167 (cit. on pp. 55, 125).
Lailly, P. and D. Sinoquet, 1996. Smooth Velocity Models in Reflection Tomography for Imaging Complex Geological Structures. Geophysical Journal International, **124**(2), 349–362 (cit. on p. 16).
Lailly, P., 1983. The Seismic Inverse Problem as a Sequence of before Stack Migrations, 206–220 (cit. on p. 15).
Lambaré, G., J. Virieux, R. Madariaga and S. Jin, 1992. Iterative Asymptotic Inversion in the Acoustic Approximation. Geophysics, **57**(9), 1138–1154 (cit. on pp. 15, 25).
Lambaré, G., 2008. Stereotomography. Geophysics, **73**(5), VE25–VE34 (cit. on p. 17).
Lambaré, G., P. Guillaume and J.-P. Montel, 2014. Recent Advances in Ray-Based Tomography. 76th EAGE Conference & Exhibition, We G103 01 (cit. on pp. 11, 17).

Lameloise, C.-A., 2015. Analyse de Vitesse Par Migration Quantitative et Introduction Des Ondes Transmises. École nationale supérieure des Mines de Paris (cit. on pp. 23, 110, 251).
Lameloise, C.-A. and H. Chauris, 2014. Quantitative Migration for a More Robust Migration Velocity Analysis. 76th EAGE Conference & Exhibition, Th G103 13 (cit. on p. 25).
– 2016. Extension of Migration Velocity Analysis to Transmitted Wavefields. Geophysical Journal International, **207**(1), 343–356 (cit. on pp. 23, 28, 43, 111, 245, 251).
Lameloise, C.-A., H. Chauris and E. Cocher, 2015. Automatic Migration Velocity Analysis Applied to Direct Waves in a Crosswell Configuration. 77th EAGE Conference & Exhibition, We P1 05 (cit. on pp. 28, 47, 111).
Lameloise, C.-A., H. Chauris and M. Noble, 2014. Improving the Gradient of the Image-Domain Objective Function Using Quantitative Migration for a More Robust Migration Velocity Analysis. Geophysical Prospecting, **63**(2), 391–404 (cit. on pp. 5, 22, 24, 25, 30, 44, 50, 51, 71, 108, 110, 111, 246).
Le Stunff, Y. and D. Grenier, 1998. Taking into Account a Priori Information in 3D Tomography. 1998 SEG Annual Meeting, 1875–1878 (cit. on p. 17).
Levander, A., 1988. Fourth-order Finite-difference P-SV Seismograms. Geophysics, **53**(11), 1425–1436 (cit. on pp. 8, 55).
Li, J. and W. W. Symes, 2007. Interval Velocity Estimation via NMO-Based Differential Semblance. Geophysics, **72**(6), U75–U88 (cit. on pp. 28, 43, 246).
Li, V., I. Tsvankin and T. Alkhalifah, 2016. Analysis of RTM Extended Images for VTI Media. Geophysics, **81**(3), S139–S150 (cit. on p. 110).
Lines, L., A. Schultz and S. Treitel, 1988. Cooperative Inversion of Geophysical Data. Geophysics, **53**(1), 8–20 (cit. on p. 17).
Lippmann, B. A., 1956. Rearrangement Collisions. Physical Review, **102**(1), 264–268 (cit. on p. 36).
Liu, Y., X. Chang, D. Jin, R. He, H. Sun and Y. Zheng, 2011. Reverse Time Migration of Multiples for Subsalt Imaging. Geophysics, **76**(5), WB209–WB216 (cit. on pp. 40, 42).
Liu, Y., W. W. Symes and Z. Li, 2014a. Extended Reflection Waveform Inversion via Differential Semblance Optimization. 2014 SEG Annual Meeting (cit. on p. 41).
– 2014b. Inversion Velocity Analysis Via Differential Semblance Optimization. 76th EAGE Conference & Exhibition, Th P02 07 (cit. on pp. 23, 111, 112, 251, 271).
Long, A., S. Lu, D. Whitmore, H. LeGleut, R. Jones, N. Chemingui and M. Farouki, 2013. Mitigation of the 3D Cross-Line Acquisition Footprint Using Separated Wavefield Imaging of Dual-Sensor Streamer Seismic. 75th EAGE Conference & Exhibition, Th 01 05 (cit. on p. 39).
Luo, S. and P. C. Sava, 2011. A Deconvolution-Based Objective Function For Wave-Equation Inversion. 2011 SEG Annual Meeting, 2788–2792 (cit. on p. 17).
Luo, Y. and G. T. Schuster, 1991. Wave-equation Traveltime Inversion. Geophysics, **56**(5), 645–653 (cit. on p. 17).
Métivier, L., R. Brossier, Q. Mérigot, E. Oudet and J. Virieux, 2016. Measuring the Misfit between Seismograms Using an Optimal Transport Distance: Application to Full Waveform Inversion. Geophysical Journal International, **205**(1), 345–377 (cit. on p. 14).
Métivier, L., F. Bretaudeau, R. Brossier, S. Operto and J. Virieux, 2014. Full Waveform Inversion and the Truncated Newton Method: Quantitative Imaging of Complex Subsurface Structures. Geophysical Prospecting, **62**(6), 1353–1375 (cit. on pp. 251, 261).
Métivier, L., R. Brossier and J. Virieux, 2015. Combining Asymptotic Linearized Inversion and Full Waveform Inversion. Geophysical Journal International, **201**(3), 1682–1703 (cit. on p. 140).
Métivier, L., R. Brossier, J. Virieux and S. Operto, 2013. Full Waveform Inversion and the Truncated Newton Method. SIAM Journal on Scientific Computing, **35**(2), B401–B437 (cit. on pp. 87, 93, 200, 251, 261, 271).
Miller, D., M. Oristaglio and G. Beylkin, 1987. A New Slant on Seismic Imaging: Migration and Integral Geometry. Geophysics, **52**(7), 943–964 (cit. on p. 13).

Mora, P., 1988. Elastic Wave-field Inversion of Reflection and Transmission Data. Geophysics, **53**(6), 750–759 (cit. on p. 13).
– 1989. Inversion = Migration + Tomography. Geophysics, **54**(12), 1575–1586 (cit. on p. 13).
Moré, J. J. and D. J. Thuente, 1994. Line Search Algorithms with Guaranteed Sufficient Decrease. ACM Trans. Math. Softw. **20**(3), 286–307 (cit. on pp. 50, 66, 149, 259).
Muijs, R., J. O. A. Robertsson and K. Holliger, 2007. Prestack Depth Migration of Primary and Surface-Related Multiple Reflections: Part I — Imaging. Geophysics, **72**(2), S59–S69 (cit. on pp. 41, 42).
Mulder, W. A., 2008. Automatic Velocity Analysis with the Two-Way Wave Equation. 70th EAGE Conference & Exhibition, P165 (cit. on p. 22).
– 2014. Subsurface Offset Behaviour in Velocity Analysis with Extended Reflectivity Images. Geophysical Prospecting, **62**(1), 17–33 (cit. on pp. 22–25, 50, 51, 71, 110, 111, 120, 246).
Mulder, W. A. and F. ten Kroode, 2002. Automatic Velocity Analysis by Differential Semblance Optimization. Geophysics, **67**(4), 1184–1191 (cit. on pp. 23, 28, 43, 111, 246).
Mulder, W. A. and T. van Leeuwen, 2008. Automatic Migration Velocity Analysis and Multiples. 2008 SEG Annual Meeting, 3128–3132 (cit. on p. 28).
Nasyrov, D., D. Kiyashchenko, Y. Kiselev, B. Kashtan and V. Troyan, 2008. Velocity Analysis for VSP Data Using Multiples. 2008 SEG Annual Meeting, 3420–3424 (cit. on p. 43).
– 2009. Multiple Migration of VSP Data for Velocity Analysis. 2009 SEG Annual Meeting, 4164–4168 (cit. on p. 43).
Nekut, A. and D. J. Verschuur, 1998. Minimum Energy Adaptive Subtraction in Surface-Related Multiple Attenuation. 1998 SEG Annual Meeting, 1507–1510 (cit. on p. 38).
Nemeth, T., C. Wu and G. Schuster, 1999. Least-squares Migration of Incomplete Reflection Data. Geophysics, **64**(1), 208–221 (cit. on pp. 15, 51, 253).
Nichols, D., 2012. Resolution in Seismic Inversion- Spectral Gap or Spectral Overlap, Which Is Harder to Handle? 74th EAGE Conference & Exhibition, E12 (cit. on pp. 11, 17).
Noble, M., A. Gesret and N. Belayouni, 2014. Accurate 3-D Finite Difference Computation of Traveltimes in Strongly Heterogeneous Media. Geophysical Journal International, **199**(3), 1572–1585 (cit. on p. 16).
Noble, M., 1992. Inversion Non Linéaire de Données de Prospection Pétroliere. Paris 7 (cit. on p. 120).
Nocedal, J., 1980. Updating Quasi-Newton Matrices with Limited Storage. Mathematics of Computation, **35**(151), 773–782 (cit. on p. 261).
Nocedal, J. and S. J. Wright, 2006. Numerical Optimization. 2nd edition. Springer (cit. on pp. 10, 19, 50, 65, 66, 142, 176, 257, 259–261).
Operto, S., Y. Gholami, V. Prieux, A. Ribodetti, R. Brossier, L. Metivier and J. Virieux, 2013. A Guided Tour of Multiparameter Full-Waveform Inversion with Multicomponent Data: From Theory to Practice. The Leading Edge, **32**(9), 1040–1054 (cit. on pp. 10, 12).
Operto, S., J. Virieux, P. Amestoy, J. L'Excellent, L. Giraud and H. Ali, 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 (cit. on p. 8).
Østmo, S., W. A. Mulder and R.-É. Plessix, 2002. Finite-Difference Iterative Migration By Linearized Waveform Inversion In the Frequency Domain. 2002 SEG Annual Meeting, 1384–1387 (cit. on pp. 15, 51).
Pérez Solano, C. A., 2013. Imagerie Sismique de La Proche Sub-Surface : Modification de l'inversion Des Formes d'onde Pour l'analyse Des Ondes de Surface. École nationale supérieure des Mines de Paris (cit. on p. 7).
Plessix, R.-É., 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 (cit. on pp. 10, 19, 46, 50, 67, 80, 83, 115, 136, 271).
– 2009. Three-Dimensional Frequency-Domain Full-Waveform Inversion with an Iterative Solver. Geophysics, **74**(6), WCC149–WCC157 (cit. on p. 112).

Plessix, R., G. Baeten, J. W. de Maag, M. Klaassen, Z. Rujie and T. Zhifei, 2010. Application of Acoustic Full Waveform Inversion to a Low-frequency Large-offset Land Data Set. 2010 SEG Annual Meeting, 930–934 (cit. on p. 14).

Plessix, R.-É., Y.-H. De Roeck and G. Chavent, 1995. Automatic and Simultaneous Migration Velocity Analysis and Waveform Inversion of Real Data Using a MBTT/WKB J Formulation. 1995 SEG Annual Meeting, 1099–1102 (cit. on p. 18).

Plessix, R.-É. and W. A. Mulder, 2004. Frequency-Domain Finite-Difference Amplitude-Preserving Migration. Geophysical Journal International, **157**(3), 975–987 (cit. on p. 15).

Podvin, P. and I. Lecomte, 1991. Finite Difference Computation of Traveltimes in Very Contrasted Velocity Models: A Massively Parallel Approach and Its Associated Tools. Geophysical Journal International, **105**(1), 271–284 (cit. on p. 16).

Polak, E. and G. Ribiere, 1969. Note Sur La Convergence de Méthodes de Directions Conjuguées. Revue française d'informatique et de recherche opérationnelle, série rouge, **3**(1), 35–43 (cit. on p. 260).

Poole, T. L., A. Curtis, J. O. Robertsson and D.-J. van Manen, 2010. Deconvolution Imaging Conditions and Cross-Talk Suppression. Geophysics, **75**(6), W1–W12 (cit. on p. 42).

Pratt, R. G., Z.-M. Song, P. Williamson and M. Warner, 1996. Two-Dimensional Velocity Models from Wide-Angle Seismic Data by Wavefield Inversion. Geophysical Journal International, **124**(2), 323–340 (cit. on p. 13).

Reinier, M., A. Cavalié, P. Guillaume, G. Lambaré and F. Blom, 2012. Building a Robust Depth Imaging Velocity Model in North Sea. 2012 SEG Annual Meeting, 1–5 (cit. on p. 33).

Reiter, E. C., M. N. Toksöz, T. H. Keho and G. M. Purdy, 1991. Imaging with Deep-Water Multiples. Geophysics, **56**(7), 1081–1086 (cit. on p. 41).

Rickett, J., 2013. The Variable Projection Method for Waveform Inversion with an Unknown Source Function. Geophysical Prospecting, **61**(4), 874–881 (cit. on p. 252).

Rickett, J. E. and P. C. Sava, 2002. Offset and Angle-domain Common Image-point Gathers for Shot-profile Migration. Geophysics, **67**(3), 883–889 (cit. on p. 21).

Robertsson, J., 1996. A Numerical Free-surface Condition for Elastic/Viscoelastic Finite-difference Modeling in the Presence of Topography. Geophysics, **61**(6), 1921–1934 (cit. on p. 55).

Rose, J. H., 2002. 'Single-Sided' Autofocusing of Sound in Layered Materials. Inverse Problems, **18**(6), 1923 (cit. on p. 43).

Ryu, J. V., 1982. Decomposition (DECOM) Approach Applied to Wave Field Analysis with Seismic Reflection Records. Geophysics, **47**(6), 869–883 (cit. on p. 37).

Sacchi, M. and T. Ulrych, 1995. High-resolution Velocity Gathers and Offset Space Reconstruction. Geophysics, **60**(4), 1169–1177 (cit. on p. 38).

Sava, P. C. and B. Biondi, 2004. Wave-Equation Migration Velocity Analysis. I. Theory. Geophysical Prospecting, **52**(6), 593–606 (cit. on p. 111).

Sava, P. C. and S. Fomel, 2003. Angle-domain Common-image Gathers by Wavefield Continuation Methods. Geophysics, **68**(3), 1065–1074 (cit. on p. 21).

– 2006. Time-Shift Imaging Condition in Seismic Migration. Geophysics, **71**(6), S209–S217 (cit. on pp. 21, 110, 113, 115).

Sava, P. C. and A. Guitton, 2005. Multiple Attenuation in the Image Space. Geophysics, **70**(1), V10–V20 (cit. on pp. 37, 38).

Sava, P. C. and I. Vasconcelos, 2011. Extended Imaging Conditions for Wave-Equation Migration. Geophysical Prospecting, **59**(1), 35–55 (cit. on pp. 21, 113).

Schuster, G. T., J. Yu, J. Sheng and J. Rickett, 2004. Interferometric/Daylight Seismic Imaging. Geophysical Journal International, **157**(2), 838–852 (cit. on p. 41).

Schuster, G. T., 2007. Basics of Seismic Wave Theory. University of Utah (cit. on p. 255).

Sen, M. K. and P. L. Stoffa, 2013. Global Optimization Methods in Geophysical Inversion. Second edition. Cambridge University Press (cit. on p. 9).

Sevink, A. G. J. and G. C. Herman, 1996. Three-Dimensional, Nonlinear, Asymptotic Seismic Inversion. Inverse Problems, **12**(5), 757 (cit. on p. 140).
Shan, G., 2003. Source-Receiver Migration of Multiple Reflections. 2003 SEG Annual Meeting, 1008–1011 (cit. on p. 41).
Shen, P., 2013. Subsurface Focusing Measurement of Diving Waves and Its Application to Reflection Tomography. 75th EAGE Conference & Exhibition, Th 10 05 (cit. on pp. 43, 111).
Shen, P. and W. W. Symes, 2008. Automatic Velocity Analysis via Shot Profile Migration. Geophysics, **73**(5), VE49–VE59 (cit. on pp. 21–23, 111).
– 2013. Subsurface Domain Image Warping by Horizontal Contraction and Its Application to Wave-Equation Migration Velocity Analysis. 2013 SEG Annual Meeting, 4715–4719 (cit. on p. 28).
– 2015. Horizontal Contraction in Image Domain for Velocity Inversion. Geophysics, **80**(3), R95–R110 (cit. on pp. 4, 23, 26, 108, 111, 120, 124, 128, 129, 134, 139, 246, 251, 271).
Shen, P., W. W. Symes and C. C. Stolk, 2003. Differential Semblance Velocity Analysis by Wave-Equation Migration. 2003 SEG Annual Meeting, 2132–2135 (cit. on p. 21).
Sheng, J., 2001. Migrating Multiples and Primaries in CDP Data by Crosscorrelogram Migration. 2001 SEG Annual Meeting, 1297–1300 (cit. on p. 41).
Sheriff, R. E. and L. P. Geldart, 2006. Exploration Seismology. 2. ed., re-issued in this digitally print. monochrome version. Cambridge Univ. Press (cit. on p. 6).
Shin, C. and Y. H. Cha, 2008. Waveform Inversion in the Laplace Domain. Geophysical Journal International, **173**(3), 922–931 (cit. on p. 14).
– 2009. Waveform Inversion in the Laplace-Fourier Domain. Geophysical Journal International, **177**(3), 1067–1079 (cit. on p. 14).
Shipp, R. M. and S. C. Singh, 2002. Two-Dimensional Full Wavefield Inversion of Wide-Aperture Marine Seismic Streamer Data. Geophysical Journal International, **151**(2), 325–344 (cit. on p. 13).
Singh, S., R. Snieder, J. Behura, J. van der Neut, K. Wapenaar and E. Slob, 2015. Marchenko Imaging: Imaging with Primaries, Internal Multiples, and Free-Surface Multiples. Geophysics, **80**(5), S165–S174 (cit. on pp. 43, 250).
Sirgue, L., O. I. Barkved, J. P. Van Gestel, O. J. Askim and J. H. Kommedal, 2009. 3D Waveform Inversion on Valhall Wide-Azimuth OBC. 71st EAGE Conference and Exhibition, U038 (cit. on p. 14).
Sirgue, L. and R. G. Pratt, 2004. Efficient Waveform Inversion and Imaging: A Strategy for Selecting Temporal Frequencies. Geophysics, **69**(1), 231–248 (cit. on pp. 13, 215).
Snieder, R., 2002. Elastic Coda Wave Interferometry, a New Tool for the Instrumented Oilfield. 2002 SEG Annual Meeting, 1650–1653 (cit. on p. 39).
Socco, L. V. and C. Strobbia, 2004. Surface-Wave Method for near-Surface Characterization: A Tutorial. Near Surface Geophysics, **2**(22), 165–185 (cit. on p. 7).
Soni, A. and D. J. Verschuur, 2014. Full-Wavefield Migration of Vertical Seismic Profiling Data: Using All Multiples to Extend the Illumination Area: Full-Wavefield Migration of Vertical Seismic Profiling Data. Geophysical Prospecting, **62**(4), 740–759 (cit. on p. 42).
Staal, X. R. and D. J. Verschuur, 2012. Velocity Estimation Using Internal Multiples. 2012 SEG Annual Meeting, 1–5 (cit. on pp. 43, 111).
Staal, X. R., 2015. Combined Imaging and Velocity Estimation by Joint Migration Inversion. Technische Universiteit Delft (cit. on p. 43).
Stolk, C. C. and M. V. de Hoop, 2005. Modeling of Seismic Data in the Downward Continuation Approach. SIAM Journal on Applied Mathematics, **65**(4), 1388–1406 (cit. on p. 21).
Stolk, C. C., M. V. de Hoop and W. W. Symes, 2009. Kinematics of Shot-Geophone Migration. Geophysics, **74**(6), WCA19–WCA34 (cit. on pp. 112, 132).
Stolk, C. C. and W. W. Symes, 2003. Smooth Objective Functionals for Seismic Velocity Inversion. Inverse Problems, **19**(1), 73–89 (cit. on p. 21).
– 2004. Kinematic Artifacts in Prestack Depth Migration. Geophysics, **69**(2), 562–575 (cit. on p. 21).

Stork, C., 1992. Reflection Tomography in the Postmigrated Domain. Geophysics, **57**(5), 680–692 (cit. on p. 16).
Symes, W. W., 2008. Migration Velocity Analysis and Waveform Inversion. Geophysical Prospecting, **56**(6), 765–790 (cit. on pp. 9, 11, 19, 110, 111, 113).
– 2014. Seismic Inverse Poblems: Recent Developments in Theory and Practice. Inverse Problems - from Theory to Applications, 2–6 (cit. on p. 52).
Symes, W. W. and J. J. Carazzone, 1991. Velocity Inversion by Differential Semblance Optimization. Geophysics, **56**(5), 654–663 (cit. on pp. 20, 110, 111).
Symes, W. W. and M. Kern, 1994. Inversion of Reflection Seismograms by Differential Semblance Analysis: Algorithm Structure and Synthetic Examples. Geophysical Prospecting, **42**(6), 565–614 (cit. on pp. 20, 21, 252).
Tang, Y., S. Lee, A. Baumstein and D. Hinkley, 2013. Tomographically Enhanced Full Wavefield Inversion. 2013 SEG Annual Meeting, 1037–1041 (cit. on p. 133).
Tarantola, A., 1984. Inversion of Seismic Reflection Data in the Acoustic Approximation. Geophysics, **49**(8), 1259–1266 (cit. on pp. 9, 12, 15).
– 2005. Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM, Society for Industrial and Applied Mathematics (cit. on p. 9).
Tatham, R. and P. Stoffa, 1976. Vp/Vs — A Potential Hydrocarbon Indicator. Geophysics, **41**(5), 837–849 (cit. on p. 12).
Tejero, C. E. J., D. Dagnino, V. Sallarès and C. R. Ranero, 2015. Comparative Study of Objective Functions to Overcome Noise and Bandwidth Limitations in Full Waveform Inversion. Geophysical Journal International, **203**(1), 632–645 (cit. on p. 14).
Ten Kroode, F., 2002. Prediction of Internal Multiples. Wave Motion, **35**(4), 315–338 (cit. on pp. 36, 39).
– 2012. A Wave-Equation-Based Kirchhoff Operator. Inverse Problems, **28**(11), 115013 (cit. on pp. 5, 44, 111–113, 132, 133, 135, 136).
Tognarelli, A., E. Stucchi, N. Bienati, A. Sajeva, M. Aleardi and A. Mazzotti, 2015. Two-Grid Stochastic Full Waveform Inversion of 2D Marine Seismic Data. 77th EAGE Conference and Exhibition, Th N104 01 (cit. on p. 13).
Trad, D., T. Ulrych and M. Sacchi, 2003. Latest Views of the Sparse Radon Transform. Geophysics, **68**(1), 386–399 (cit. on pp. 35, 38).
Tu, N. and F. J. Herrmann, 2015. Fast Imaging with Surface-Related Multiples by Sparse Inversion. Geophysical Journal International, **201**(1), 304–317 (cit. on p. 42).
Tu, N., T. van Leeuwen and F. J. Herrmann, 2013. Limitations of the Deconvolutional Imaging Condition for Two-Way Propagators. 2013 SEG Annual Meeting, 3916–3920 (cit. on p. 42).
Van Groenestijn, G. J. and D. J. Verschuur, 2009. Estimating Primaries by Sparse Inversion and Application to Near-Offset Data Reconstruction. Geophysics, **74**(3), A23–A28 (cit. on p. 39).
Van Leeuwen, T., R. Kumar and F. J. Herrmann, 2015. Affordable Full Subsurface Image Volume - An Application to WEMVA. 77th EAGE Conference & Exhibition, WS05–C01 (cit. on pp. 23, 111).
Van Leeuwen, T. and W. A. Mulder, 2008a. Velocity Analysis Based on Data Correlation. Geophysical Prospecting, **56**(6), 791–803 (cit. on pp. 18, 43).
– 2008b. Velocity Analysis with Multiples — NMO Modeling for Layered Velocity Structures. 2008 SEG Annual Meeting, 1925–1929 (cit. on p. 43).
– 2009. A Variable Projection Method for Waveform Inversion. 71st EAGE Conference and Exhibition, U024 (cit. on p. 252).
– 2010a. A Comparison of Seismic Velocity Inversion Methods for Layered Acoustics. Inverse Problems, **26**(1), 015008 (cit. on p. 21).
– 2010b. A Correlation-Based Misfit Criterion for Wave-Equation Traveltime Tomography. Geophysical Journal International, **182**(3), 1383–1394 (cit. on p. 17).

Verschuur, D. J., 2013. Seismic Multiple Removal Techniques: Past, Present and Future. Revised edition. 1. EAGE Publication (cit. on pp. 9, 33–35, 37).
Verschuur, D. J. and A. J. Berkhout, 1997. Estimation of Multiple Scattering by Iterative Inversion, Part II: Practical Aspects and Examples. Geophysics, **62**(5), 1596–1611 (cit. on p. 4).
Verschuur, D. J. and A. J. Berkhout, 2011. Seismic Migration of Blended Shot Records with Surface-Related Multiple Scattering. Geophysics, **76**(1), A7–A13 (cit. on p. 42).
Verschuur, D. J., A. J. Berkhout and C. P. A. Wapenaar, 1992. Adaptive Surface-Related Multiple Elimination. Geophysics, **57**(9), 1166–1177 (cit. on pp. 36, 38).
Verschuur, D. J. and X. R. Staal, 2014. Using Primaries and Multiples in Time-Lapse Imaging and Velocity Estimation. 2014 SEG Annual Meeting, 4955–4959 (cit. on p. 39).
Vi Nhu Ba, E., 2014. Détection Des Zones de Failles Par Tomographie En Transmission : Application à La Station Expérimentale de Tournemire. École nationale supérieure des Mines de Paris (cit. on p. 16).
Vidale, J., 1988. Finite-Difference Calculation of Travel Times. Bulletin of the Seismological Society of America, **78**(6), 2062–2076 (cit. on p. 16).
Vigh, D. V., W. E. S. Starr and K. D. Kenneth Dingwall, 2009. 3D Prestack Time Domain Full Waveform Inversion. 71st EAGE Conference and Exhibition, U043 (cit. on p. 14).
Virieux, J., 1986. P-SV Wave Propagation in Heterogeneous Media: Velocity-stress Finite-difference Method. Geophysics, **51**(4), 889–901 (cit. on p. 8).
Virieux, J. and S. Operto, 2009. An Overview of Full-Waveform Inversion in Exploration Geophysics. Geophysics, **74**(6), WCC1–WCC26 (cit. on pp. 10, 12, 215).
Virieux, J., H. Calandra and R.-É. Plessix, 2011. A Review of the Spectral, Pseudo-Spectral, Finite-Difference and Finite-Element Modelling Techniques for Geophysical Imaging. Geophysical Prospecting, **59**(5), 794–813 (cit. on p. 8).
Vyas, M. and Y. Tang, 2010. Gradients for Wave-equation Migration Velocity Analysis. 2010 SEG Annual Meeting, 4077–4081 (cit. on pp. 25, 111, 112, 120, 129).
Wang, H., S. C. Singh and H. Calandra, 2014. Integrated Inversion Using Combined Wave-Equation Tomography and Full Waveform Inversion. Geophysical Journal International, **198**(1), 430–446 (cit. on p. 17).
Wapenaar, K., F. Broggini and R. Snieder, 2011. A Proposal for Model–independent 3D Wave Field Reconstruction from Reflection Data. 2011 SEG Annual Meeting, 3788–3792 (cit. on p. 43).
Wapenaar, K., J. Thorbecke, J. van der Neut, F. Broggini, E. Slob and R. Snieder, 2014. Marchenko Imaging. Geophysics, **79**(3), WA39–WA57 (cit. on pp. 42, 250).
Warner, M. and L. Guasch, 2014. Adaptive Waveform Inversion - FWI Without Cycle Skipping - Theory. 76th EAGE Conference & Exhibition, We E106 13 (cit. on p. 17).
– 2015. Adaptive Waveform Inversion Using Incomplete Physics, Imperfect Data, and an Incorrect Source. 75th EAGE Conference & Exhibition, Th N104 10 (cit. on p. 17).
Warner, M., J. Morgan, A. Umpleby, I. Stekl and L. Guasch, 2012. Which Physics for Full-Wavefield Seismic Inversion? 74th EAGE Conference & Exhibition, W029 (cit. on p. 12).
Weglein, A. B., F. A. Gasparotto, P. M. Carvalho and R. H. Stolt, 1997. An Inverse-scattering Series Method for Attenuating Multiples in Seismic Reflection Data. Geophysics, **62**(6), 1975–1989 (cit. on pp. 36, 39).
Weibull, W. and B. Arntsen, 2011. Automatic Migration Velocity Analysis Using Reverse Time Migration. 73rd EAGE Conference & Exhibition, B012 (cit. on pp. 24, 25).
– 2013. Anisotropic Migration Velocity Analysis Using Reverse-Time Migration. Geophysics, **79**(1), R13–R25 (cit. on pp. 111, 133).
Weibull, W., B. Arntsen and M. Houbiers, 2012a. Reverse Time Migration Velocity Analysis - A Real Field Data Example. 74th EAGE Conference & Exhibition, P234 (cit. on p. 23).
Weibull, W., B. Arntsen, M. Houbiers and J. Mispel, 2012b. Automatic Anisotropic Migration Velocity Analysis for Reverse-Time Migration. 2012 SEG Annual Meeting (cit. on p. 23).

Whitmore, N. D., A. A. Valenciano and W. Söllner, 2010. Imaging of Primaries and Multiples Using a Dual-Sensor Towed Streamer. 2010 SEG Annual Meeting, 3187–3192 (cit. on pp. 41, 42).
Wiggins, J. W., 1988. Attenuation of Complex Water-bottom Multiples by Wave-equation-based Prediction and Subtraction. Geophysics, **53**(12), 1527–1539 (cit. on p. 36).
– 1999. Multiple Attenuation by Explicit Wave Extrapolation to an Interpreted Horizon. The Leading Edge, **18**(1), 46–54 (cit. on p. 36).
Williamson, P. R., 1990. Tomographic Inversion In Reflection Seismology. Geophysical Journal International, **100**(2), 255–274 (cit. on p. 16).
Wirgin, A., 2004. The Inverse Crime (cit. on p. 210).
Wong, M., B. Biondi and S. Ronen, 2014. Imaging with Multiples Using Least-Squares Reverse Time Migration. The Leading Edge, **33**(9), 970–976 (cit. on p. 42).
Woodward, M. J., D. Nichols, O. Zdraveva, P. Whitfield and T. Johns, 2008. A Decade of Tomography. Geophysics, **73**(5), VE5–VE11 (cit. on p. 17).
Wu, R. and L. Chen, 2006. Directional Illumination Analysis Using Beamlet Decomposition and Propagation. Geophysics, **71**(4), S147–S159 (cit. on p. 133).
Wu, R.-S., J. Luo and B. Wu, 2014. Seismic Envelope Inversion and Modulation Signal Model. Geophysics, **79**(3), WA13–WA24 (cit. on p. 14).
Wu, Z. and T. Alkhalifah, 2015. Simultaneous Inversion of the Background Velocity and the Perturbation in Full-Waveform Inversion. Geophysics, **80**(6), R317–R329 (cit. on pp. 112, 119).
Xu, S., D. Wang, F. Chen, G. Lambaré and Y. Zhang, 2012. Inversion on Reflected Seismic Wave. 2012 SEG Annual Meeting, 1–7 (cit. on pp. 17, 18).
Xu, S., H. Chauris, G. Lambaré and M. Noble, 2001. Common-angle Migration: A Strategy for Imaging Complex Media. Geophysics, **66**(6), 1877–1894 (cit. on p. 21).
Al-Yahya, K., 1989. Velocity Analysis by Iterative Profile Migration. Geophysics, **54**(6), 718–729 (cit. on pp. 9, 18, 19).
Yang, T., J. Shragge and P. C. Sava, 2013. Illumination Compensation for Image-Domain Wavefield Tomography. Geophysics, **78**(5), U65–U76 (cit. on pp. 24, 133).
Yang, T. and P. C. Sava, 2011. Wave-Equation Migration Velocity Analysis with Time-Shift Imaging. Geophysical Prospecting, **59**(4), 635–650 (cit. on p. 21).
– 2015. Image-Domain Wavefield Tomography with Extended Common-Image-Point Gathers. Geophysical Prospecting, **63**(5), 1086–1096 (cit. on pp. 22, 23, 30).
Yilmaz, Ö., 2001. Seismic Data Analysis. Society of Exploration Geophysicists (cit. on pp. 6, 9).
Zhang, D. and G. T. Schuster, 2014. Least-Squares Reverse Time Migration of Multiples. Geophysics, **79**(1), S11–S21 (cit. on p. 42).
Zhang, Y., G. Zhang and N. Bleistein, 2005. Theory of True-Amplitude One-Way Wave Equations and True-Amplitude Common-Shot Migration. Geophysics, **70**(4), E1–E10 (cit. on p. 136).
Zhang, Y., S. Xu, B. Tang, B. Bai, Y. Huang and T. Huang, 2010. Angle Gathers from Reverse Time Migration. The Leading Edge, **29**(11), 1364–1371 (cit. on pp. 20, 21).
Zhou, H., L. Amundsen and G. Zhang, 2012. Fundamental Issues in Full Waveform Inversion. 2012 SEG Annual Meeting, 1–5 (cit. on p. 17).
Zhou, W., 2016. Velocity Model Building by Full Waveform Inversion of Early Arrivals & Reflections and Case Study with Gas Cloud Effect. Université de Grenoble Joseph Fourier (cit. on pp. 10, 18, 251).
Zhou, W., R. Brossier, S. Operto and J. Virieux, 2015. Full Waveform Inversion of Diving & Reflected Waves for Velocity Model Building with Impedance Inversion Based on Scale Separation. Geophysical Journal International, **202**(3), 1535–1554 (cit. on pp. 10, 18, 251).

Résumé

Les expériences de sismique active sont couramment utilisées pour estimer la valeur d'un modèle de vitesse de propagation des ondes P dans le sous-sol. Les méthodes dites d'« analyse de vitesse par migration » ont pour but la détermination d'un macro-modèle de vitesse, lisse, et responsable de la cinématique de propagation des ondes. Dans une première étape de « migration », une image de réflectivité est obtenue à partir des données enregistrées en utilisant une première estimation du macro-modèle. Cette image dépend d'un paramètre additionnel permettant dans un second temps d'estimer la qualité du macro-modèle puis de l'améliorer. Les images de réflectivité obtenues par les techniques de migration classiques sont cependant contaminées par des artefacts, altérant la qualité de la remise à jour du macro-modèle. En particulier, elles ne prennent pas en compte les réflexions multiples, habituellement retirées des données avant traitement. Cette étape reste cependant délicate et on se prive alors de l'information supplémentaire contenue dans les multiples.

Nous proposons dans cette étude une stratégie d'optimisation imbriquée en itérant l'étape de migration avant de remettre à jour le macro-modèle. La migration itérative produit des images de réflectivité satisfaisantes pour l'analyse de vitesse et s'étend naturellement aux réflexions multiples. Un désavantage de la méthode est son coût de calcul. Un pseudo-inverse de l'opérateur de modélisation est alors utilisé comme préconditionneur pour limiter le nombre d'itérations dans la boucle interne. Une autre difficulté est l'instabilité de la remise à jour du modèle de vitesse calculée pour des modèles de réflectivité successifs proches les uns des autres. Une nouvelle approche plus robuste est proposée, valide aussi dans le cas de multiples. Son efficacité est testée sur des jeux de données synthétiques 2D.

Mots Clés

imagerie sismique • analyse de vitesse par migration • réflexions multiples • migration itérative • problèmes inverses

Abstract

Active seismic experiments are commonly used to recover a model of the P-wave propagation velocity in the subsurface. "Migration Velocity Analysis" techniques aim at deriving a smooth background velocity model controlling the kinematics of wave propagation. First, a reflectivity image is obtained by "migration" of observed data using a first estimate of the background velocity. This image depends on an additional "subsurface-offset" parameter allowing to assess the quality of the background velocity model with a focusing criterion and to correct it. However classical migration techniques do not provide a sufficiently accurate reflectivity image, leading to inconsistent velocity updates. In particular they do not take into account multiple reflections, usually regarded as noise and removed from the data before processing. Multiple removal is however a difficult step, and additional information contained in multiples is discarded.

In this thesis, we propose to determine the reflectivity model by iterative migration before subsequent velocity analysis, leading to a nested optimisation procedure. Iterative migration yields accurate reflectivity image and extends naturally to the case of multiples. One of its disadvantages is the associated increased computational cost. To limit the number of iterations in the inner loop, a preconditioner based on a pseudo-inverse of the modelling operator is introduced. Another difficulty is the instability of the velocity update obtained with very close successive reflectivity models. We propose a modified approach, valid in the presence of multiples, and discussed through applications on 2D synthetic data sets.

Keywords

seismic imaging • migration velocity analysis • multiple reflections • iterative migration • inverse problems